AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology:
Human Perception and Performance

ISSN: 0096-1523

2021, Vol. 47, No. 2, 252–268
https://doi.org/10.1037/xhp0000887

Prioritization in Visual Attention Does Not Work the Way You Think It Does

Gavin J. P. Ng, Simona Buetti, Trisha N. Patel, and Alejandro Lleras
Department of Psychology, University of Illinois at Urbana–Champaign

A common assumption in attention theories is that attention prioritizes search items based on their similarity to the target. Here, we tested this assumption and found it wanting. Observers searched through displays containing *candidates* (distractors that cannot be confidently differentiated from the target by peripheral vision) and *lures* (distractors that can be). Candidates had high or low similarity to the target. Search displays were either candidate-homogeneous (all items of same similarity) or candidate-heterogeneous (equal numbers of each similarity). Response times to candidate-heterogeneous displays were equivalent to the average of high- and low-similarity displays, suggesting that attention was allocated randomly, rather than toward the high-similarity candidates first. Lures added a response time cost that was independent of the candidates, suggesting they were rejected prior to candidates being inspected. These results suggest a "reverse" prioritization process: Distributed attention discards least target-similar items first, while focused spatial attention is randomly directed to target-similar items.

Public Significance Statement
Most theories of attention propose that attention visits locations in a scene in descending order of their priority, with attentional priority reflecting an object's similarity to the target object or feature that the observer is looking for. Here we propose that the inverse is in fact true. Attention starts by evaluating peripheral information in parallel and rejecting unlikely targets as a function of their dissimilarity to the target; that is, attention moves up the similarity scale, not down. Furthermore, we propose that given the processing limitations of peripheral vision, attention cannot be properly guided by visual similarity at the top of the priority scale: Once similarity is sufficiently high, attention simply visits potential target locations at random. This new bottom-to-top conceptualization of attentional processing should be of wide interest to anyone working in an attention-related field, applied or theoretical.

Keywords: visual search, attention, prioritization, open data, open materials

Central to most theories of visual selection is the concept of attentional prioritization: the idea that early visual processing produces an ordered list of locations in the visual scene for attention to examine. The specifics of how this ordered list of locations is arrived at vary between theories, as does the term used to describe the priority "score" of a location. For instance, in Wolfe's (1994, 2006) Guided Search, this "score" is referred to as "activation", while in Zelinsky's (2008) Target Acquisition Model it is referred to as "priority", and it is referred to as saliency in Itti and Koch's (2000) saliency model (and its various later modifications, e.g., Navalpakkam & Itti, 2005, 2007). The higher this saliency score, the higher the attentional priority of that location. Attention is thought to inspect the scene by going down a priority list starting from the location with the highest activation. Locations already visited are "scratched off" the list through variants of an inhibition of return mechanism with varying degrees of memory for inspected locations (Itti & Koch, 2000; McCarley et al., 2003). Finally, the ranking in this priority list can be impacted by factors such as eye movements (because the resolution with which an item is processed depends on where it falls on the retina; Balas et al., 2009; Rosenholtz, Huang, Raj, et al., 2012; Zelinsky, 2008) and noise (Wolfe, 1994, 2006).

The key assumptions underlying this prioritization account are that (a) the visual system is able to compute (even if noisily) a priority score for each item in the scene that veridically reflects that item's attentional importance (determined by that item's sim-

