Psychon Bull Rev
DOI 10.3758/s13423-015-0978-1

THEORETICAL REVIEW

Serial vs. parallel models of attention in visual search: accounting for benchmark RT-distributions

Rani Moran[1,4] · **Michael Zehetleitner**[2] · **Heinrich René Liesefeld**[2] · **Hermann J. Müller**[2,3] · **Marius Usher**[1]

Abstract Visual search is central to the investigation of selective visual attention. Classical theories propose that display items are identified as focal attention is deployed *serially* to their locations based on their salience. While this accounts for set-size effects over a continuum of task difficulties, it has been suggested that parallel models can account for such effects equally well. We compared the serial Competitive Guided Search model with a parallel model in their ability to account for RT distributions and error rates from a large visual search data-set featuring three classical search tasks: 1) a spatial configuration search (2 vs. 5); 2) a feature-conjunction search; and 3) a unique feature search (Wolfe, Palmer & Horowitz *Vision Research*, *50*(14), 1304-1311, 2010). In the parallel model, each item is represented by a diffusion to two boundaries (target-present/absent); the search corresponds to a parallel race between these diffusors. The parallel model was highly flexible in that it allowed both for a parametric range of capacity-limitation and for set-size adjustments of identification boundaries. Furthermore, a quit unit allowed for a continuum of search-quitting policies when the target is not found, with "single-item inspection" and exhaustive searches comprising its extremes. The serial model was found to be superior to the parallel model, even before penalizing the parallel model for its increased complexity. We discuss the implications of the results and the need for future studies to resolve the debate.

Keywords Visual search · Attention · Parallel processing · Serial processing · Computational models · Model comparison · RT distributions · Search termination

Electronic supplementary material The online version of this article (doi:10.3758/s13423-015-0978-1) contains supplementary material, which is available to authorized users.

✉ Rani Moran
rani.moran@gmail.com

1 School of Psychological Sciences and Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel

2 General and Experimental Psychology, Ludwig-Maximilians-Universität München, Munich, Germany

3 School of Psychology, Birkbeck College, University of London, London, UK

4 Department of Psychology, Tel Aviv University, Ramat Aviv, POB 39040, Tel Aviv 69978, Israel

Visual search is ubiquitous in daily life, as when we look for a particular object (target) in a crowded scene containing numerous other objects (distractors) and also is central to the investigation of the nature of selective visual attention. Classical theories of selective attention suggest that *two stages* or modes of processing are involved in visual search: 1) a parallel, preattentive, and capacity-unlimited stage in which all visual items are processed to extract a search-guiding "*master*" or "*salience*"[1] map, and 2) a serial, capacity-limited stage during which *focal* attention is allocated *serially* to locations flagged on the salience map to identify selected items (Treisman & Gelade, 1980; Treisman, 1988; Wolfe, 1994, 2007). Henceforth, we will refer to the classical two-stage theory of attention as the "serial model," due to the nature of its attentional component. In contrast, according to *single-stage*, mostly signal-detection-based, *parallel* theories—henceforth referred to as the "parallel model"— attention is distributed diffusively and all items are identified simultaneously (Cameron, Tai, Eckstein & Carrasco, 2004;

[1] Note that salience, or search 'priority', may be based on bottom-up stimulus driven properties (visual gradients) or a combination of bottom-up with top-down properties (matches to a target 'template' held in some search-guiding memory).

Eckstein, Thomas, Palmer & Shimozaki, 2000; Palmer, 1995; Palmer & McLean, 1995; Palmer, Verghese, & Pavel, 2000; Shaw, 1984; Verghese, 2001; Ward & McClelland, 1989). Single-stage parallel models, in turn, are in contention with respect to whether the diffuse-attention mode of processing is capacity-limited (Snodgrass & Townsend, 1980; Thornton & Gilden, 2007; Ward & McClelland, 1989) or unlimited (Palmer & McLean, 1995; Verghese, 2001); we will discuss this important distinction below.

The most important empirical pattern that has been taken to be critical for distinguishing between serial and parallel (1- vs. 2-stage) attentional-allocation theories of visual search is the effect of set size on mean search RT, in particular the slope of the function relating RT to set size. Thus, the classical result of zero slopes in easy search tasks has been interpreted as an indication of parallel, "pop-out" search, whereas positive slopes[2] in more difficult tasks have been interpreted as an indication of a serial search process that relies on focal attention (Treisman & Gelade, 1980; Treisman, 1988). According to the theory of Guided Search (GS), a continuum of search slopes can be obtained by varying the salience of the target among distractors, which is a function of the target-distractor contrast. In GS, target salience controls the probability with which the target item is chosen and identified in each (serial-step) deployment of focal attention: When target salience is very high, the target is invariably selected and identified first, irrespective of the set size, thus accounting for the flat RT/set-size slopes in pop-out tasks. On the other extreme, when target salience is very low, the target is not more salient than any of the distractors and hence item-selection is random. Accordingly, more items need to be searched as set size increases, resulting in steep slopes. Finally, in-between these two extremes, intermediate slopes result from moderate levels of target saliency (Wolfe, 1994, 1998, 2007; Liesefeld et al., 2015).

This classic account, however, has been challenged by supporters of parallel models, who pointed out that under certain assumptions, single-stage parallel search models also can account for the set-size regularities discussed above. In particular, it has been argued that mean RT × set size functions are inadequate to discriminate between serial and parallel search mechanisms (Thornton & Gilden, 2007; Townsend, 1972, 1976, 1990; Palmer, 1995; Palmer & McLean, 1995; Palmer et al., 2000; Verghese, 2001; Ward & McClelland, 1989). That is, shallow search slopes in easy searches and steep slopes in difficult searches can be generated both from serial and parallel mechanisms. For example, a parallel search across all the items in the display can display positive slopes if attentional capacity is limited, so that the amount of processing resources that can be allocated towards each item decreases with set size (Ward & McClelland, 1989; Snodgrass & Townsend, 1980). Moreover, even unlimited-capacity parallel models can account for set-size effects, as a consequence of increased decision criteria to mitigate increases in the influence of decision noise with increasing set size: without such decision-criteria change, the more elements there are to identify, the higher are the chances that one of the distractors will be misidentified as a target (Palmer & McLean, 1995; Verghese, 2001). Recently, Williams, Eidels, and Townsend (2014) have challenged another alleged marker of seriality, namely: bimodality of RT distributions (Cousineau & Shiffrin, 2004), by demonstrating that such distributions could be generated by parallel models with attentional gradients. In sum, patterns of effects on RTs and RT distributions that on first sight appear characteristic for serial models can also be explained by purely parallel models. Accordingly, the serial/parallel controversy is far from being settled. So far, however, no formal quantitative comparison of serial and parallel visual search models with respect to RT-distribution data has been performed.

The purpose of the present paper was to compare a serial-search with a parallel-search model in their ability to account for the full distribution of search RTs (for both target-present and target-absent displays) and error rates, as a function of set size. The serial model exemplar is the Competitive Guided Search model (CGS), which was recently fitted to the data of Wolfe, Palmer, and Horowitz (2010), providing a satisfactory account of the RT distributions, error rates, and their dependence on set size (Moran, Zehetleitner, Müller, & Usher, 2013). The parallel model was developed by us as an extension and integration of proposals made in previous studies (Palmer & McLean, 1995; Thornton & Gilden, 2007; Ward & McClelland, 1989). In particular, it includes a family of models that have flexibility with respect to capacity (which can be limited or non-limited, on a continuum), strategic set-size adjustments of the decision criteria, and the search-termination policy (i.e., how soon does one quit the search when the target has not been found).

We fit RT distributions, following Wolfe et al.'s (2010) demonstration that RT distributions are more informative and constraining with respect to visual-search theories than mean RTs alone (Balota & Yap, 2011; Ratcliff, 1978). The models were fitted to an extensive benchmark data-set (of more than 100K search trials) collected by Wolfe et al., which includes three of the most prevalent tasks in the visual-search literature: a color feature, a color-orientation conjunction, and a spatial-configuration search task. Whereas the spatial configuration (2 vs. 5) and the conjunction tasks produce positive set-size effects and have traditionally been considered to be indicative of a 'serial' architecture, the color-feature task produces flat set-size slopes and has thus customarily been considered to be indicative of a parallel architecture. We start with a brief description of the two models, followed by our computational methods and results. To anticipate, we find that the parallel model is limited in its ability to fit the qualitative data

[2] Typically, the slope is roughly twice as steep for target-absent compared to target-present displays.

patterns from these search tasks and that quantitative formal model comparisons consistently favor CGS. Finally we discuss interpretations and potential follow-up studies.

Computational models

Serial search model: competitive-guided search

Competitive-guided search (CGS) is an instantiation of the Guided Search framework, which, like GS (Wolfe, 1994, 2007), conceives of the search process as a sequence of selection–identification iterations. In each iteration, all visual items compete for selection by the limited-capacity identification process, with weights that are proportional to item salience. Once an item is selected, it is correctly identified (with probability 1), with a Wald-distributed identification latency (reflecting noisy accumulation to a single boundary; Luce, 1986). If the target is selected, search terminates with a "target-present" response. If a distractor has been selected and identified as such, it is inhibited to prevent future reselection of the same item. Additionally, CGS features a quit-unit that competes with the visual items for selection (Fig. 1). The activation of this unit increases over the course of the search with each identified distractor item. When the quit-unit is selected, the search is terminated and a "target-absent" response is given. This allows the model to terminate search in a probabilistic way before all items are searched even when the target is not found, accounting for the large overlap in RT distributions between target-present and target-absent responses (Wolfe et al., 2010). Together with residual time and motor-error parameters, the model features a total of 8 parameters (see Moran et al., 2013, for full details). An attractive property of this model is that only a single set of parameters is needed for all set-size conditions; that is, the number of parameters is independent of the number of set-size conditions.

