Signal detection evidence for limited capacity in visual search

Evan M. Palmer,
Department of Psychology, Wichita State University, 1845 North Fairmount, Wichita, Kansas 67260-0034, USA

David E. Fencsik,
Department of Psychology, California State University, East Bay, Hayward, CA, USA

Stephen J. Flusberg,
Department of Psychology, Stanford University, Stanford, CA, USA

Todd S. Horowitz, and
Visual Attention Laboratory, Brigham and Women's Hospital, Cambridge, MA, USA. Department of Ophthalmology, Harvard Medical School, Boston, MA, USA

Jeremy M. Wolfe
Visual Attention Laboratory, Brigham and Women's Hospital, Cambridge, MA, USA. Department of Ophthalmology, Harvard Medical School, Boston, MA, USA

Evan M. Palmer: evan.palmer@wichita.edu

Abstract

The nature of capacity limits (if any) in visual search has been a topic of controversy for decades. In 30 years of work, researchers have attempted to distinguish between two broad classes of visual search models. *Attention-limited models* have proposed two stages of perceptual processing: an unlimited-capacity preattentive stage, and a limited-capacity selective attention stage. Conversely, *noise-limited models* have proposed a single, unlimited-capacity perceptual processing stage, with decision processes influenced only by stochastic noise. Here, we use signal detection methods to test a strong prediction of attention-limited models. In standard attention-limited models, performance of some searches (feature searches) should only be limited by a preattentive stage. Other search tasks (e.g., spatial configuration search for a "2" among "5"s) should be additionally limited by an attentional bottleneck. We equated average accuracies for a feature and a spatial configuration search over set sizes of 1–8 for briefly presented stimuli. The strong prediction of attention-limited models is that, given overall equivalence in performance, accuracy should be better on the spatial configuration search than on the feature search for set size 1, and worse for set size 8. We confirm this crossover interaction and show that it is problematic for at least one class of one-stage decision models.

Keywords

Theoretical and computational attention models; Visual search; Signal detection theory

Correspondence to: Evan M. Palmer, evan.palmer@wichita.edu.

In some visual searches, accuracy declines and response time (RT) slows as more stimuli are added to the display, while in others, performance is fairly constant as a function of the number of stimuli (the set size). There have been two seemingly fundamentally different approaches to understanding these *set size effects* in visual search. Two-stage theories of attention (e.g., Broadbent, 1958; Treisman & Gelade, 1980; Wolfe, Cave, & Franzel, 1989) involve *preattentive* and *attentive* stages in human visual processing. In the preattentive stage, a limited set of visual attributes are processed (e.g., color, orientation, motion; see Wolfe & Horowitz, 2004) with essentially unlimited capacity. Binding more than one feature to the same object (e.g., specifying the color and orientation of an object) requires serial access to a limited-capacity "attentional bottleneck," leading to set size effects. We will refer to this class of models as *attention-limited models*, since their defining characteristic is that access to a second stage of attentional processing is capacity limited.

Two-stage, attention-limited models have two loci at which performance can be influenced: the unlimited-capacity preattentive stage and the severely limited attentional bottleneck. Some stimuli are largely unaffected by the bottleneck. "Feature searches," in which the target is defined by a salient feature, such as orientation, will produce a strong preattentive signal that is essentially independent of set size. The preattentive stage delivers a single item, the target, to the attentional bottleneck. Under some circumstances, the preattentive stage may even generate a response on its own (e.g., Chan & Hayward, 2009; Treisman & Gelade, 1980). In any case, processing of these stimuli is not limited by the attentional bottleneck. Other stimuli are not differentiable in the preattentive stage, for instance a "spatial configuration search" for a "2" among "5"s. In this case, the properties of the preattentive stage are moot, and the limited-capacity attentional bottleneck governs performance. Thus, attention-limited models predict qualitatively different performance for feature and spatial configuration stimuli, even within the same yes/no detection task structure.

In contrast, unlimited-capacity models do away with the attentional bottleneck, and instead conceptualize visual search as detection of a signal among multiple noise sources in a single stage (e.g., Carrasco & McElree, 2001; Davis et al., 2006; Dosher, Han, & Lu, 2004; Eckstein, Thomas, Palmer, & Shimozaki, 2000; Kinchla, 1974; Palmer, 1995; Palmer & McLean, 1995; Shaw, 1980; Verghese, 2001). The observer compares a signal, derived from processing of the entire display, to a criterion. If the criterion is exceeded, the observer concludes that the target is present; if not, the observer concludes that the target is absent. These models assume unlimited processing capacity, with set size effects arising from the influence of noise at the decision stage. As set size increases, so do the number of noise sources from which a target must be distinguished. With more noise sources come more chances that some noise in a target-absent display will surpass the observer's criterion, causing a false alarm. We will refer to this class of models as *noise-limited models*, since they propose that set size effects arise only from the presence of more noise sources in higher-set-size displays.

One-stage noise-limited models have been successfully applied to simple feature searches and moderately difficult conjunction searches. For targets and distractors that differ along a

single feature dimension (i.e., feature search), set size effects have been successfully modeled by noise-limited theories for luminance increments, color differences, and size increments (Palmer, 1994); contrast and orientation differences (Eckstein, 1998); line length, rectangle aspect ratio and line orientation (Palmer, Ames, & Lindsey, 1993); mirror-symmetrical orientation differences (Davis et al., 2006); and asymmetrical orientation differences (Dosher, Han, & Lu, 2010). Converging evidence for the above findings comes from Verghese and Nakayama (1994), who argued that set size effects in feature searches for orientation, spatial frequency, and color differences could *not* be accounted for by a limited-capacity model (i.e., a two-stage attention-limited model). For search targets defined by a conjunction of two features, noise-limited models can account for set size effects in searches for luminance × orientation conjunctions (Eckstein, 1998; Eckstein et al., 2000) and for color × form conjunctions (McElree & Carrasco, 1999).

However, despite their success at accounting for set size effects in feature and conjunction search, in several cases in the literature, noise-limited models have not been able to fully account for performance decrements as a function of set size. Palmer (1994, 1995) tested observers on both line bisection stimuli (e.g., search for an "L" among "T"s) and point orientation stimuli, and he found that for both searches the increase in discrimination thresholds as a function of set size was larger than would be predicted by a decision integration model (i.e., noise-limited model), but not as large as that predicted by a perceptual-coding model (i.e., attention-limited model). Põder (1999) found that stimuli differing in the relative position of their components yielded set size effects larger than similar stimuli with a distinctive visual feature, consistent with limited-capacity processing of the relative position stimuli. Davis et al. (2006) examined baseline search for a tilted line among vertical lines, symmetrical search for a right-tilted line among left-tilted lines, and asymmetrical search for a vertical line among tilted lines. They found that a noise-limited model could account for set size effects for baseline and symmetrical search, but not for asymmetrical search: Decrements in performance as set size increased were larger than predicted by an unlimited-capacity parallel model of asymmetrical search. Davis et al. (2006) concluded that a serial-process limited-capacity model was consistent with the asymmetrical search data. Finally, Shaw (1984) showed that set size effects in searches for letter stimuli were larger than could be accounted for by a noise-limited model and were consistent with an attention-limited serial-process model.