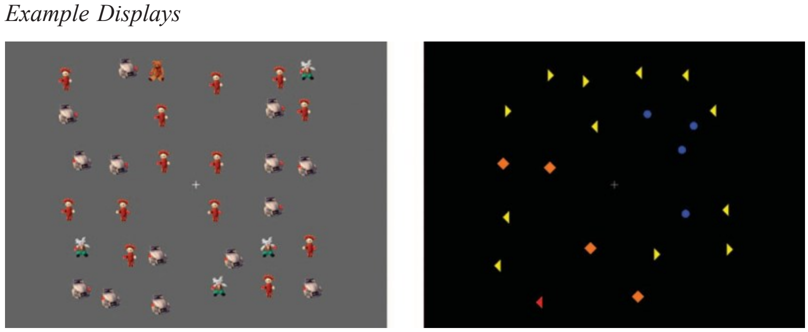
This article was published Online First December 14, 2020.
Gavin J. P. Ng https://orcid.org/0000-0003-3858-194X
Simona Buetti https://orcid.org/0000-0002-1718-450X
Trisha N. Patel https://orcid.org/0000-0002-0962-0287
Alejandro Lleras https://orcid.org/0000-0003-0391-1355
Alejandro Lleras and Simona Buetti developed the study concept. Alejandro Lleras, Simona Buetti and Trisha N. Patel developed Experiment 2, and Alejandro Lleras, Simona Buetti, and Gavin J. P. Ng developed Experiments 1 and 3. Gavin J. P. Ng performed the data analysis and interpretation under the supervision of Alejandro Lleras and Simona Buetti. The manuscript was drafted by Gavin J. P. Ng, and Alejandro Lleras and Simona Buetti provided revisions. All authors approved the final version of the manuscript. This material is partly based upon work supported by the National Science Foundation under Grant BCS1921735 to Simona Buetti.
Correspondence concerning this article should be addressed to Gavin J. P. Ng, Department of Psychology, University of Illinois at Urbana–Champaign, 603 East Daniel Street, Champaign, IL 61820, United States. Email: gavin.jp.ng@gmail.com

ilarity to the target or its saliency, depending on the specific model), and (b) attention works down the list of priority scores. The goal of the present study is to demonstrate that there are problems with these assumptions. First, resolution and processing limitations in peripheral vision make the computation of the list difficult, if not impossible; as objects become more and more similar to the target template, peripheral analysis of those items becomes less and less reliable. Second, recent evidence suggests that the orderly discarding of items starts from the bottom of the similarity scale, not at the top, with least similar items being discarded faster than less dissimilar items (Lleras et al., 2019; Wang et al., 2017). Indeed, take displays like the ones shown in Figure 1. Wang et al. (2017) demonstrated through computational simulations that the reaction time (RT) to find the teddy bear in the scene in Figure 1a can be predicted by assuming that search starts by initially processing all the items in the scene. As time progresses, peripheral analysis of the scene allows for items to be rejected (as nontargets) in parallel. This parallel rejection occurs in an orderly fashion with the least target-similar items (the white cars) being rejected first and the more target-similar items (the red dolls) being rejected later. Lleras et al. (2019) extended those findings to scenes composed of simpler colored geometric shapes (Figure 1b). The time to find the red triangle target in the scene is also predicted by assuming the same parallel rejection process of all nontarget items in Wang et al. (2017). Blue circles (lowest target-distractor similarity) are rejected first, followed by yellow triangles (medium target-distractor similarity) and then orange diamonds (highest target-distractor similarity). The important take-away from these results is that, even in the presence of higher-similarity items, RTs to find the target are impacted by the time taken to reject lower similarity items, suggesting these items are not "glossed over" by attentive processing. If attention had truly started "at the top" of the similarity scale and moved down from there, the lower similarity items ought to almost never have impacted RTs since the target should almost always be found before attention even visits the low-similarity distractors. Therefore, these results suggest attention starts at the bottom of the similarity scale, initially considering all search items as likely targets, and then it moves up that scale, rejecting more and more items that are unlikely to be the target.