Parallel search model

We developed a parallel search model as an extension and integration of a number of previous models (Palmer & McLean, 1995; Thornton & Gilden, 2007; Ward & McClelland, 1989). The core assumption of the model is that all items are identified in parallel. For each item in the display, we thus assume a corresponding item identifier that accumulates evidence for and against the hypothesis that the item is a target. One such identifier is illustrated in Fig. 2, modeled as a two-boundary noisy diffusion process, whose upper boundary corresponds to a match (item is the target) and the lower boundary to a mismatch (item is not the target). We assume that all diffusers race in parallel and that they have the same boundary separation a and starting point z. We additionally assume that the target diffuser (if a target is present) has a drift rate v and that all distractor diffusers have the same absolute drift rate but with the opposite sign, $-v$.[3] This two-boundary diffusion is a standard way to extend signal-detection theory to account for RTs and speed-accuracy tradeoffs (Ratcliff, 1978; see also Ratcliff et al., 2007, for a dual-diffusion model based on a race of diffusion processes).[4] The model also includes decision noise, an essential component in the account of set-size effects in parallel search models (Palmer & McLean, 1995).

For a display of set size n, we assume that n such diffusors run independently in parallel. We now describe the search-termination rule. A "target-present" decision is made as soon as one of the diffusors reaches the upper boundary (self-termination on matches). By contrast, 'target-absent' decisions are triggered by a quit unit, whose activation rises as more diffusers reach the lower boundary. Typically, parallel search models postulate that the search is exhaustive when a target is not found (Palmer & McLean, 1995, Ward & McClelland, 1989; Williams et al., 2014). Importantly, our quit-unit-based termination rule reduces to an exhaustive search for certain parameters (see below). However, for other parameters, our termination-rule will quit the search "early," i.e., before full-display inspection. Thus, our termination-rule augments the model with further flexibility, with exhaustiveness comprising a special case, thus offering similar flexibility to that which is present in the CGS model.

To elaborate, we assume that when the k'th item reaches the lower identification boundary (note that k does not index a spatial position but the fact that $k-1$ items have already reached at the distractor boundary before the focal item), the search quits with probability $\left({}^{k}\!/_{n}\right)^{q}$, where $q \geq 0$ is the quit-unit exponent and a free parameter of the model. ${}^{k}\!/_{n}$ is the proportion of display items that have reached the lower boundary. Accordingly, the tendency to quit the search becomes stronger as the proportion of the display items identified as distractors increases (see Donkin & Shiffrin, 2011, for a similar search termination rule). Note that if the n'th item reaches the lower boundary (and assuming the search has not already terminated), the quit unit is triggered with probability 1. For very high quit-unit exponents ($q \to \infty$) the search is

[3] This assumption implies that observers set non-biased drift rate criteria for interpreting target-match vs. mismatch evidence. In principle, observers could bias their drift rate criterion so that distractors and targets generate drift rates that are unequal in their magnitude, implementing a 'dynamic integration bias' (e.g., Moran, 2015). However, here we assume that any bias in identification is fully reflected in the diffusion starting point (see the parameter z below), but otherwise integration proceeds in a non-biased manner.

[4] The notion of a race between diffusion processes captures the intuition that each item-identification is competitive in terms of evidence-accumulation for or against the target, while the different diffusors operate independently of each other (except for the capacity constraint on the drift, which we discuss below).

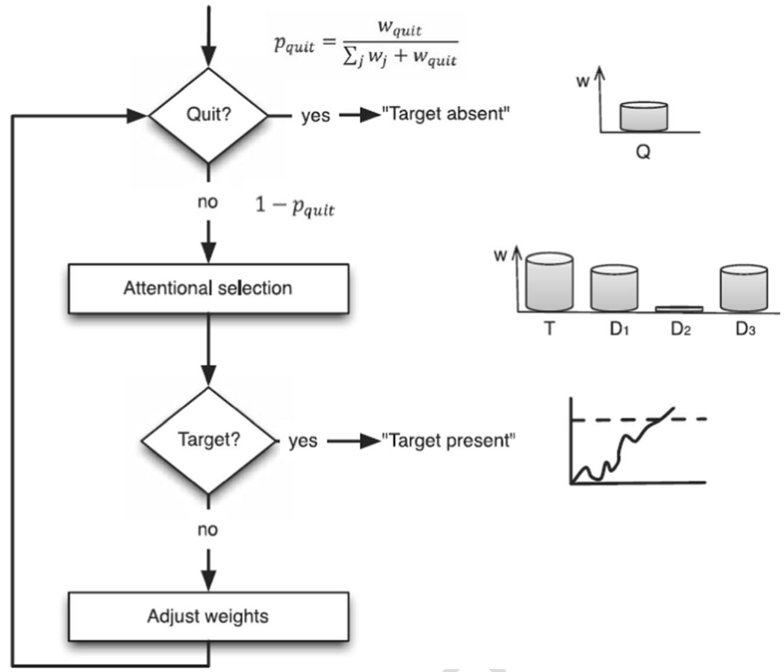

Fig. 1 CGS model (reproduced from Moran et al., 2013). Flow chart depicts the sequence of decisions. When a trial is started, first a "quit-or-continue" decision is made. The probability of quitting is described by the equation for p_{quit}, which is equal to the weight associated with the quit unit relative to the summed weights associated with the quit unit and the display items, w_j. If search is not terminated, an item is selected for inspection. If the target is selected, a "target-present" response is issued. If a nontarget has been selected, the weights are adjusted, that is, w_{quit} is increased and the weight of the just inspected item is set to zero, after which the sequence starts over with the next quit-or-continue decision. Responses are subject to a small proportion of motor errors. The icons to the right of the quitting decision and the attentional selection unit denote the weights for the quit unit as well as the weights of one target, T, and three distractors, D1 through D3. D2 has already been identified as a nontarget and its weight was reset to zero. Also, the quit weight has already been increased. The example illustrates some "target guidance," as the target weight is slightly higher than the distractor weights

exhaustive, because for any $k < n$ the quitting probability is negligible ($\lim_{q\to\infty}\left({}^{k}/_{n}\right)^{q} = 0$). The other extreme is obtained when $q = 0$, where the quit unit is deterministically triggered by the first element to reach the lower bound (i.e., single-item inspection). Intermediate levels of q control the tendency to quit the search earlier or later. This choice of quit unit shares important similarities with the operation of the quit unit in the CGS: In both models, the search-termination probability increases as a function of the number of rejected distractors and decreases as a function of set size.

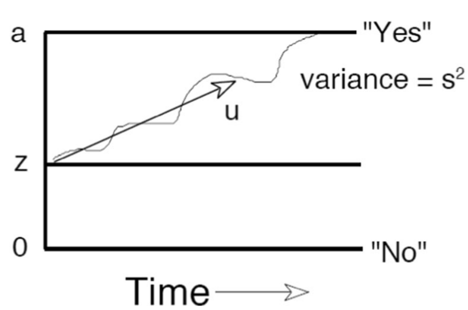

Fig. 2 Noisy target match in the parallel model. Targets have a positive drift v pointing towards the upper (yes) boundary; for nontargets, we assume a symmetric diffusion process (drift -v) towards the negative (no) boundary. Integration is subject to a diffusion noise denoted s

In addition to the search time, the model includes a uniformly distributed 'residual-time' component that captures the time consumed by 'non-search' processes, such as the initial perceptual encoding of the display and the motor production of the response. This choice of a uniform residual time is typical for applications of the diffusion model.[5] In the Appendix, we provide analytical derivations of the RT densities and error rates, for both target-present and target-absent conditions, based on the assumptions described above.

To account for set-size effects on RTs and error rates, parallel models of this type must assume that set size affects either the drift rates and/or the response boundaries. If processing capacity is limited (Ward & McClelland, 1989), then the drift rate should decrease as a function of set size. By contrast, if capacity is unlimited, then the drift rate would be invariant

[5] The mapping of diffusion-model parameters to psychological constructs has been demonstrated behaviorally in perceptual-decision paradigms (Schwarz, 2001; Voss et al., 2004). Additionally, there is electrophysiological evidence (Philiastides et al., 2014; van Vugt et al., 2014) that the non-decision time parameter maps onto neuronal non-decision processing components.

with respect to set size and set-size effects are attributable to strategic changes in decision criteria mitigating the increasing influence of noise. To equip the parallel model with ample flexibility, we allowed the drift rates, starting points, and boundary separations to vary with set size. Specifically, we let the starting point, z, and the boundary separation, a, vary freely as a function of set size (yielding 8 free model parameters for the four values of set size in the experiments). Because in Wolfe et al.'s (2010) experiment, set size is randomized across trials, this assumption entails that (highly experienced) observers are able to rapidly estimate the set size and use this information to adjust decision criteria.[6] To allow a flexible amount of capacity limitation, we assume that drift rates vary with set size (n) as a power function, $v(n) = \frac{v}{n^c}$, whose drift rate "v" and exponent "c" are additional free parameters; note that $c = 0$ corresponds to unlimited capacity and $c = 1/2$ to a signal-detection-based derivation of limited capacity (Ward & McClelland, 1989; Palmer, 1990; Smith & Sewell 2013). As customary, the diffusion noise s was maintained fixed at a constant level $s = 0.1$ (but see Donkin, Brown, & Heathcote, 2009). In addition to the two residual-time parameters, the mean T_{er} and the range s_{er}, the model thus included 13 free parameters. The number of free parameters (n_p) depends on the number of different set sizes k empirically tested in the experiment, as $n_p = 5 + 2 * k$.