The debate about capacity limits has persisted for decades because it is quite hard to produce unequivocal evidence for or against entire model classes. This has been true for the experiments, reviewed above, involving briefly presented stimuli. It has also been true for experiments involving RT measures. The bulk of studies of attention-limited models with a serial component have been RT experiments (Dukewich & Klein, 2005; Klein, 1988; Kwak, Dagenbach, & Egeth, 1991; Luck & Hillyard, 1990). However, it has been known for many years that the most common measures based on mean RTs are ambiguous with respect to the capacity limit question (Townsend, 1990; Townsend & Wenger, 2004). A few studies have applied more sophisticated tools to RT data (Bricolo, Gianesini, Fanini, Bundesen, & Chelazzi, 2002; Gilden, Thornton, & Marusich, 2010), but the matter remains unresolved.

We do not propose to settle the debate here. Our goal is to test a clear, qualitative prediction of two-stage attention-limited models, using the signal detection methods that have typically been deployed to study predictions of one-stage noise-limited models. The term *qualitative* is important here. Rather than asking what class of model better fits the data, we describe a situation in which a two-stage model *must* produce a crossover interaction in the data. If it did not produce this interaction, two-stage models of the type described above would be falsified. We did find such a crossover. Of course, this does not mean that two-stage attention-limited models are true, only that they have dodged this particular bullet. Similarly, the crossover does not prove all one-stage noise-limited models to be false. For instance, Cameron, Tai, Eckstein and Carrasco (2004) modeled two tasks, a target identification task and yes/no detection task, and found a similar interaction in their data that was consistent with their proposed noise-limited model. However, they were modeling two different tasks on the same stimuli with two levels of difficulty, while we are modeling the same task on two different stimuli with the same overall difficulty. Thus, observing a crossover interaction in our data in and of itself does not invalidate the noise-limited approach, but it does pose a challenge to those models, and does eliminate some classes of such models, notably those that propose the same decision rule (defined below) for both feature and spatial configuration stimuli.

The crossover prediction

Consider a briefly presented display of some small number of items that can all be resolved without moving the eyes. How do observers make the "yes" or "no" decision to respond "target present" or "target absent"? According to signal detection theory (SDT), an observer might sample some perceptual evidence from each stimulus in the display and then make a decision about target presence or absence on the basis of the largest signal (MAX rule). If this maximum value is greater than some criterion, the observer responds "target present"; otherwise, a "target absent" response is given. Under a MAX rule, performance decreases as set size increases (both hits and false alarms rise, but false alarms rise faster), because more stimuli means more chances for a distractor noise sample to exceed criterion (Green & Swets, 1966). Figure 1a shows simulated results for such a decision rule.[1] Different curves are produced by different levels of discriminability between target and distractors. For searches in which the target can be identified by a single perceptual attribute (e.g., orientation or size), a MAX rule captures the data quite well (Davis et al., 2006; Eckstein et al., 2000; Palmer et al., 1993; Verghese, 2001). Furthermore, Green and Swets (1966) demonstrated that the MAX rule yields the highest probability of a correct response, which may explain why it is the most widely used decision rule in one-stage noise-limited modeling of visual search. Other rules are possible, however. For example, one could use a SUM rule in which all signals are added and a "target present" response is produced if the sum exceeds some criterion (Baldassi & Burr, 2000). In this case, the shape of the function in Fig. 1a would change but the parallel nature of the family of curves would not.

---

[1]We use the standard detectability (d') measure here and in the Results simply to summarize performance in the search task (hit rates and false alarm rates), not to estimate underlying perceptual sensitivity, which requires taking into account the increase in the number of noise sources with set size. Decreases in this d' measure with set size should not, in and of themselves, be interpreted as evidence for limited capacity. In the Appendix, we test the fit of different models to provide a more formal analysis; this analysis confirms the findings based on d'.

Now suppose that there is a capacity limit that prevents the processing of all items at the same time, as might happen under an attention-limited model of search. For purposes of illustration, let us suppose that the visual system processes some items serially and that this search display is presented briefly enough that only a single item can be processed. If the display was presented for a long time, this two-stage model would propose that more items would be selected and processed. If the display set size is 1, the noise-limited one-stage and attention-limited two-stage models would converge on a simple signal detection problem: Is the signal from this one item above or below my criterion for saying "target present"? If there are two or more items, however, the models would diverge. As noted, the noise-limited model makes more errors when there are more noise samples, since there is a higher chance that one of those samples will exceed the criterion. The two-stage model, on the other hand, still has only the one sample from the one item that was selected. If the one selected item is not the target, the model must guess or just respond "target absent." In either case, the prediction must be that performance falls more dramatically as set size increases for attention-limited models than it would for the noise-limited model situation depicted in Fig. 1a. This steeper performance decline is shown in Fig. 1b, which assumes that only one item is selected. If more than one item were selected in a brief exposure, the steep decline in performance would occur when the set size exceeded the number of selected items.

Now, suppose that we arrange to have two search tasks produce the same overall performance, averaged across set sizes. If one of them is governed by one rule (e.g., the MAX rule of Fig. 1a) and the other by another rule that produces a different decline of performance with set size (e.g., LIMITED CAPACITY, as in Fig. 1b), the two performance × set size functions must cross (Fig. 1c). Specifically, if average performance is equated across set sizes, performance will be better for the attention-limited search task than for the noise-limited search task at set size 1 and worse at the highest set size.

Two-stage models, such as an attention-limited model, mandate this crossover interaction. Under these models, some searches will be performed entirely by the unlimited-capacity preattentive stage, while other searches will require access to a second, attention-limited stage of processing. In our guided search model (Wolfe, 1994, 2007), for example, the selection of one item would be "guided" by the parallel, noise-limited preattentive stage. In a search for a basic feature (such as a search for an item of unique orientation), the limit on that selection, and thus the limit on performance of the task, would be in the preattentive stage, and assuming a MAX rule, performance would fall off as shown in Fig. 1a. In a spatial configuration task, the preattentive stage in a two-stage model would provide no information about the presence or location of the target. Selection would be random, and performance would follow Fig. 1b. If overall performance is equated for the two tasks, then there must be a crossover interaction in the performance × set size functions, as illustrated in Fig. 1c. We tested this prediction in two experiments.

Experiment 1

The same observers performed the same yes/no decision task with two different search stimuli. In one condition, observers searched for a tilted bar among vertical bars. Orientation discriminations, even for stimuli degraded by noise, are thought to be processed in parallel

across multiple items (Baldassi & Verghese, 2002; Davis et al., 2006; Morgan & Solomon, 2006; Palmer et al., 1993). In the other condition, observers searched for a digital "2" among digital "5"s. A character discrimination such as "2" versus "5" may be easy to perform under focal attention, but in attention-limited theories, the characters are thought to be preattentively indistinguishable (Wolfe & Bennett, 1997), requiring limited-capacity serial deployments of attention (Kwak et al., 1991). Typically, an orientation stimulus difference would be easier to perceive than a "2" versus "5" stimulus difference. However, we degraded the orientation stimuli, moving performance below "2" versus "5" search at set size 1. If performance in the yes/no decision task for the two different stimuli was governed by the same decision rule, then orientation stimuli would remain harder to detect than "2" versus "5" stimuli at all set sizes, as in Fig. 1a and b. If *and only if* the two stimuli are governed by different decision rules, the performance × set size functions will cross, as in Fig. 1c.