This orderly rejection of items that are unlikely to be the target continues until the remaining distractors are relatively similar to the target, at which point focused attention inspects these items in a random order (Buetti et al., 2016; Lleras et al., 2020). This proposal follows the target contrast signal theory, which posits that the output of early visual processing is a contrast value that indexes how dissimilar items in the scene are to the target. Rejection of dissimilar items, having large contrast values (referred to as *lures*) takes time; this time cost increases logarithmically with the number of items that are rejected (e.g., Buetti et al., 2016; Lleras et al., 2020; Ng et al., 2018; Wang et al., 2017; Wang et al., 2018). The logarithmic increase in response times as a function of lures is indicative of a parallel, stochastic process for evaluating items in the display (Buetti et al., 2016; Lleras et al., 2020; Townsend & Ashby, 1983). Because of the resolution limitations of peripheral vision, rejection of similar items, having small contrast values (referred to as *candidates*) cannot occur in parallel. Candidate rejection requires focused attention and incurs a linear cost with the number of candidates. These items are not ordered by contrast values precisely because of the unreliability of those small contrast values. Thus, target contrast signal theory proposes that the inspection of these target-similar items by attention is not guided by similarity but occurs in random order.

There has yet to be direct evidence to support the major claim in target contrast signal theory that attentional scrutiny is random, or at least, not guided by similarity or "attentional priority." Such evidence is crucial for the theory because one could have just as easily specified a model where the accumulated contrast (accumulated during the initial processing and rejection of lures) drives the deployment of attention toward accumulators with lower contrast values (i.e. toward locations with higher target-distractor similarity values). Such a "contrast-guided" selection mechanism would produce more efficient search by increasing the probability of selecting the target location (if, on average, there is a lower contrast value at the target location than at candidate locations), and more generally, by increasing the rate of evidence accumula-

Figure 1
Example Displays



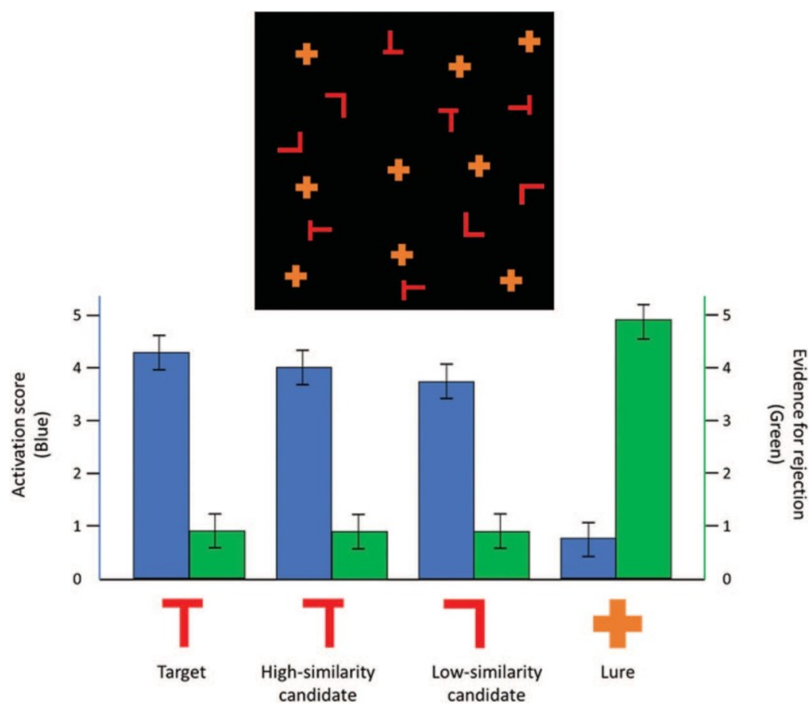
Note. Example displays from Wang et al. (2017)—left panel—and Lleras et al. (2019)—right panel. In these heterogeneous search displays, RTs to find the target (a teddy bear on the left and a red triangle on the right) are almost perfectly predicted by a model that incorporates the time taken to discount in parallel all the nontarget items in the display, with less similar items being rejected faster than more similar items. See the online article for the color version of this figure.

tion at the lowest-contrast locations (i.e. the slowest accumulating locations). Indeed, the major models of visual search subscribe to the account whereby the deployment of attention is prioritized toward items that are most similar to the target. One could also further imagine other kinds of rules that would prioritize some accumulators over others based on the information that was accumulated during the lure-rejection process. Yet, target contrast signal theory proposes that selection of items for focused attention unfolds without taking into account any of the information accumulated at the nonrejected location: It is indeed random with respect to the contrast (or similarity) to the target.