Methods

Sketch of the experimental methods of Wolfe et al. (2010) Wolfe et al. (2010) collected data from a total of 28 participants for three classic search tasks: nine participants in a feature search (with target defined by color), 10 in a conjunction search (with target defined by a combination of color and orientation), and nine in a spatial configuration search (with a target digit-2 among distractor digit-5s). In each task, four set sizes (3, 6, 12, and 18 items) were crossed with two trial types (target present vs. absent) to create a factorial design with a total of eight conditions. For each participant, approximately 500 trials were run for each of the eight factorial cells. Both factors were intermixed within experimental blocks, that is, they varied randomly from trial to trial.

Model fitting Our full method for fitting the CGS model has been reported in detail elsewhere (Moran et al., 2013).[7] In fitting the parallel model, we repeated the same steps. Accordingly, our method is only sketched here. In brief, we adopted the Quantile Maximal Probability Estimation (QMPE; Heathcote, Brown, & Mewhort, 2002) procedure to our purpose. To utilize QMPE, each of the eight set-size (4) * target-presence (2) experimental conditions was separated into seven bins: six bins defined by the 0.1, 0.3, 0.5, 0.7, and 0.9 quantiles for correct RTs and one bin for all error trials. Thus, the data from each search task provided 8 (conditions) * (7-1) = 48 free empirical observations. In essence, QMPE consists of Maximum-Likelihood Estimation (MLE) once the precise RT is censored and only bin identity is maintained. We fitted the model separately to the data of each participant as well as to the "average observer" obtained by averaging accuracy rates and correct-RT quantiles across participants. Further details are provided in the appendix.[8]

Results

Spatial-configuration search (2 vs. 5)

The best fits of the two models for the hardest task in the benchmark data of Wolfe et al. (2010) are illustrated in Fig. 3. The figure (as the following similar figures) depicts model predictions based on the fit for the average observer. As can be seen in Table 1, the capacity-limitation parameter for all observers (range 0.21-0.42; mean = 0.33) falls in the range between the unlimited capacity (c = 0) and the signal-detection notion of limited capacity (c = 0.5). Additionally, the quit-unit exponent (range 9.45-42.76, mean = 21.59) indicates that participants tend to search the display deeply but not exhaustively when the target is not found. Figure 4 (left panel) displays the distribution over the number of items that were identified as distractors before the search was terminated on correct-rejection trials.

As evident in the upper panels of Fig. 3, both models were able to account satisfactorily for the slowdown of RT with set size, for the skew in the RT distributions (larger distance between the upper quantile symbols) and for the substantial overlap between the target-present and target-absent RT distributions (Wolfe et al., 2010). In the parallel model, this slowdown is accounted for by both the decrease in drift rate ($c > 0$) and the increase in the boundary separation as functions of set size (Table 1). For hit trials (top left panel), however, there is a tendency for the parallel-model RT distributions to be too wide for the smaller set sizes (3, 6 items) and too narrow for the larger set sizes (12, 18 items). Additionally, the parallel model (red symbols) shows discrepancies in the false-alarm (FA) rates, particularly for set sizes 12 and 18 items (bottom

[6] Adjusting two decision criteria is mathematically equivalent to adjusting the boundary separation and the starting point.

[7] Moran et al. (2013) fitted several sub-models and more constrained CGS model variants: a "no-guidance" model for the 2-vs.-5 task, a 'half set size' variant for the conjunction task, a model with a unique residual time shift parameter for all tasks, or a model where a minimal mean identification time was enforced. Here, we focus on the fits of the non-constrained general 8-free parameter model.

[8] Matlab simulation code for both models is provided in the Supplemental Information.

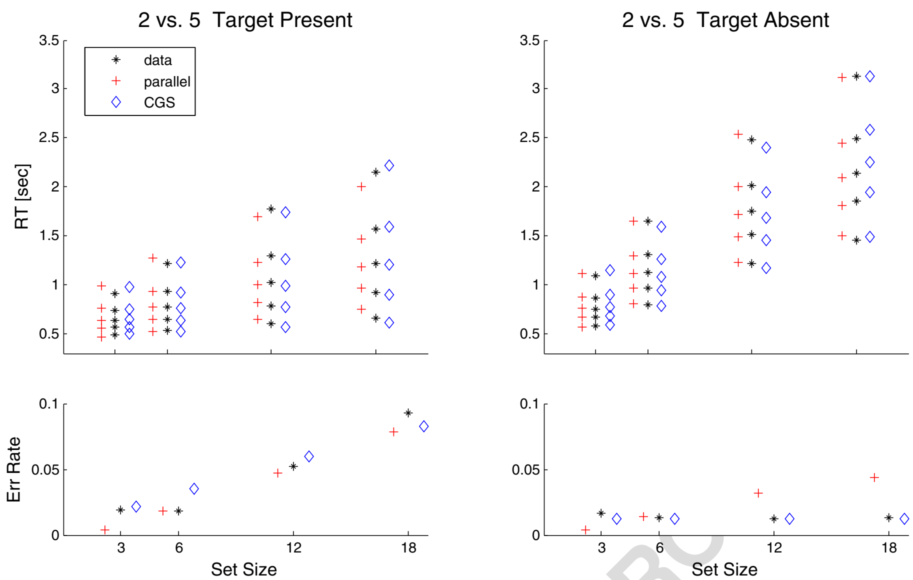

Fig. 3 Data fits of the serial and parallel models to the Wolfe et al. (2010) "2 vs. 5" average-observer data. Empirical data are denoted with black * symbols, the parallel-model predictions with red + symbols, and the CGS predictions with blue diamonds. Right and left panels correspond to target-present and target-absent trials, respectively. Upper and lower panels correspond to quantile correct RTs and error rates, respectively

right panel). Indeed, the model predicts an increase in the FA rate as a function of set size, whereas the empirical FA rate is constant. The reason for the predicted increase in FA with set size is that as set size increases, so does the probability that one of the distractors will mistakenly hit the upper "target" diffusion boundary in a target-absent display. Notably, this tendency is mitigated by the search being non-exhaustive, so that the effective set size that is searched when a target is not found is smaller than the nominal set size. Furthermore, the set-size-related increase in boundary separation acts to reduce FAs. Still, these influences are overruled by the decrease in drift rate, which acts to increase FAs. Regarding miss rates, the parallel and serial models seem to be "on par" (bottom left panel).

To compare the goodness of fit for both models, we calculated (Table 2) the difference between the parallel and CGS models with respect to deviance (i.e., minus twice the log likelihood of the data, under the QMPE parameters), AIC (Akaike, 1973), and BIC (Schwarz 1978). Strikingly, even without the additional penalty imposed by information criteria (AIC, BIC) for the five extra parameters of the parallel model, the CGS model fits the data better for seven individual participants (i.e., all except participants 4 and 7) as well as for the group as a whole, as evidenced by the positive Δ Dev values.

Table 1 Best fitting parameters for the parallel model in the 2-vs.-5-search task

Participant	v	a_3	a_6	a_{12}	a_{18}	z_3	z_6	z_{12}	z_{18}	T_{er}	s_{er}	c	q
1	0.40	0.20	0.24	0.28	0.31	0.09	0.11	0.13	0.14	0.23	0.00	0.30	21.68
2	0.46	0.22	0.22	0.28	0.33	0.10	0.10	0.14	0.17	0.24	0.03	0.29	23.29
3	0.47	0.30	0.33	0.39	0.43	0.15	0.17	0.20	0.21	0.35	0.00	0.42	14.80
4	0.30	0.26	0.29	0.31	0.33	0.12	0.13	0.13	0.12	0.25	0.00	0.21	19.66
5	0.44	0.23	0.25	0.30	0.31	0.11	0.11	0.13	0.13	0.28	0.29	0.36	9.45
6	0.47	0.17	0.17	0.20	0.22	0.08	0.06	0.07	0.07	0.24	0.00	0.31	42.76
7	0.46	0.29	0.32	0.39	0.42	0.13	0.16	0.20	0.21	0.31	0.00	0.39	10.70
8	0.34	0.17	0.21	0.26	0.29	0.07	0.09	0.11	0.11	0.22	0.14	0.30	19.66
9	0.52	0.32	0.31	0.33	0.36	0.15	0.14	0.16	0.17	0.18	0.00	0.39	32.28
Avg. Obs.	0.39	0.22	0.24	0.30	0.33	0.10	0.11	0.13	0.15	0.27	0.01	0.31	16.81
Mean	0.43	0.24	0.26	0.30	0.33	0.11	0.12	0.14	0.15	0.26	0.05	0.33	21.59

Avg. Obs. row presents fits to averaged (RT-quantile and accuracy) data, whereas the "mean" row presents the parameters averaged across individual participants. Subscripts refer to set size

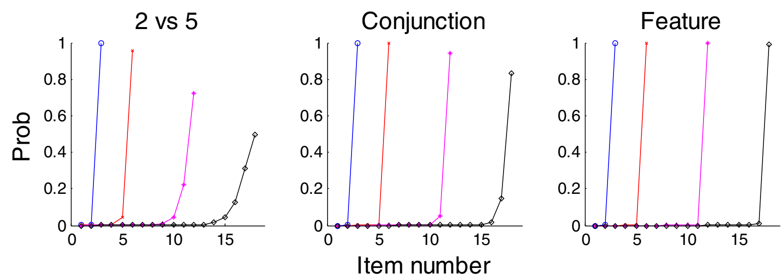

Fig. 4 Distributions of the number of identified distractors before quit-unit triggering on correct rejection trials for the different tasks of Wolfe et al. (2010), based on the best-fitting parameters for the average observers. The different colors indicate the different set sizes

Penalizing the models for complexity, AIC still prefers the parallel model for Participants 4 and 7. According to BIC, CGS is inferior only for participant 4, whereas for participant 7 the models are tied.