To create conditions for a crossover interaction, we developed a search task in which the difficulty of identifying stimuli (i.e., the detectability of targets) could be adjusted independently of the search stimuli or the set size. We accomplished this by adding visual noise to each stimulus (see Fig. 2). Participants performed a yes/no decision task with set size varying from trial to trial. We used a weighted staircase procedure (Kaernbach, 1991) to adjust noise opacity, increasing it after each correct response and decreasing it after errors, regardless of set size or target presence or absence. Since one staircase was run over all set sizes, this procedure controlled average performance. In this case, the staircase yielded average performance of 80% correct across all set sizes. With the same amount of added noise for all set sizes, accuracy would be above 80% for low set sizes and below 80% for high set sizes. If performance on each stimulus set was governed by the same decision rule, this method would produce the same performance × set size function for both displays. Otherwise, this method adjusted overall search performance on each stimulus set into a range in which the functions could exhibit a crossover interaction. Each participant completed the staircase procedure, followed by data collection at a fixed noise level determined by the staircase.

We performed two versions of the experiment, one with a physical variation of set size, and the other with an attentional variation of set size. In the attentional set size manipulation (Palmer, 1994), eight items were displayed on every trial, but only a subset of the items (the relevant set size for that trial) were cued before the search display appeared (Fig. 3). The physical set size manipulation followed the same procedure, except that only the items that were cued were displayed. All participants completed both manipulations in random order, in different blocks. If observers could not restrict attention to the relevant items, these two manipulations should yield very different results.

Method

Participants—A group of 5 observers (1 female, 4 male) between the ages of 24 and 38 years served as participants. One of the participants (E.M.P.) was an author, while the other 4 were naïve. All had normal or corrected-to-normal visual acuity, passed the Ishihara color vision test, and gave informed consent. The naïve participants received $10/h compensation.

Apparatus—The stimuli were presented and the data gathered on an Apple Macintosh G4 450-MHz computer driving a 20-in. (diagonal) monitor at a resolution of 1,024 × 768 pixels. Responses were recorded with an Apple Macintosh USB keyboard. The experiment was controlled using MATLAB 5.2.1 and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997).

Stimuli—The display background was dark gray with a central white square that served as a focal point and subtended 0.8° × 0.8° of visual angle. On each trial, eight dark gray placeholder squares (each 4.7° × 4.7° of visual angle) surrounded the central white square in a circle formation, with a center-to-center distance of 10° of visual angle between the focal square and the larger stimulus squares. For orientation search, the target was an oblique light-gray rectangle (subtending 4.2° × 0.5° visual angle, tilted 20° clockwise from vertical), and the distractors were vertical light-gray rectangles. For spatial configuration search, the target was a light-gray digital "2" and the distractors were digital "5"s, both subtending 4.2° × 4.2° of visual angle. Stimuli were separated by at least 2° of visual angle to reduce crowding effects (Levi, 2008).

The level of visual noise was manipulated by varying the percentage opacity of a uniform noise field superimposed over the stimuli. Noise level was determined separately for each observer and each stimulus condition by a staircase procedure. The staircase procedure increased the opacity by 2.5% after every correct response and decreased it by 10% after every incorrect response, regardless of set size or target presence or absence; these step sizes converged on an accuracy level of 80% (Kaernbach, 1991). The staircase was run until 40 reversals, then the noise level was set to the average of the previous 20 reversals. Across the 5 participants, the average noise levels were 62% (*SD* 3.5%) for the orientation stimuli and 31% (*SD* 18.5%) for the "2" versus "5" stimuli (higher percentages indicate greater opacity of the noise field, meaning more obscured stimuli).

Procedure—The trial procedure is illustrated in Fig. 3. Participants sat at a computer in a dark testing room. Responses were entered on the keyboard, and error feedback was given after each trial. The observers were told to prioritize accuracy over speed.

The orientation stimuli and the "2" versus "5" stimuli, as well as the attentional and physical set size manipulations, were run in separate blocks in random order. For each stimulus type and set size, participants practiced at a low noise level to familiarize them with the task; next, they completed the staircase procedure; finally, they completed the experimental trials. Each block of experimental trials consisted of 100 target-present and 100 target-absent trials for each set size of 1, 2, 4, and 8, for both the physical and attentional set size manipulations. Participants completed 3,200 experimental trials among well over 4,000 total trials, including practice and the staircase procedure.

Data analysis—For convenience, we converted accuracy to d', defined as z(hit rate) − z (false alarm rate). For cells with perfect performance, we added $1 \div (2 \times$ number of trials) errors (Wickens, 2002).

Results

Figure 4 plots performance by set size, averaged across the 5 observers. Both the physical set size (panel A) and attentional set size (panel B) manipulations produced a crossover effect; performance was better on the "2" versus "5" stimuli than on the orientation stimuli at set size 1, but was worse on the "2" versus "5" stimuli at set size 8.

These observations were confirmed by a $2 \times 2 \times 4$ (display manipulation $\times$ search stimuli $\times$ set size) within-subjects ANOVA. The crossover effect manifested in a significant interaction between search stimuli (oriented bars or "2"s vs. "5"s) and set size, $F(3, 12) = 40.25$, $p < .001$, $\eta_p^2 = .91$. This result, combined with the lack of a three-way interaction ($F = 1.45$, $p = .28$, $\eta_p^2 = .27$), indicates that the crossover interaction was observed independently of the display manipulation method (physical vs. attentional cue). The two methods of manipulating set size produced the same result: Neither the display manipulation effect ($F = 2.91$, $p = .163$, $\eta_p^2 = .42$) nor the interaction with set size ($F = 1.28$, $p = .32$, $\eta_p^2 = .24$) was significant. The lack of a main effect of search stimuli indicated that our staircase procedure successfully brought overall performance on the two stimulus sets into the same range ($F = 1.26$, $p = .33$, $\eta_p^2 = .24$). There was a significant main effect of set size, $F(3, 12) = 75.86$, $p < .001$, $\eta_p^2 = .95$, and a reliable interaction of display manipulation $\times$ search stimuli, $F(1, 4) = 78.85$, $p < .005$, $\eta_p^2 = .95$.

To more closely examine the critical interaction of search stimuli by set size, we averaged performance across the display manipulations and performed *t* tests comparing the orientation stimuli and the "2" versus "5" stimuli for set sizes 1 and 8. The *t* tests show that observers performed reliably better on "2" versus "5" stimuli than on orientation stimuli at set size 1, $t(8) = 2.71$, $p < .05$, $d = 1.92$, and reliably worse on "2" versus "5" stimuli than on orientation stimuli at set size 8, $t(8) = 2.51$, $p < .05$, $d = 1.77$. These two results, taken in combination with the significant search stimuli $\times$ set size interaction, indicate an unambiguous crossover interaction.[2]

Experiment 2

Experiment 1 established that the same observers performing the same yes/no visual search task with stimuli of equal overall difficulty used different decision rules for the different search tasks. However, these search tasks were performed in different blocks. Perhaps the decision rules were not a property of the visual system, but were merely strategies that observers adopted when faced with blocks of different stimuli. To address this possibility, we replicated the attentional set size condition of Experiment 1 but randomly intermixed the orientation and "2" versus "5" search stimuli on a trial-by-trial basis. Thus, on any given trial, observers did not know whether they would be performing an orientation search or a spatial configuration search. Observers would need to be *very* flexible in order to differentially apply one decision rule on orientation search trials and another on "2" versus "5" trials.

---

[2]For more detailed model fitting to these data, refer to the Appendix.

Method

Participants—A group of 5 observers (1 female, 4 male) between the ages of 22 and 37 years served as participants. One of the participants (E.M.P.) was an author, while the other 4 were naïve. All had normal or corrected-to-normal visual acuity, passed the Ishihara color vision test, and gave informed consent.

Apparatus—Stimuli were presented and responses gathered on the same computers as in Experiment 1 for observers A. M.S. and C.A.C., but A.T.M., C.M.B., and E.M.P. completed the experiment on a 2-GHz Apple Mac Pro computer driving a 17-in. (diagonal) Dell M991 monitor at a resolution of $1,400 \times 1,050$ pixels. The experiment was programmed using MATLAB (version 7.5) and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997).

Stimuli—The same stimuli and timing parameters were used as in Experiment 1. Only the attentional set size manipulation was run, meaning that set sizes of 1, 2, 4, and 8 were precued as in Experiment 1, and eight stimuli were shown on every trial. Thus, any differences in performance as a function of set size could be attributed to attentional selection and not sensory differences (Palmer, 1994).

As in Experiment 1, the level of visual noise was manipulated by varying the percentage opacity of a uniform noise field superimposed over the stimuli. Noise levels for each stimulus type were determined separately for each observer by a staircase procedure. Across the 5 participants, the average noise levels were 67.5% (*SD* 3.5%) for the orientation stimuli and 17% (*SD* 13%) for the "2" versus "5" stimuli.

Procedure—Observers completed several sessions of training over 2–3 days before the experiment began. In the first phase of training, each participant performed several hundred trials of practice on the orientation and "2" versus "5" stimuli separately, with no noise, until they could consistently achieve greater than 90% correct on a block of 100 trials. Next, the two stimulus sets were randomly intermixed on a trial-by-trial basis, again with no noise, until observers could consistently perform at greater than 90% correct. Finally, the staircase procedure described in Experiment 1 was run on the intermixed stimulus sets to adjust the level of noise separately for the orientation and "2" versus "5" stimuli. The overall visual noise opacity for each stimulus set was independently adjusted to yield performance of about 80% correct, averaged across set sizes.

In the experimental phase, the stimuli were presented with fixed noise opacity, as in Experiment 1. Data was collected in four sessions. In each session, observers completed 50 practice trials and then 25 trials per cell (4 set sizes × 2 TP/TA × 2 stimuli), or 450 trials per session, with a programmed break every 50 trials. Observers were only allowed to complete two sessions per day, and data collection was completed over 2–3 days. There were 1,600 experimental trials, but when combined with practice and training, each observer completed over 2,400 trials in this experiment.

Results

Figure 5 plots performance by set size averaged across the 5 observers. The averaged performance of the 5 participants showed a clear crossover interaction, with higher performance on "2" versus "5" search than on orientation search at set size 1, but lower performance on "2" versus "5" search at set sizes 4 and 8.

These observations were confirmed by the statistical analyses. We performed a 2×4 (search stimuli $\times$ set size) within-subjects ANOVA, which indicated that the main effect of set size, $F(3, 12) = 144.75$, $p < .001$, $\eta_p^2 = .97$, and the interaction of search stimuli by set size $F(3, 12) = 10.513$, $p < .01$, $\eta_p^2 = .72$, were both significant. The main effect of search stimuli was not significant ($F = 6.13$, $p = .069$, $\eta_p^2 = .61$), indicating that our efforts to equalize the overall perceptibility for the two sets of search stimuli were successful.

We performed a more detailed analysis of the critical performance $\times$ set size crossover interaction via t tests between the "2" versus "5" and orientation stimuli at each set size. Those analyses indicated that overall performance was reliably better for the "2" versus "5" search stimuli than for the orientation stimuli at set size 1, $t(8) = 2.71$, $p < .05$, $d = 1.77$, performance was equivalent at set size 2, $t(8) = 0.047$, $p = .96$, and performance on the "2" versus "5" stimuli was reliably worse at set sizes 4, $t(8) = 2.67$, $p < .05$, $d = 1.89$, and 8, $t(8) = 2.76$, $p < .05$, $d = 1.95$. As in the last experiment, observers were significantly better at discriminating a "2" versus "5" at set size 1 but better at discriminating a tilted from a vertical line at set sizes 4 and 8 (see note 2).

General discussion

The central finding of this article is that two-stage attention-limited models survive a test carried out using signal detection methods that have been more typically supportive of one-stage noise-limited models. Two-stage attention-limited models are required to predict that an orientation feature search and a "2" versus "5" spatial configuration search are constrained in different ways. In the context of this experimental design, this means that the two tasks behave as though operating under different decision rules. When overall performance is equated, this situation demands a crossover interaction for the two functions that relates performance to set size. Specifically, performance needed to be better for the "2" versus "5" task at set size 1 and worse at set size 8. This is what we observed.

This finding does not mean that one-stage noise-limited models are necessarily incorrect. It would, however, appear to falsify one-stage models with a single decision rule for both of these tasks. If both tasks were being performed under the same rule, the performance $\times$ set size functions should have been identical, or at least parallel, depending on whether or not we succeeded in equating overall task difficulty. If "2" versus "5" performance was better at set size 1, it should have remained better at set size 8. Working within the one-stage noise-limited framework, Davis et al. (2006) proposed that multiple decision rules were at work; in this context, our result highlights the fact that any one-stage noise-limited model should specify a priori which decision rules apply to which stimuli, even when the two search tasks are intermixed.

Though the emphasis of this article has been on the qualitative prediction of a crossover interaction, it is possible to ask quantitative questions about the ability of specific decision rules to explain the pattern of results. The results of one such exercise can be found in the Appendix to this article, which shows that an unlimited-capacity MAX rule predicts performance data better from orientation search than from "2" versus "5". Additionally, the Appendix presents individual data for the two experiments and the details of the model fitting.

Discussions of the existence of capacity limitations in visual search and whether one-stage or two-stage process models are better at predicting human search performance have often been hindered by methodological differences. Experiments evaluating two-stage, attention-limited models tend to use unlimited exposure durations and to gather RTs as the dependent measure of interest, whereas experiments evaluating one-stage, noise-limited models tend to use short exposure durations and to gather accuracy as the primary dependent measure. However, a useful theory of visual search should be successful in both experimental domains. We hope that the present effort to evaluate attention-limited models of visual search using traditional noise-limited methods will help promote a more productive exploration of the relative strengths and weaknesses of these two approaches to understanding visual search.