Conjunction search

For the conjunction-search task, too, the model fits provide strong support for the CGS model (see Table 3 for the best-fitting parameters). As shown in Fig. 5, the parallel model provides a good fit for the target-present RTs (top left panel). However, the model fails with respect to target-absent displays: it underestimates the inter-quantile range of RTs for the large set size (12, 18 items; top right panel) and falsely predicts an increasing FA rate with increasing set size (bottom right panel). Additionally, CGS accounts better for the miss rates (bottom left panel). A model comparison (Table 2) showed that for all participants (except for Participant 1) as well as for the group as a whole, CGS yielded lower deviance values despite its lower number of parameters. Even for Participant 1, after adding the penalty term, CGS was preferred according to AIC (let alone, BIC).

Feature search

For the feature task, too, the model fits provide strong support for the CGS model (see Table 4 for the best-fitting parameters). To understand the reasons for this, we focus below on the fits of the parallel model. Considering first the target-present displays, we find that the parallel model provides a good fit for the hit RTs (Fig. 6, top left panel). Remarkably, as in the data, there are no observable set size effects on the predicted hit RTs. Table 4 shows that with increasing set size, the threshold separation hardly changes, while the starting point moves closer to the lower target-absent boundary. With everything else being equal, this effect would lead to an

Table 2 Model comparison measures for the parallel vs. the CGS model for the different tasks

Task	2 vs. 5			Conjunction			Feature		
Participant	ΔDev	ΔAIC	ΔBIC	ΔDev	ΔAIC	ΔBIC	ΔDev	ΔAIC	ΔBIC
1	98.2	108.2	139.6	−7.1	2.9	34.4	216.9	226.9	258.4
2	172.7	182.7	214.2	3.8	13.8	45.2	33.1	43.1	74.5
3	21.7	31.7	63.0	473.3	483.3	514.8	42.2	52.2	83.7
4	−25.7	−15.7	15.7	113.5	123.5	154.9	77.7	87.7	119.2
5	168.2	178.2	209.6	132.3	142.3	173.7	150.3	160.3	191.7
6	307.1	317.1	348.6	396.7	406.7	438.2	−6.1	3.9	35.3
7	−41.4	−31.4	0.0	21.5	31.5	62.9	66.4	76.4	107.8
8	222.6	232.6	264.1	212.3	222.3	253.7	121.0	131.0	162.4
9	34.3	44.3	75.8	13.3	23.3	54.8	108.9	118.9	150.3
10				227.1	237.1	268.6			
Avg. Obs.	971.6	981.6	1024.0	1732.7	1742.7	1785.7	704.5	714.5	757.0
Group	957.5	1047.5	1429.4	1586.6	1686.6	2116.4	810.2	900.2	1282.2

The Δ (Dev, AIC, BIC) is calculated by subtracting the respective values for CGS from the parallel-model values. For the average observer ("Avg. Obs."), the number of observations was taken to be the total number of observations summed across participants. For the "Group" row, the entire set of fits for the individual participants was considered as a single "group fit" for the entire data. The likelihood of this group fit was the product of the likelihoods across individual observers. Additionally, the number of parameters and observations for the group fit were obtained by summing the number of parameters and observations, respectively, across participants. Dev, deviance

Table 3 Best fitting parameters for the parallel model in the conjunction-search task

Participant	v	a_3	a_6	a_{12}	a_{18}	z_3	z_6	z_{12}	z_{18}	T_{er}	s_{er}	c	q
1	0.44	0.15	0.14	0.14	0.15	0.07	0.06	0.04	0.04	0.30	0.07	0.08	118.79
2	0.52	0.13	0.13	0.14	0.14	0.05	0.04	0.04	0.04	0.34	0.00	0.12	62.34
3	0.66	0.30	0.24	0.21	0.23	0.13	0.09	0.07	0.07	0.19	0.00	0.41	86.97
4	0.55	0.08	0.08	0.09	0.10	0.04	0.03	0.04	0.04	0.26	0.01	0.08	20.86
5	0.39	0.11	0.12	0.13	0.14	0.06	0.05	0.04	0.05	0.29	0.00	0.08	26.43
6	0.49	0.16	0.15	0.17	0.20	0.07	0.06	0.06	0.07	0.22	0.00	0.31	23.89
7	0.62	0.20	0.18	0.17	0.18	0.09	0.07	0.06	0.06	0.33	0.00	0.24	115.31
8	0.33	0.16	0.15	0.17	0.19	0.08	0.06	0.07	0.08	0.27	0.00	0.08	20.05
9	0.48	0.15	0.14	0.15	0.16	0.06	0.05	0.05	0.04	0.31	0.07	0.12	174.80
10	0.63	0.22	0.18	0.20	0.23	0.10	0.07	0.08	0.10	0.26	0.00	0.34	46.38
Avg. Obs.	0.47	0.15	0.14	0.15	0.17	0.07	0.05	0.05	0.06	0.29	0.00	0.18	33.60
Mean	0.51	0.17	0.15	0.16	0.17	0.07	0.06	0.06	0.06	0.28	0.02	0.19	69.58

"Avg. Obs." row presents the parameters for the average observer (data averaged across observers) and the "mean" row displays the parameters averaged across individual participants

increase in hit RTs since the target has to traverse a longer distance to reach the upper boundary. However, this effect is offset by a weak tendency for "super-capacity," that is: a negative capacity exponent, which results in set size increasing drift rates. This weak super-capacity could arise if the larger number of display items increased the target's (bottom-up) salience.[9] Note, however, that this drift/starting-point trade-off is less successful in accounting for the miss rates: unlike the data, the parallel model predicts a large increase in miss rates as a function of set size (bottom left panel). Why does the starting point move downwards, however?

To understand this, we need to consider the target-absent condition. As shown in Fig. 6, in this condition, unlike in the data, the model predicts a speed-up in correct rejections (top right panel) in the form of shrinkage of the upper part of the RT range. Figure 4 shows that when the target is not found, the search is exhaustive. Thus, all other things being equal (including boundaries and starting point), RTs for CRs would increase with set size (it takes longer for more distractors to reach the lower boundary). However, this "exhaustiveness effect" is offset by the set-size-dependent decrease of the starting point and the super-capacity. These "opposing effects" balance each other almost perfectly with respect to the three lowest RT quantiles (including the median), and maintain a satisfactorily low stable rate of FAs (bottom right panel). However, unlike the data, the model predicts a speed-up in the two upmost (0.7, 0.9) CR quantiles. This intricate trade-off provides further demonstration for why stronger model constraints can be gleaned by fitting search models to RT distributions, rather than only to central-tendency measures (Wolfe et al., 2000).

Finally, a model comparison (Table 2) showed that for all participants (except for Participant 6) as well as for the group as a whole, CGS yielded lower deviance values despite its lower number of parameters. Even for Participant 6, CGS was preferred according to AIC (let alone BIC). This finding is striking, taking into account that for a long time, feature search has been considered the prototypical task for a parallel search architecture.

The parallel-model fits to the feature task that we presented above correspond to a highly flexible model, which assumes that boundaries, starting points, and drift rates can vary with set size and which also included the capacity and the quit-termination parameters. Interestingly, the inclusion of the latter did not help the parallel model in this case because the fit always converged to large q-values that correspond to exhaustive search (Table 4; Fig. 4, rightmost panel). To better understand the reason for this intruding behavior, we explored a more constrained model variant, which was obtained by setting a moderate upper bound on the quit parameter ($q \leq 5$) that prevented a fully exhaustive search. As expected, the fits were worse than for the flexible model that we presented in Fig. 5. Notably, this model was able to account for the traditional property of flat mean-RT with set size, but not for the full RT distribution and the error rate functions (see Supplemental information).