Acknowledgments

References

Baldassi S, Burr DC. Feature-based integration of orientation signals in visual search. Vision Research. 2000; 40:1293–1300.10.1016/S0042-6989(00)00029-8 [PubMed: 10788640]

Baldassi S, Verghese P. Comparing integration rules in visual search. Journal of Vision. 2002; 2(8):3, 559–570.10.1167/2.8.3

Brainard DH. The psychophysics toolbox. Spatial Vision. 1997; 10:433–436.10.1163/156856897X00357 [PubMed: 9176952]

Bricolo E, Gianesini T, Fanini A, Bundesen C, Chelazzi L. Serial attention mechanisms in visual search: A direct behavioral demonstration. Journal of Cognitive Neuroscience. 2002; 14:980–993.10.1162/089892902320474454 [PubMed: 12419123]

Broadbent, D. Perception and communication. New York: Oxford University Press; 1958.

Burnham, KP.; Anderson, DR. Model selection and multimodel inference: A practical information-theoretic approach. 2. New York: Springer; 2002.

Cameron EL, Tai JC, Eckstein MP, Carrasco M. Signal detection theory applied to three visual search tasks—identification, yes/no detection and localization. Spatial Vision. 2004; 17:295–325.10.1163/1568568041920212 [PubMed: 15559107]

Carrasco M, McElree B. Covert attention accelerates the rate of visual information processing. Proceedings of the National Academy of Sciences. 2001; 98:5363–5367.10.1073/pnas.081074098

Chan LKH, Hayward WG. Feature integration theory revisited: Dissociating feature detection and attentional guidance in visual search. Journal of Experimental Psychology Human Perception and Performance. 2009; 35:119–132.10.1037/0096-1523.35.1.119 [PubMed: 19170475]

Cousineau D. Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. Tutorials in Quantitative Methods for Psychology. 2005; 1:42–45.

Davis ET, Shikano T, Main K, Hailston K, Michel RK, Sathian K. Mirror-image symmetry and search asymmetry: A comparison of their effects on visual search and a possible unifying explanation. Vision Research. 2006; 46:1263–1281.10.1016/j.visres.2005.10.032 [PubMed: 16376402]

Dosher BA, Han S, Lu ZL. Parallel processing in visual search asymmetry. Journal of Experimental Psychology Human Perception and Performance. 2004; 30:3–27.10.1037/0096-1523.30.1.3 [PubMed: 14769065]

Dosher BA, Han S, Lu ZL. Information-limited parallel processing in difficult heterogeneous covert visual search. Journal of Experimental Psychology Human Perception and Performance. 2010; 36:1128–1144.10.1037/a0020366 [PubMed: 20873936]

Dukewich KR, Klein RM. Implications of search accuracy for serial self-terminating models of search. Visual Cognition. 2005; 12:1386–1403.10.1080/13506280444000788

Eckstein MP. The lower visual search efficiency for conjunctions is due to noise and not serial attentional processing. Psychological Science. 1998; 9:111–118.10.1111/1467-9280.00020

Eckstein MP, Thomas JP, Palmer J, Shimozaki SS. A signal detection model predicts the effects of set size on visual search accuracy for feature, conjunction, triple conjunction, and disjunction displays. Perception & Psychophysics. 2000; 62:425–451.10.3758/BF03212096 [PubMed: 10909235]

Gilden DL, Thornton TL, Marusich LR. The serial process in visual search. Journal of Experimental Psychology Human Perception and Performance. 2010; 36:533–542.10.1037/a0016464 [PubMed: 20515187]

Green, DM.; Swets, JA. Signal detection theory and psychophysics. New York: Wiley; 1966.

Kaernbach C. Simple adaptive testing with the weighted up–down method. Perception & Psychophysics. 1991; 49:227–229.10.3758/BF03214307 [PubMed: 2011460]

Kinchla RA. Detecting target elements in multielement arrays: A confusability model. Perception & Psychophysics. 1974; 15:149–158.10.3758/BF03205843

Klein R. Inhibitory tagging system facilitates visual search. Nature. 1988; 334:430–431.10.1038/334430a0 [PubMed: 3405288]

Kwak HW, Dagenbach D, Egeth H. Further evidence for a time-independent shift of the focus of attention. Perception & Psychophysics. 1991; 49:473–480. [PubMed: 2057313]

Levi DM. Crowding—An essential bottleneck for object recognition: A mini-review. Vision Research. 2008; 48:635–654.10.1016/j.visres.2007.12.009 [PubMed: 18226828]

Luck SJ, Hillyard SA. Electrophysiological evidence for parallel and serial processing during visual search. Perception & Psychophysics. 1990; 48:603–617. [PubMed: 2270192]

McElree B, Carrasco M. The temporal dynamics of visual search: Evidence for parallel processing in feature and conjunction searches. Journal of Experimental Psychology Human Perception and Performance. 1999; 25:1517–1539.10.1037/0096-1523.25.6.1517 [PubMed: 10641310]

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

Morgan MJ, Solomon JA. Attentional capacity limit for visual search causes spatial neglect in normal observers. Vision Research. 2006; 46:1868–1875.10.1016/j.visres.2005.11.026 [PubMed: 16430942]

Palmer J. Set-size effects in visual search—The effect of attention is independent of the stimulus for simple tasks. Vision Research. 1994; 34:1703–1721. [PubMed: 7941377]

Palmer J. Attention in visual search: Distinguishing four causes of a set-size effect. Current Directions in Psychological Science. 1995; 4:118–123.10.1111/1467-8721.ep10772534

Palmer J, Ames CT, Lindsey DT. Measuring the effect of attention on simple visual search. Journal of Experimental Psychology Human Perception and Performance. 1993; 19:108–130.10.1037/0096-1523.19.1.108 [PubMed: 8440980]

Palmer, J.; McLean, J. *Imperfect, unlimited-capacity, parallel search yields large set-size effects*. Paper presented at the annual meeting of the Society for Mathematical Psychology; Irvine, California. 1995.

Pelli DG. The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spatial Vision. 1997; 10:437–442.10.1163/156856897X00366 [PubMed: 9176953]

Põder E. Search for feature and for relative position: Measurement of capacity limitations. Vision Research. 1999; 39:1321–1327.10.1016/S0042-6989(98)00253-3 [PubMed: 10343845]

Shaw, ML. Identifying attentional and decision-making components in information processing. In: Nickerson, RS., editor. Attention and performance VIII. Hillsdale: Erlbaum; 1980. p. 277-296.

Shaw, ML. Division of attention among spatial locations: A fundamental difference between detection of letters and detection of luminance increments. In: Bouma, H.; Bouwhuis, DG., editors. Attention and performance X: Control of language processes. Hillsdale: Erlbaum; 1984. p. 109-121.

Townsend JT. Serial vs. parallel processing: Sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished. Psychological Science. 1990; 1:46–54.10.1111/j. 1467-9280.1990.tb00067.x

Townsend JT, Wenger MJ. The serial-parallel dilemma: A case study in a linkage of theory and method. Psychonomic Bulletin & Review. 2004; 11:391–418. [PubMed: 15376788]

Treisman AM, Gelade G. A feature-integration theory of attention. Cognitive Psychology. 1980; 12:97–136.10.1016/0010-0285(80)90005-5 [PubMed: 7351125]

Verghese P. Visual search and attention: A signal detection theory approach. Neuron. 2001; 31:523–535. [PubMed: 11545712]

Verghese P, Nakayama K. Stimulus discriminability in visual search. Vision Research. 1994; 34:2453–2467.10.1016/0042-6989(94)90289-5 [PubMed: 7975284]

Wickens, T. Elementary signal detection theory. New York: Oxford University Press; 2002.