[9] This is plausible for the present data set, because displays with more items were also more densely packed. Importantly, higher item density with featurally homogeneous displays entails more "iso-feature suppression" (i.e., suppression of the activation of a detector tuned to a particular feature within its receptive field by the presence of objects possessing similar features picked up by detectors in neighboring fields; e.g., Li, 1999), rendering the distractors less salient—which means that the target becomes relatively more salient. In line with such an increase in iso-feature suppression, several studies actually reported a decrease (instead of the more typical increase!) of RTs with the number of distractors in singleton feature ('pop-out') search, as in the current task (Bravo & Nakayama, 1992; Rangelov, Müller, & Zehetleitner, 2013).

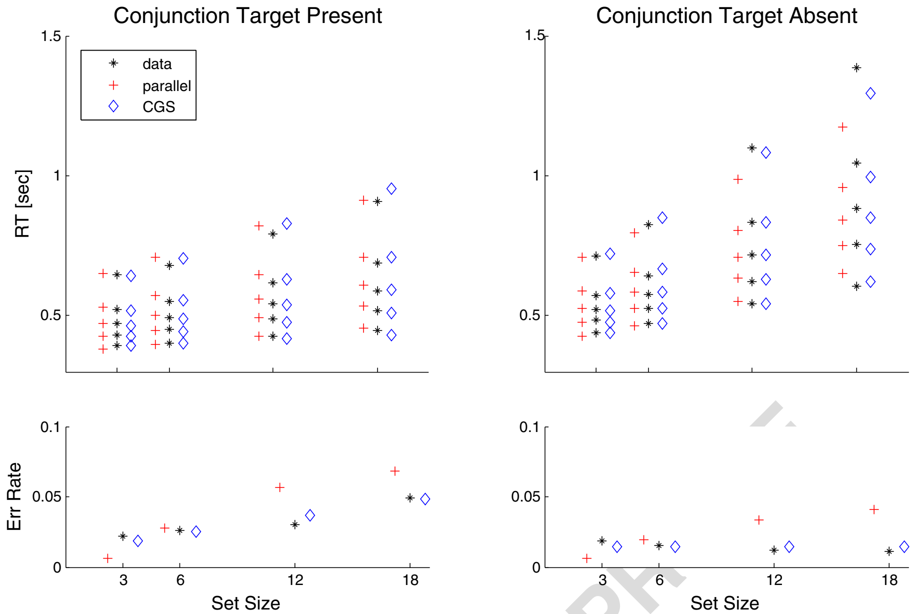

Fig. 5 Model fits for the conjunction-search task of Wolfe et al. (2010) to average observer data. The arrangement of the figure is identical to Fig. 3

Discussion

Despite the remarkable support for the two-stage Guided Search model in accounting for visual search data (Wolfe, 1994, 2007), it has been suggested that the typical set-size effects on mean RT (positive slopes) also are consistent with a number of parallel search models (Palmer & McLean, 1995; Thornton & Gilden, 2007; Verghese, 2001; Ward & McClelland, 1989). Such models could, in principle, account for the positive slopes as a result of either a reduced rate of item processing due to limited capacity (Snodgrass & Townsend, 1980; Shaw, 1984) or an increase in the decision boundary necessitated to maintain error rates at reasonable levels; without such a boundary change, the FA-rate would dramatically increase with set size (Palmer & McLean, 1995). The purpose of our investigation was to develop, based on an extension and integration of prior suggestions (Palmer & McLean, 1995; Thornton & Gilden, 2007; Ward & McClelland, 1989), such a parallel model that combines both capacity limitations and flexible decision-boundary settings and to assess how well it accounts for visual search data compared with the serial CGS model, which has recently been shown to account well for RT-distribution data (Moran et al., 2013). To endow the parallel model with ample flexibility, we even introduced a "quit unit" that allows for pre-exhaustive-search termination when the target is not found. We focused

t4.1 **Table 4** Best-fitting parameters for the parallel model in the feature-search task

	Participant	v	a_3	a_6	a_{12}	a_{18}	z_3	z_6	z_{12}	z_{18}	T_{er}	s_{er}	c	q
t4.2														
t4.3	1	0.32	0.11	0.11	0.11	0.11	0.05	0.04	0.03	0.02	0.29	0.07	-0.13	200.1
t4.4	2	0.34	0.11	0.10	0.10	0.10	0.05	0.04	0.03	0.03	0.22	0.08	-0.12	178.4
t4.5	3	0.72	0.08	0.07	0.07	0.07	0.04	0.03	0.03	0.02	0.25	0.08	-0.05	44.2
t4.6	4	0.38	0.18	0.17	0.17	0.17	0.08	0.06	0.05	0.04	0.21	0.00	-0.05	204.6
t4.7	5	0.47	0.12	0.11	0.11	0.11	0.05	0.04	0.03	0.02	0.26	0.07	-0.05	170.1
t4.8	6	0.53	0.10	0.09	0.09	0.09	0.05	0.04	0.03	0.02	0.24	0.06	-0.06	111.5
t4.9	7	0.34	0.11	0.11	0.11	0.11	0.04	0.03	0.03	0.02	0.21	0.00	-0.13	159.5
t4.10	8	0.35	0.14	0.13	0.12	0.12	0.06	0.05	0.03	0.03	0.23	0.01	-0.10	159.2
t4.11	9	0.36	0.10	0.10	0.10	0.11	0.04	0.03	0.03	0.02	0.26	0.13	-0.12	163.9
t4.12	Avg. Obs.	0.39	0.12	0.11	0.11	0.11	0.05	0.04	0.03	0.03	0.24	0.07	-0.09	78.8
t4.13	Mean	0.42	0.12	0.11	0.11	0.11	0.05	0.04	0.03	0.03	0.24	0.06	-0.09	154.6

"Avg. Obs." row presents the parameters for the average observer (data averaged across observers) and the "mean" row displays the parameters averaged across individual participants

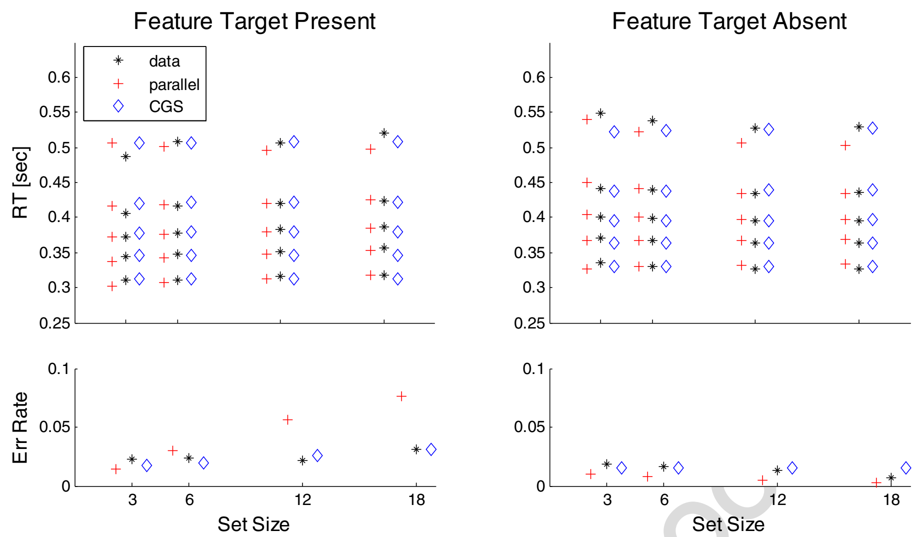

Fig. 6 Model fits for the feature task of Wolfe et al.'s (2010) to average observer data. Arrangement of the figure is identical to that of Fig. 3

on three classical search tasks from a rich data set (Wolfe et al., 2010) that provided reliable estimators of the full RT distributions for individual observers. Importantly, both the spatial configuration and the conjunction tasks exhibit robust set-size effects, thus allowing for probing the origin(s) of those effects. Methodologically, we embraced Wolfe et al.'s (2010) call for fitting the models to RT distributions, rather than simply to RT means, as RT distributions provide enhanced constraints on the nature of the generating search mechanism(s).

Consider first the more difficult ("serial") search tasks. The results showed that the fits of the parallel model were problematic. In the 2-vs.-5 task, the model erroneously predicted a set-size-dependent increase in FA rate and failed in accounting for the set-size-related range expansion in hit RTs. For the conjunction task, the parallel model failed to account for performance on target-absent displays with respect to both RT distributions and error rates. Formal model-comparison procedures using AIC and BIC consistently favored CGS for almost all participants and for the group as a whole (Table 2). Importantly, the superiority of CGS was not a consequence of "over-parameterizing" the parallel model and hence subjecting it to heavier AIC/BIC penalties. Indeed, despite its larger number of free parameters (13 vs. 8 for CGS), the parallel model performed worse based on a goodness-of-fit deviance measure, which does not apply number-of-parameters-related penalties. This finding is striking, especially when taking into account that CGS provided adequate fits with parameters that were invariant with respect to set size, whereas the parallel model allowed for flexible set-size adjustments in boundary separation and identification bias. Furthermore, by introducing a capacity parameter "c," the parallel model was equipped with the ability to behave in a capacity-limited (e.g., Ward & McClelland, 1989) as well as in a capacity-unlimited (Palmer & McLean, 1995; Verghese, 2001) and even in a super-capacity manner. Thus, our model comparisons show that the serial, two-stage CGS model (Moran et al., 2013) performs better than a family of parallel models that vary along the degree of capacity limitation.

Having compared the models with respect to these traditional serial search tasks, we next compared the models based on their fits to the feature task. Given that this task has traditionally been considered to epitomize a parallel search architecture, it provides a stringent test for the serial model. Strikingly, we found consistent superiority for CGS (Table 2), especially in its ability to provide a better account for miss rates and correct-rejection RTs.

Differences between the serial and the parallel model with respect to the feature task

As explained by Moran et al. (2013), CGS accounts for the feature-task data by assigning very high weights to both target saliency and to the quit-unit boost. This has two desirable consequences with respect to the model's capability in fitting the data. On target-present trials, for any set size the target is almost certainly identified as the first item. On target-absent trials, for any set size, the quit unit is almost always selected after the rejection of the first distractor. In other words, if the target failed to pop out, observers safely terminate the search, deciding that a target is absent. Thus, in both target-present and target-absent displays, a single item is identified; consequently, there are no set-size effects. The rightmost panel in Fig. 4, however, shows that according to the parallel-model account, feature search was exhaustive when the target was not found. This finding appears puzzling at first glance, because by setting the quit-unit exponent to 0 ($q = 0$), the model can also quit after a single item is inspected. Why, then, can it not successfully mimic the serial model?

Interestingly, this question highlights a fundamental distinction between the serial and parallel models compared. Whereas in our CGS parameterization, identification time was invariant with respect to set size, this was not the case with the parallel model. Imagine we would maintain all the parameters of the parallel model invariant across set size and set a zero quit-unit exponent so that it quits after the first distractor reaches the lower boundary. Even in this case, the parallel model would not mimic the serial model because the event of “identifying the first item” in the parallel model is an event that is sensitive to statistical facilitation: With more display items, ceteris paribus, the first item would be identified faster, producing negative RT slopes (this is obvious for target-absent responses, but also would occur with target-present responses, as the target diffusor needs to be faster than all of the distractor diffusors in order for the response to amount to a “hit”).

Notably, everything else is not necessarily equal as the parallel model was endowed with ample flexibility to apply set size modulations with respect to threshold separation, starting point, and capacity. The results of our quantitative model fits show, however, that the empirical RT distributions and error rates provide strict constraints such that a policy of exhaustive search (when the target is not found) yielded the best fits. Ceteris paribus, search-exhaustiveness induces a plethora of set-size effects: a slowdown in CRs (due to the 'need to wait' for the last distractor to reach the lower boundary), a speedup in hits (due to statistical facilitation; note that a 'hit' can be triggered by a distractor, rather than the target, mistakenly reaching the upper bound) an increase in FA-rate (higher likelihood that one of the distractors will mistakenly reach the upper boundary in target absent displays) and a reduction in miss rates (once more, due to the higher probability that one of the distractors will mistakenly reach the upper boundary in a target present display and will trigger a correct hit response). In its best fits, the parallel model tried compensating for these effects with set size reductions of starting points and with increasing drift (“super-capacity”). These fits, alas, were inferior to those produced by CGS, because they failed to provide a satisfactory tradeoff in accounting for miss rates and they generated an unobserved set-size related speedup in the high quantiles of the CR distributions. As shown in the Supplement, more constrained fits (which impose a limit on the quit parameter) failed to improve the model fits.

Qualifications and future directions

While our results favor the two-stage serial CGS model, they need to be taken with caution with regard to concluding an unequivocal superiority for a serial over a parallel architecture of attentional selection. First, extensions of our parallel model need to be explored. For example, within the framework of parallel-diffusor models, it would be important to probe the possibility that different items are processed with different drift rates due to attentional gradients. Such gradients (Cheal, Lyon & Gottlob, 1994; Downing, 1988; LaBerge & Brown, 1989; Müller & Humphreys, 1991) may play an important role in parallel models because, as recently shown by Williams et al. (2014), they allow parallel models to produce mixture-RT distributions, an important characteristic of serial search models (e.g., Moran et al., 2013). Because such investigations will depend on a number of critical assumptions (e.g., the magnitude of attentional gradients, the within-trial dynamics of such gradients, etc.), they will require a dedicated investigation. Additionally, in the current model, we adopted the simplifying assumption that people set nonbiased drift rate criteria for the item-identification process and that any identification-biases are reflected in the starting point (see also Footnote 3). Consequently, the identification drift-rates for the target and the distractors are equal in magnitude. This assumption, however, could be relaxed in future studies to allow for different target-distractors drifts. Future investigations may also explore alternative termination rules for target-absent responses. While our quit unit constitutes one approach for implementing an “urgency signal” (the tendency to quit the search increases as more distractors are rejected), alternative mechanisms could be explored, for example, by collapsing decision boundaries (Drugowitsch et al., 2012; Moran, 2015; Thura et al., 2012; but see Hawkins et al., 2015; Moran, Teodorescu, & Usher, 2015).

It should be noted that our model comparison study is parametric in that it makes specific distributional assumptions with respect to the components of the model (e.g., item-identification and residual times). In this respect, our approach is modest in its ambition as compared with non-parametric, model-free attempts to identify the visual-search cognitive architecture. Alas, prior model-free attempts have produced inconclusive conclusions, because they highlighted the possibility for serial-parallel mimicry (for a recent review, see Algom et al. 2014). Still, one limitation of our study is that it cannot rule out the possibility that different distributional assumptions in future serial and parallel models will improve visual search models and that such future parallel models will outperform future serial models. This, however, does not imply that our current findings are trivial. On the contrary, we contend that the advantage of our approach is that—as a consequence of making parametric assumptions—it avoids the risk of model mimicry. Furthermore, our parametric assumptions are well motivated: By grounding our parallel model on a diffusion-type architecture—the currently most popular approach for modeling speeded decisions across a wide range of cognitive tasks—we believe that our findings are highly informative in the context of current research. Finally, these results provide a challenge that more sophisticated parallel models will need to rise to if they wish to compete with Guided-Search type serial models in accounting for visual search data.

An alternative approach to testing the adequacy of the parallel model of visual search that may avoid the pitfall of specific assumptions associated with the model we explored here (e.g., termination rule on target-absent trials) could rely on the development of ideal-observer-inspired models (Palmer, Verghese & Pavel, 2000). One such promising signal-detection model was developed by Verghese (2001) to account for accuracy with brief search displays. Unlike the current parallel model, in the Verghese model, the individual items are not separately identified. Rather, search decisions are based on a global match between the search display and a target template. This global match in turn is based on the maximal value of the local matches between each display item and the target. Decisions are based on comparing the global match with a "signal-detection" criterion (see also Cameron et al., 2004; Eckstein et al., 2000). While this model was shown to account for set-size effects on accuracy, it has not yet been formally extended and tested on its ability to account of RT distributions.[10]

Furthermore, conclusions (favoring the serial model) may need to be qualified to visual search displays that are available until response. In a set of studies, Palmer and colleagues (2000) showed that in a paradigm in which the display was presented very briefly and the dependent variable was accuracy (rather than RT), a parallel signal-detection model with unlimited capacity provided a better account of the data than either a parallel model with limited capacity or a serial model (Palmer, 1994; Palmer, Ames, & Lindsey, 1993; Dosher, Han, & Lu, 2004, 2010). It thus is possible that the strategy that observers rely on in visual search varies with task contingencies: While for briefly presented displays observers may rely on the maximal value of saliency, with time-unlimited and difficult search displays, they may use the salience map to engage in serial attentional selections that guide a high-resolution identification process to verify target presence. With more difficult displays still, observer also may need to use eye movements to explicitly search through the display (Bloomfield, 1979; Zelinsky & Sheinberg, 1997).

To better understand the nature of the operating processes in visual search, future studies comparing serial and parallel models are required. Such studies should examine additional data-sets based on experimental manipulations that are designed to differentiate between these types of models. For example, it would be important to test how these types of models account for visual-search performance in displays in which target salience is manipulated on a continuum (Liesefeld et al., 2015) or in which target prevalence is manipulated (Wolfe & Van Wert, 2010). Furthermore, the understanding of the nature of attentional processes in visual search will have to include efficiency considerations. For example, once attentional gradients are assumed (Williams et al; 2014), a two-stage serial model, such as Guided Search, which shifts its high-resolution attentional resources across the display, may just be the best way to use the visual system to optimize search performance under its constraints.

Finally, while serial and parallel theories describe two prototypical search mechanisms, future research also should consider the possibility of hybrid mechanisms. For example, an attentional spotlight (Eriksen & Yeh, 1985; LaBerge and Brown 1989; Posner & Petersen, 1990) might be deployed serially *between* spatial locations in the search display, while processing items simultaneously (i.e., in parallel) *within* locations. Another possibility is that the search mechanism is analogous to a "car washing" pipeline, wherein several items are identified in parallel and the identification of another item can begin only after the identification of an 'engaged' item completes (Wolfe, 2007). Exploring such possibilities in future formal models of visual search, and evaluating these models based on RT distributional data, may yield "middle-ground theories" with respect to the serial-parallel search debate.

Acknowledgments This work was supported by grant 158/2011 from the German-Israeli Foundation (M. Z., M. U., & H. J. M.).

Appendix: fitting the parallel model

As explained in the main text, the parallel model was fitted using QMPE. To utilize QMPE one has to calculate, for a given ensemble of model parameters and for each experimental condition, predictions with respect to the proportion of trials terminating in each of the seven bins (i.e., RT quantiles 0.1-0.9 and errors). These calculations, for the parallel model, were based on first calculating probabilities of events for a single diffuser, and then deriving predictions for the ensemble of n parallel diffusers based on the assumption of stochastic independence across the different diffusers. We describe each of these steps in turn.