Wolfe JM. Guided Search 2.0: A revised model of visual search. Psychonomic Bulletin & Review. 1994; 1:202–238.10.3758/BF03200774 [PubMed: 24203471]

Wolfe, JM. Guided Search 4.0: Current progress with a model of visual search. In: Gray, WD., editor. Integrated models of cognitive systems. New York: Oxford University Press; 2007. p. 99-119.

Wolfe JM, Bennett SC. Preattentive object files: Shapeless bundles of basic features. Vision Research. 1997; 37:25–43.10.1016/S0042-6989(96)00111-3 [PubMed: 9068829]

Wolfe JM, Cave KR, Franzel SL. Guided search: An alternative to the feature integration model for visual search. Journal of Experimental Psychology Human Perception and Performance. 1989; 15:419–433.10.1037/0096-1523.15.3.419 [PubMed: 2527952]

Wolfe JM, Horowitz TS. What attributes guide the deployment of visual attention and how do they do it? Nature Reviews Neuroscience. 2004; 5:495–501.10.1038/nrn1411

Appendix

In the main text, we use the standard d' measure (Green & Swets, 1966) to summarize accuracy in our visual search task and to demonstrate the crossover interaction. Figure 1 demonstrates that d' is a valid measure for testing the crossover prediction. In that figure, we plot d' from performance predicted by MAX rule models with and without capacity limits. The figure demonstrates that $d' \times$ set size functions will cross over each other only under different decision rules. Thus, we can use it to test for a change in decision rules across different sets of search stimuli.

In this Appendix, we formally test our conclusions by fitting different models of visual search to our data. We demonstrate that, for most observers, an UNLIMITED-CAPACITY MAX rule model explains performance with the orientation stimuli, but a LIMITED-CAPACITY model is more appropriate for the "2" versus "5" stimuli.

We used the MAX rule formula described by Palmer et al. (1993), developing a version of it with a capacity limit. Under this model, each distractor generates activation drawn from the distribution $F(x) = \Phi(x)$, where Φ is the cumulative normal distribution, while target activity

is drawn from the distribution $G(x) = F(x − s) = \Phi(x − s)$, where s is the difference between the target and distractor distributions (i.e., sensitivity).

The LIMITED-CAPACITY version of the model selects k of the n stimuli, where k is capacity and n is set size, and computes the maximum value. If the maximum value is greater than a criterion, c, then a target-present response is given; otherwise, a target-absent response is given. Thus, a false alarm will occur whenever any of the k selected distractors are greater than c on a target-absent trial, so the false alarm rate will be

$$P_{\text{false alarm}} = 1 - [F(c)]^{k}. \quad \text{(A1)}$$

A hit will occur in one of two circumstances: when the target is among the k sampled stimuli and either it or one of the $k − 1$ distractors exceeds c, or when the target is not sampled but one of the k distractors exceeds c (in the equation below, we calculate 1 minus the opposite of this probability). Thus, the hit rate will be

$$P_{\text{hit}} = 1 - \left\{ P(t \in K) F(c-s) [F(c)]^{k-1} + P(t \notin K) [F(c)]^{k} \right\}, \quad \text{(A2)}$$

where $t \in K$ refers to when the target t is in K, the set of k sampled stimuli, and $t \notin K$ is the opposite. We can calculate the probability that t is not in K:

$$P(t \notin K) = \frac{\binom{n-1}{k}}{\binom{n}{k}} = \frac{n-k}{n}$$

Since $P(t \in K) = 1 - P(t \notin K)$, we can express Eq. A2 as follows:

$$P_{\text{hit}} = 1 - \left\{ \frac{k}{n} F(c-s) [F(c)]^{k-1} + \frac{n-k}{n} [F(c)]^{k} \right\}. \quad \text{(A3)}$$

Note that when capacity is unlimited, $k = n$, and Eqs. A1 and A3 reduce to the unlimited-capacity max-rule model derived by Palmer et al. (1993, Eq. A9).

First, we fit both UNLIMITED-CAPACITY and LIMITED-CAPACITY versions of Eqs. A1 and A3 to each observer's results from Experiments 1 and 2. We treated s (sensitivity) and c (criterion) as free parameters, allowing both to vary across observers and stimulus sets, and allowing c only to vary across set size. We estimated the best-fitting parameter values using maximum likelihood. For the UNLIMITED-CAPACITY model, we fixed $k = n$. For the LIMITED-CAPACITY model, we treated k as a free parameter, allowing it to vary across observers and stimulus sets, but not set size.

Figure 6 plots the model fits against the observed data for both Experiments 1 and 2. The first column for each experiment shows the results from orientation search, along with the fit

of the UNLIMITED-CAPACITY model; adding the capacity parameter did not improve the fit (the fits were identical), so that model is not shown. The second column of each experiment includes the results of the "2" versus "5" search with the predictions of the UNLIMITED-CAPACITY model. The third column in each graph plots the same results with the predictions of the LIMITED-CAPACITY model. To illustrate the quality of each model's predictions, we calculated Akaike's information criterion (AIC; Burnham & Anderson, 2002) for each model fit and included it in each plot. AIC measures the goodness of fit of a model to observed data, accounting for the number of free parameters in the model. Lower values of AIC indicate a better fit.

To formally test the goodness of fit of each model, we conducted generalized likelihood ratio tests (GLRTs) comparing the fits of the UNLIMITED-CAPACITY and LIMITED-CAPACITY models. In essence, the GLRT indicates whether the addition of the capacity parameter reliably improves the model fit. Table 1 reports the GLRT results for each observer under each set of search stimuli. For the orientation search, the UNLIMITED-CAPACITY model fit best for all but 1 of the observers across the two experiments. However, for the "2" versus "5" search, the LIMITED-CAPACITY model fit best for all but 1 observer. Thus, the UNLIMITED-CAPACITY model appears to explain performance on the orientation search task, but the LIMITED-CAPACITY model is better for the "2" versus "5" search task.

We also investigated the usefulness of varying criterion with set size. In most cases, the model with variable criterion provided a reliably better fit. The only exceptions were subjects A.M.S. and C.A.C. in the "2" versus "5" condition of Experiment 2. However, in both cases, the LIMITED-CAPACITY model was still a better fit than the UNLIMITED-CAPACITY model, even when the criterion was held constant across set size.