To calculate the probability that a single diffuser hits its upper boundary by time t, we used the infinite sum formulas for the cumulative density function (henceforth CDF) of the diffusion process (Busemeyer & Diederich, 2010; Cox & Miller, 1965; Feller, 1968; Luce, 1986; Ratcliff, 1978; Smith, 1990, 2000). In practice, we calculated the first 1000 terms in the sum.[11] The resulting CDF is of the form $F(t; a, v, z, up/down)$: the probability that a diffuser starting at z, with

[10] We have performed preliminary explorations of a sequential-sampling extension of this model, which yielded a lower poorer fits as compared with the CGS model (Moran et al., 2013). One challenge that the test of this model involves is that, unlike the one presented, it does not allow analytical calculations and thus requires more laborious, slow and noisy, model simulations.

[11] In fitting the feature task we sometimes encountered numerical problem (e.g., when drift rates get very high). Thus, we also fit the average observer of this task with an alternative methods were the (single diffuser) CDF was estimated based on a simulation of 100K diffusion trials, obtaining similar results (i.e., superiority of CGS over the parallel model

drift v, has reached the upper (or lower) threshold with threshold separation a by time t. These CDFs were used to calculate target and distractor events as follows. The probabilities that the target diffusor has reached the upper or lower threshold by time t are given by $F_{T,up} = F(t; a, v, z, up)$ and $F_{T,down} = F(t; a, v, z, down)$, respectively. The probabilities that a distractor diffusor has reached the upper or lower threshold by time t is given by $F_{D,up} = F(t; a, v, -z, down)$ and $F_{D,down} = F(t; a, v, -z, up)$, respectively. The density probability functions (henceforth pdf) $f_{T,up}, f_{T,down}, f_{D,up}, f_{D,down}$ were calculated based on numerical derivations of the corresponding CDFs.

Based on these distributions, we calculated the probabilities for Hits and Correct Rejections (CRs) with $t_s \leq t$, as described next. Note that t_s denotes the "search time," which is different from the full RT that includes an additional residual component. Miss and False Alarm (FA) rates were calculated as the complements for Hits and CRs, respectively, with $t_s = \infty$.[12]

Hits The pdf was calculated as follows. Hits are composed of three disjoint events, each described in turn. The first event is that the target diffusor is the first to reach the upper boundary. In this case any number of $0 \leq k \leq n-1$ distractors could have already reached the lower boundary, failing to trigger the quit unit. Hence this event contributes the term:

$$\sum_{k=0}^{n-1}\left[\binom{n-1}{k} f_{T,up}(t_s) F_{D,down}{}^{k}(t_s)\left(1-F_{D,down}(t_s)-F_{D,up}(t_s)\right)^{n-k-1}\prod_{l=1}^{k}\left(1-\left({}^{l}/_{n}\right)^{q}\right)\right]$$

Q4

The second event is that any of the $n-1$ distractor diffusors is the first element to reach the upper boundary, while the target has already reached the lower boundary in addition to $0 \leq k \leq n-2$ distractors. This event contributes the term:

$$(n-1)\sum_{k=0}^{n-2}\left[\binom{n-2}{k} f_{D,up}(t_s) F_{T,down}(t_s) F_{D,down}{}^{k}(t_s)\left(1-F_{D,down}(t_s)-F_{D,up}(t_s)\right)^{n-k-2}\prod_{l=1}^{k+1}\left(1-\left({}^{l}/_{n}\right)^{q}\right)\right]$$

The third event is that any of the $n-1$ distractor diffusors is the first to reach the upper boundary, while $0 \leq k \leq n-2$ distractors have already reached the lower boundary and the target is still diffusing between boundaries. This event contributes the term:

$$(n-1)\sum_{k=0}^{n-2}\left[\binom{n-2}{k} f_{D,up}(t_s)\left(1-F_{T,down}(t_s)-F_{T,up}(t_s)\right) F_{D,down}{}^{k}(t_s)\left(1-F_{D,down}(t_s)-F_{D,up}(t_s)\right)^{n-k-2}\prod_{l=1}^{k}\left(1-\left({}^{l}/_{n}\right)^{q}\right)\right]$$

In sum,

$$\begin{aligned} f_{Hit}(t_s) = &\sum_{k=0}^{n-1}\left[\binom{n-1}{k} f_{T,up}(t_s) F_{D,down}{}^{k}(t_s)\left(1-F_{D,down}(t_s)-F_{D,up}(t_s)\right)^{n-k-1}\prod_{l=1}^{k}\left(1-\left({}^{l}/_{n}\right)^{q}\right)\right] \\ &+(n-1)\sum_{k=0}^{n-2}\left\{\binom{n-2}{k} f_{D,up}(t_s) F_{D,down}{}^{k}(t_s)\left(1-F_{D,down}(t_s)-F_{D,up}(t_s)\right)^{n-k-2}\prod_{l=1}^{k}\left(1-\left({}^{l}/_{n}\right)^{q}\right)\left[F_{T,down}(t_s)\left(1-\left(\frac{k+1}{n}\right)^{q}\right)+1-F_{T,down}(t_s)-F_{T,up}(t_s)\right]\right\} \end{aligned}$$

The CDF $F_{Hit}(t_s)$ was found by numerically integrating this term with a resolution of 0.1 ms. $F_{Hit}(\infty)$ was found by integrating up to a large t_s after which the result hardly changed (typically, 5 sec, 3 sec and 1.5 sec in the spatial configuration, conjunction and the feature task, respectively).

Correct rejections The pdf for a CR was calculated as follows. The quit could be triggered by the k'th distractor, which

[12] After fitting our models, we verified that our analytical derivations (described below) yield predictions (in terms of RT quantiles and error rates) that are very similar to these produced by a mechanistic trial-by-trial simulation of the model based on the best fitting parameters.

means that the prior $k-1$ distractors (which have already reached the lower bound) failed to trigger it. Thus,

$$f_{CR}(t_s) = n\sum_{k=1}^{n}\left[\binom{n-1}{k-1} f_{D,down}(t_s)\,F_{D,down}{}^{k-1}(t_s)\left(1-F_{D,down}(t_s)-F_{D,up}(t_s)\right)^{n-k}\prod_{l=1}^{k-1}\left(1-{}^{l}/_{n}\right)^{q}\left({}^{k}/_{n}\right)^{q}\right]$$

As before, the CDF $F_{CR}(t_s)$ and $F_{CR}(\infty)$ were found by numerical integration.

Incorporating residual time In order to apply QMPE, we need to calculate the proportions of Hits and CRs in a given temporal bin $[x, y]$, taking into account that the reaction time (RT) is the sum of the search time (t_s) and a residual component. This was achieved as follows: The residual time interval $[T_{er}-S_{er}/2,\ T_{er}+S_{er}/2]$ was represented with the 30 equally distant points $t_i = T_{er} - \frac{S_{er}}{2} + \frac{(2i-1)S_{er}}{60}$, $i = 1, 2, \ldots, 30$, and we calculated

$$P_{Hit}(x\le RT\le y) = \frac{\sum_{i=1}^{30} P_{Hit}(x-t_i\le t_s\le y-t_i)}{30} = \frac{\sum_{i=1}^{30}\left[P_{Hit}(t_s\le y-t_i)-P_{Hit}(t_s\le x-t_i)\right]}{30}$$

Similar calculations were performed for Correct Rejections.

Fitting the full model in steps Before fitting the full (13 parameter) model we fit the exhaustive search variant which contained one less free parameters ($q=\infty$). Note that in this case the above derivations simplify to

$$F_{Hit}(t_s) = 1-\left[\left(1-F_{T,up}(t_s)\right)\left(1-F_{D,up}(t_s)\right)^{n-1}\right]$$
$$F_{CR}(t_s) = F_{D,down}{}^{n}(t_s)$$

We then augmented the best fitting parameters of the exhaustive variant with a moderate quit unit exponent ($q=5$) as a starting point for fitting the full model (see Donkin, Brown & Heathcote, 2011 for a similar approach).

References

Akaike, H. (1973). *Information theory and an extension of the maximum likelihood principle*. In International Symposium on Information Theory, 2nd, Tsahkadsor, Armenian SSR (pp. 267-281).

Algom, D., Eidels, A., Hawkins, R., Jefferson, B., & Townsend, J. T. (2014). Features of response times: Identification of cognitive mechanisms through mathematical modeling. In J. R. Busemeyer, J. T. Townsend, Z. J. Wang, & A. Eidels (Eds.), *Oxford Handbook of Computational and Mathematical Psychology*. New York: Oxford University Press.

Bloomfield, J. R. (1979). Visual search with embedded targets: Color and texture differences. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 21*(3), 317–330.

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

Bravo, M. J., & Nakayama, K. (1992). The role of attention in different visual-search tasks. *Perception & Psychophysics, 51*(5), 465–472. doi:10.3758/BF03211642

Busemeyer, J. R., & Diederich, A. (2010). *Cognitive modeling*. New York, NY: Sage.

Cameron, E. L., Tai, J. C., Eckstein, M. P., & Carrasco, M. (2004). Signal detection theory applied to three visual search tasks—identification, yes/no detection and localization. *Spatial Vision, 17*(4), 295–325.

Cheal, M., Lyon, D. R., & Gottlob, L. R. (1994). A framework for understanding the allocation of attention in location-precued discrimination. *The Quarterly Journal of Experimental Psychology, 47A,* 699–739.

Cousineau, D., & Shiffrin, R. M. (2004). Termination of a visual search with large display size effects. *Spatial Vision, 17*(4/5), 327–352. doi:10.1163/1568568041920104

Cox, D. R., & Miller, H. D. (1965). *The theory of stochastic processes*. New York, NY: Chapman and Hall.

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

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

Donkin, C., & Shiffrin, R. (2011). Visual search as a combination of automatic and attentive processes. In Expanding the space of cognitive science: *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society* (pp. 2830-2835).

Dosher, B. A., Han, S., & Lu, Z.-L. (2004). Parallel processing in visual search asymmetry. *Journal of Experimental Psychology: Human Perception and Performance, 30*(1), 3–27. doi:10.1037/0096-1523.30.1.3

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

Downing, C. J. (1988). Expectancy and visual-spatial attention: Effects on perceptual quality. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 188–202.

Drugowitsch, J., Moreno-Bote, R., Churchland, A. K., Shadlen, M. N., & Pouget, A. (2012). The cost of accumulating evidence in perceptual decision making. *The Journal of Neuroscience, 32*(11), 3612–3628.

Eckstein, M. P., Thomas, J. P., Palmer, J., & Shimozaki, S. S. (2000). A signal detection model predicts the effects of set size on visual search accuracy for feature, conjunction, triple conjunction, and disjunction displays. *Perception & psychophysics, 62*(3), 425–451.

Eriksen, C. W., & Yeh, Y. Y. (1985). Allocation of attention in the visual field. *Journal of Experimental Psychology: Human Perception and Performance, 11*(5), 583–597.

Feller, W. (1968). *An introduction to probability theory and its applications*. New York, NY: Wiley.

Hawkins, G. E., Forstmann, B. U., Wagenmakers, E.-J., Ratcliff, R., & Brown, S. D. (2015). Revisiting the evidence for collapsing boundaries and urgency signals in perceptual decision-making. *Journal of Neuroscience, 35*(6), 2476–2484.

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2002). Quantile maximum likelihood estimation of response time distributions. *Psychonomic Bulletin & Review, 9*(2), 394–401.

LaBerge, D. (1995). *Attentional processing: The brain's art of mindfulness*. Harvard University Press.

LaBerge, D., & Brown, V. (1989). Theory of attentional operations in shape identification. *Psychological Review, 96,* 101–124.

Li, Z. (1999). Contextual influences in V1 as a basis for pop out and asymmetry in visual search. *Proceedings Of The National Academy Of Sciences Of The United States Of America, 96*(18), 10530–10535.

Liesefeld, H. R., Moran, R., Usher M., Müller, H. J., Zehetleitner, M. (2015). Search efficiency as a function of target saliency: The transition from inefficient to efficient search and beyond. Manuscript submitted for publication.

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

Moran, R. (2015). Optimal decision making in heterogeneous and biased environments. *Psychonomic Bulletin & Review, 22*(1), 38–53.

Moran, R., Teodorescu, A. R., & Usher, M. (2015). Post choice information integration as a causal determinant of confidence: Novel data and a computational account. *Cognitive psychology, 78,* 99–147.

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), 1–31. doi:10.1167/13.8.24.doi

Müller, H. J., & Humphreys, G. W. (1991). Luminance increment detection: Capacity limited or not? *Journal of Experimental Psychology: Human Perception and Performance, 17,* 107–124.

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

Palmer, J. (1995). Attention in Visual Search: Distinguishing Four Causes of a Set-Size Effect. *Current Directions In Psychological Science (Wiley-Blackwell), 4*(4), 118–123. doi:10.1111/1467-8721.ep10772534

Palmer, J., Ames, C. T., & Lindsey, D. T. (1993). Measuring the effect of attention on simple visual search. *Journal of Experimental Psychology: Human Perception and Performance, 19*(1), 108–130.

Palmer, J., & McLean, J. (1995). *Imperfect, unlimited-capacity, parallel search yields large set-size effects.* Paper presented at the annual meeting of the Society for Mathematical Psychology. Irvine, CA. Retrieved from. http://faculty.washington.edu/jpalmer/files/palmer_mclean1995talk.pdf

Palmer, J., Verghese, P., & Pavel, M. (2000). The psychophysics of visual search. *Vision Research, 40*(10-12), 1227–1268. doi:10.1016/S0042-6989(99)00244-8

Philiastides, M. G., Heekeren, H. R., & Sajda, P. (2014). Human scalp potentials reflect a mixture of decision-related signals during perceptual choices. *The Journal of Neuroscience, 34*(50), 16877–16889.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience, 13,* 25–42.

Rangelov, D., Müller, H. J., & Zehetleitner, M. (2013). Visual search for feature singletons: Multiple mechanisms produce sequence effects in visual search. *Journal Of Vision, 13(3)*, doi:10.1167/13.3.22

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*(2), 59–108.

Ratcliff, R., Hasegawa, Y. T., Hasegawa, Y. P., Smith, P. L., & Segraves, M. A. (2007). Dual diffusion model for single-cell recording data from the superior colliculus in a brightness-discrimination task. *Journal of Neurophysiology, 97,* 1756–1774.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics, 6*(2), 461–464.

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

Shaw, M. L. (1984). Division of attention among spatial locations-a fundamental difference between detection of letters and detection of luminance increments. *Attention and performance, 10,* 109–121.

Smith, P. L. (1990). A note on the distribution of response times for a random walk with Gaussian increments. *Journal of Mathematical Psychology, 34,* 445–459.

Smith, P. L. (2000). Stochastic dynamic models of response time and accuracy: A foundational primer. *Journal of Mathematical Psychology, 44,* 408–463.

Smith, P. L., & Sewell, D. K. (2013). A competitive interaction theory of attentional selection and decision making in brief, multielement displays. *Psychological review, 120*(3), 589–627.

Snodgrass, J. G., & Townsend, J. T. (1980). Comparing parallel and serial models: Theory and implementation. *Journal Of Experimental Psychology: Human Perception And Performance, 6*(2), 330–354. doi:10.1037/0096-1523.6.2.330

Thornton, T. L., & Gilden, D. L. (2007). Parallel and serial processes in visual search. *Psychological Review, 114*(1), 71–103. doi:10.1037/0033-295X.114.1.71

Thura, D., Beauregard-Racine, J., Fradet, C. W., & Cisek, P. (2012). Decision making by urgency gating: Theory and experimental support. *Journal of Neurophysiology, 108*(11), 2912–2930.

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

Townsend, J. T. (1976). Serial and within-stage independent parallel model equivalence on the minimum completion time. *Journal of Mathematical Psychology, 14,* 219–238.

Townsend, J. T. (1990). Serial vs. parallel processing: Sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished. *Psychological Science, 1,* 46–54. doi:10.1111/j.1467-9280.1990.tb00067.x

Treisman, A. (1988). Features and objects: The fourteenth Bartlett memorial lecture. *The Quarterly Journal of Experimental Psychology, 40*(2), 201–237.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97–136. doi:10.1016/0010-0285(80)90005-5

van Vugt, M. K., Simen, P., Nystrom, L., Holmes, P., & Cohen, J. D. (2014). Lateralized readiness potentials reveal properties of a neural mechanism for implementing a decision threshold. *PloS One, 9*(3), e90943.

Verghese, P. (2001). Visual search and attention: A signal detection theory approach. *Neuron, 31*(4), 523–535.

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition, 32*(7), 1206–1220.

Ward, R., & McClelland, J. L. (1989). Conjunctive search for one and two identical targets. *Journal Experimental Psychology: Human Perception And Performance, 15*(4), 664–672. doi:10.1037/0096-1523.15.4.664

Williams, P., Eidels, A., & Townsend, J. T. (2014). The resurrection of Tweedledum and Tweedledee: Bimodality cannot distinguish serial and parallel processes. *Psychonomic Bulletin & Review, 21*(5), 1165–1173. doi:10.3758/s13423-014-0599-0

Wolfe, J. M. (1994). Guided search 2.0 A revised model of visual search. *Psychonomic Bulletin & Review, 1*(2), 202–238.

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

Wolfe, J. M. (2007). Guided search 4.0. In W. D. Gray (Ed.), *Integrated models of cognitive systems (Cognitive Models and Architectures. pp. 99–120)*. Oxford: Integrated models of cognitive systems.

Wolfe, J. M., Palmer, E. M., & Horowitz, T. S. (2010). Reaction time distributions constrain models of visual search. *Vision Research, 50*(14), 1304–1311. doi:10.1016/j.visres.2009.11.002

Wolfe, J. M., & Van Wert, M. J. (2010). Varying target prevalence reveals two dissociable decision criteria in visual search. *Current Biology, 20,* 121–124. doi:10.1016/j.cub.2009.11.06

Zelinsky, G. J., & Sheinberg, D. L. (1997). Eye movements during parallel–serial visual search. *Journal of Experimental Psychology: Human Perception and Performance, 23*(1), 244–262.