In summary, the data and arguments from the main article show that averaged $d' \times$ set size functions for orientation and "2" versus "5" searches produce a crossover interaction, indicating that different decision rules are engaged by the different stimulus sets. The detailed modeling reported in this Appendix shows that orientation feature search data are captured quite well by an UNLIMITED-CAPACITY MAX decision rule, but "2" versus "5" spatial configuration search data are not. Rather, "2" versus "5" spatial configuration search data are better described by a LIMITED-CAPACITY decision rule.
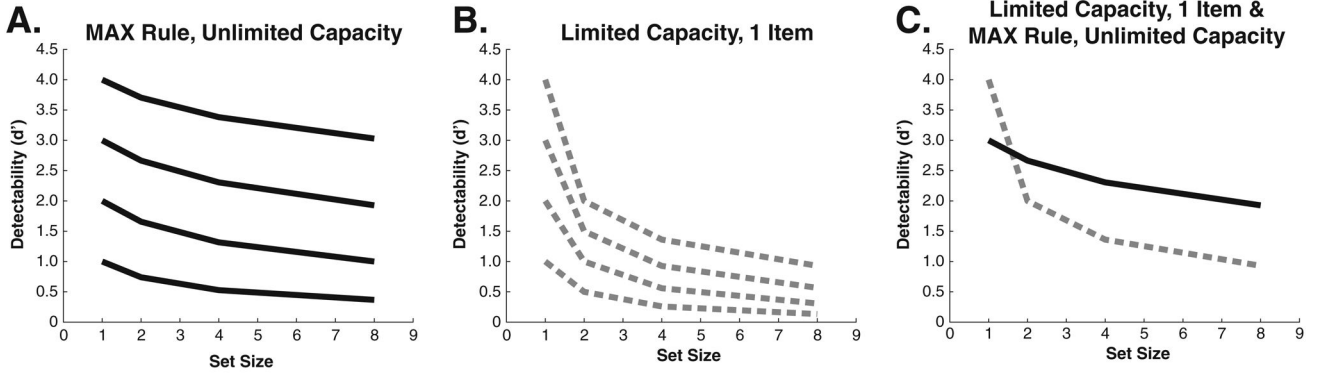
Fig. 1.
Simulated detectability (d') as a function of the number of items in the display and the type of decision rule being used. (**a**) The $d' \times$ set size functions for four different values of signal and noise distribution distances using the max rule. (**b**) The $d' \times$ set size functions for the same signal and noise distribution distances using the limited capacity, one-item rule. (**c**) Notice that $d' \times$ set size functions only cross over each other if they arise from different decision rules
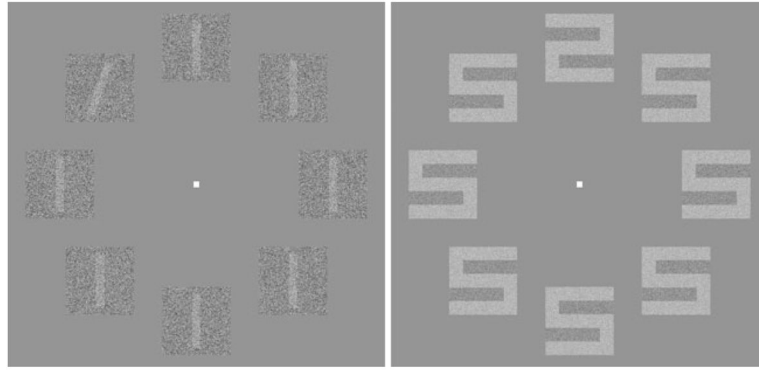
Fig. 2.
The two types of stimuli used in the experiment. Orientation stimuli are depicted on the left, and "2" vs. "5" stimuli are depicted on the right. Different amounts of noise were added to each stimulus in order to bring overall performance for both stimuli into the same range of average performance
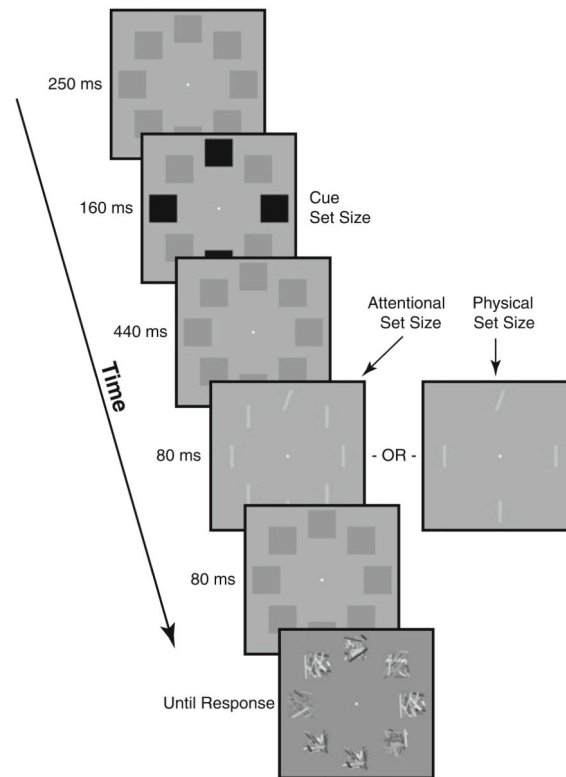
Fig. 3.
The timing and display sequence of a set size 4 orientation stimulus trial. Stimuli were presented for 80 ms and then masked after an 80-ms ISI. Participants had an unlimited amount of time to respond. Both the attentional and physical set size manipulations are depicted
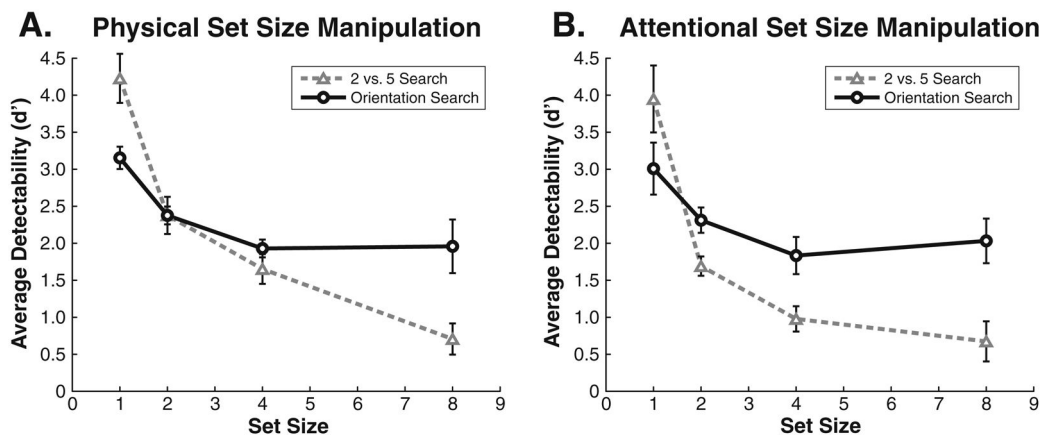
Fig. 4.

d' as a function of set size for Experiment 1. Physical set size was changed in panel A, while an attentional cue was used to manipulate set size in panel B. In both cases, the detectability functions cross over each other, indicating that different decision rules are necessary to account for performance on the two stimuli. Display manipulation and search stimuli varied across blocks, while set size varied within blocks. Error bars represent within-subjects confidence intervals (Cousineau, 2005; Morey, 2008)
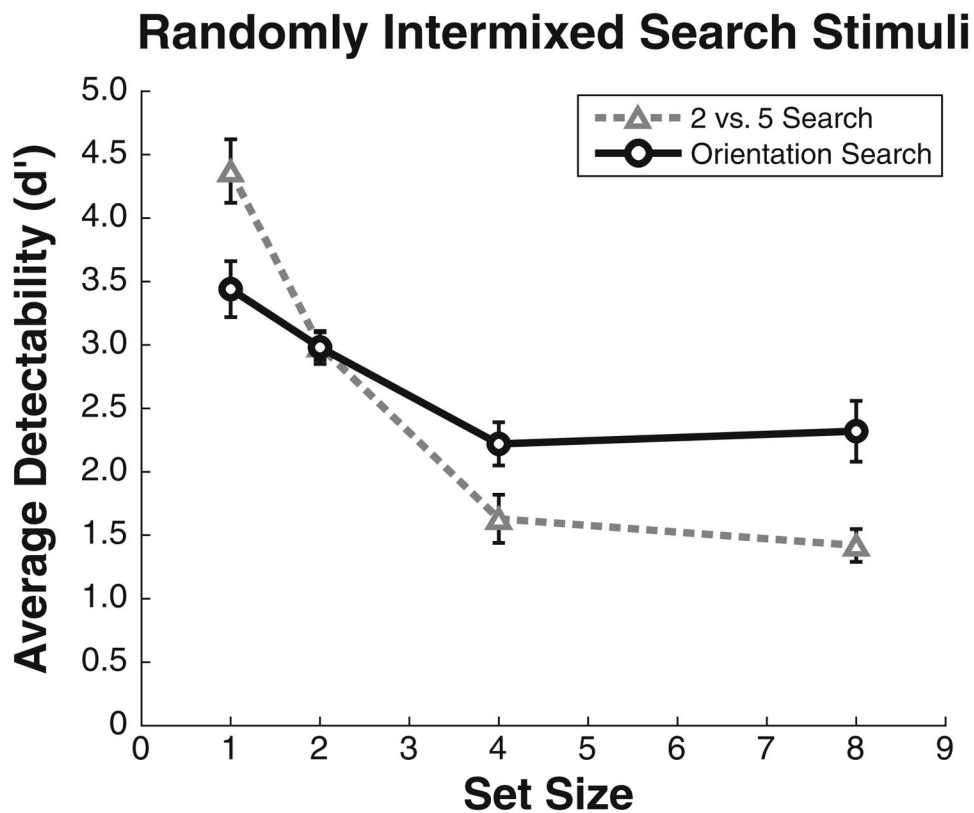
Randomly Intermixed Search Stimuli

Fig. 5.
d' as a function of set size for Experiment 2. Search stimuli ("2" vs. "5" or orientation) and set size were randomly intermixed on a trial-by-trial basis. Error bars represent within-subjects confidence intervals (Cousineau, 2005; Morey, 2008)
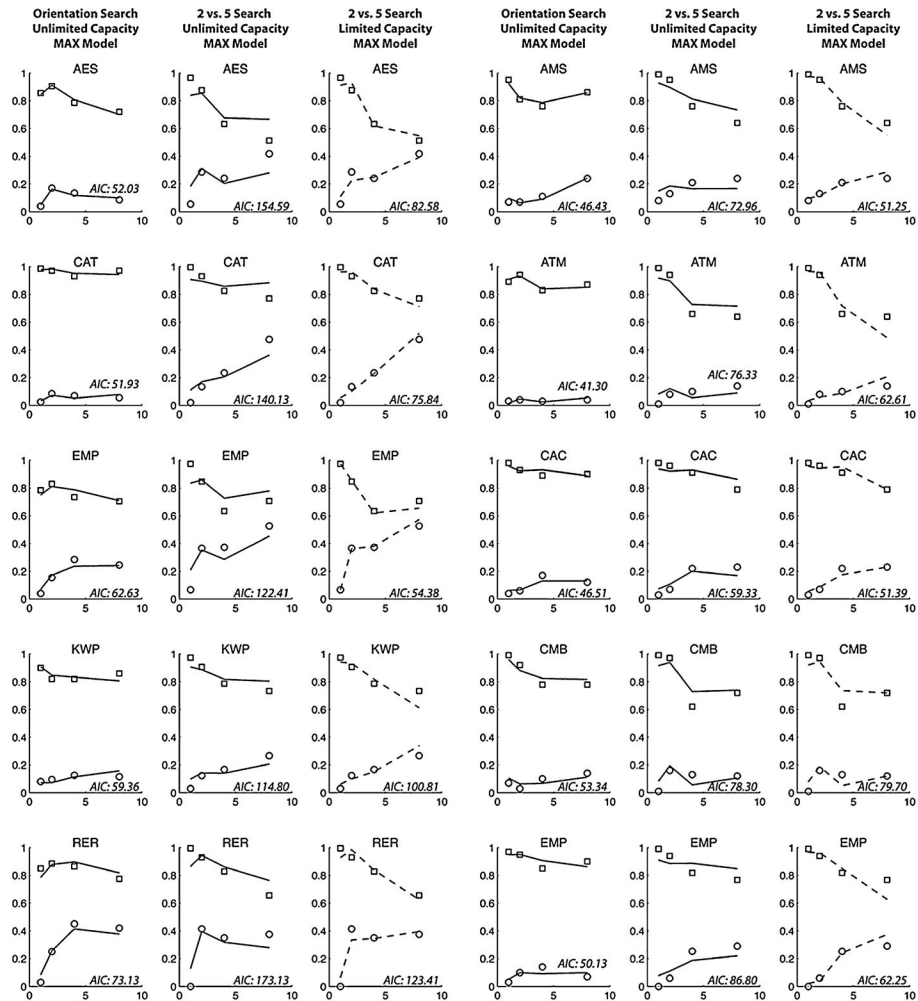
Fig. 6.

Hit rates (squares) and false alarm rates (circles) for Experiments 1 and 2 for all observers. For each experiment, the left column of graphs shows data for orientation feature search, which is well fit by the UNLIMITED-CAPACITY MAX model (solid lines). The middle column shows data from the spatial configuration "2" versus "5" experiment, UNLIMITED-CAPACITY MAX model fits. The right column shows the same "2" versus "5" data, but fit with a LIMITED-CAPACITY MAX model (dashed lines). The Akaike information criterion (AIC) for each graph is reported in the lower right. Lower AIC scores indicate a better fit of the model to the data. For 9 of the 10 observers, the LIMITED-CAPACITY model fit the "2" versus "5" data better than the UNLIMITED-CAPACITY MAX model

Table 1

Generalized likelihood ratio tests of the UNLIMITED-CAPACITY and LIMITED-CAPACITY MAX rule models

	Observer	χ_1^2
Orientation Stimuli		
Experiment 1	A.E.S.	0.000
	C.A.T.	0.000
	E.M.P.	0.047
	K.W.P.	0.000
	R.E.R.	5.253^*
Experiment 2	A.M.S.	0.000
	A.T.M.	0.000
	C.A.C.	0.000
	C.M.B.	2.498
	E.M.P.	0.000
"2" vs. "5" Stimuli		
Experiment 1	A.E.S.	74.002^{***}
	C.A.T.	66.298^{***}
	E.M.P.	70.028^{***}
	K.W.P.	15.982^{***}
	R.E.R.	51.716^{***}
Experiment 2	A.M.S.	23.714^{***}
	A.T.M.	15.716^{***}
	C.A.C.	9.939^{**}
	C.M.B.	0.606
	E.M.P.	26.553^{***}

The generalized likelihood ratio test (GLRT) statistic is used to compare the LIMITED-CAPACITY to the UNLIMITED-CAPACITY model. A significant GLRT indicates that the LIMITED-CAPACITY model is a better fit.

[*]
$p < .05.$

[**]
$p < .01.$

[***]
$p < .001$