In the present study, we put this proposal to the test by examining the extent to which participants prioritize candidates from high-to-low similarity or whether these candidates are instead inspected randomly. Consider a task in which an observer searches for a red T in a display of orange crosses (lures) and red Ls and "offset-Ls" (low- and high-similarity candidates, respectively; Figure 2). Attentional prioritization theories (e.g., Ehinger et al., 2009; Najemnik & Geisler, 2005, 2008; Navalpakkam & Itti, 2005, 2007; Wolfe, 2006; Zelinsky, 2008) predict that attention should first visit the high-similarity candidates and will likely find the target even before visiting the low-similarity candidates and lures. That is to say, the target ought to have a priority score that is comparable to the high-similarity candidates but much higher than the low-similarity candidates and lures (blue bars in Figure 2). If this is the case, search times should not be impacted by the presence of lures and low-similarity candidates. The functional set size (Neider & Zelinsky, 2008) should thus be the number of high-similarity candidates, plus the target. We refer to this as the *ideal prioritization model*. Note however that this is an "ideal" and unlikely scenario because activation values are inherently noisy (Wolfe, 1994, 2006). Indeed, if this were not the case, then the target would always have the highest activation value and would be found with a single deployment of attention. Thus, the ideal prioritization model refers to a maximum prioritization (given noisy activation values) and represents the best a priority model could perform, and as such it represents an informative boundary condition to which we can compare human performance. Furthermore, even if activation values are noisy, this model still ought to predict that attention completely ignores lures (when they are present) given their distinctiveness from the target template.

Figure 2

Example of Search Display (Top) and Depiction of Model Mechanisms (Bottom)



Note. *Top*: A display with a target (red [gray] T), low-similarity candidates (red [gray] Ls), high-similarity candidates (red [gray] offset-Ls), and lures (thick orange [light gray] crosses). *Bottom*: Prioritization theories propose that early visual processing involves the calculation of activation values with some noise (blue[dark gray] bars; units are arbitrary). This results in a top-to-bottom prioritization of attention (even if imperfect) based on target-distractor similarity. On the other hand, target contrast signal theory proposes that early visual processing instead involves an evidence accumulation process toward a nontarget threshold with the goal of rejecting in parallel items that are visually distinct from the target. The evidence that is accumulated depends on target-distractor dissimilarity (green [gray] bars; units are arbitrary). See the online article for the color version of this figure.

In contrast, Target Contrast Signal Theory predicts that distributed attention will spend time processing lures in parallel across the display. After rejecting those stimuli, attention ought to randomly scrutinize candidates, irrespective of their similarity relation to the target. In other words, the functional set size of attentional scrutiny would be that of all candidates (plus the target) regardless of their similarity to the target. This is because early visual processing involves a rejection of locations that have accumulated sufficient evidence to reach nontarget thresholds (green bars in Figure 2). Lure locations reach threshold in a systematic way, whereas candidate locations do not (Buetti et al., 2016; Lleras et al., 2020). These nonrejected locations are then randomly scrutinized by focused attention. We refer to this as the *random scrutiny model.*

Note that the two models represent the two extreme possibilities of how attention scrutinizes items: either completely randomly, or perfectly guided by the level of (noisy) activation of each item. In the Results sections we calculate a *prioritization score* that reflects the degree to which attention prioritizes items according to activation levels, which can vary from 1 (perfect prioritization) to 0 (random scrutiny), with negative numbers representing situations where revisitations to already scrutinized locations occur.

We first present an experiment demonstrating that observers can differentiate between the two types of candidates (low- and high-similarity) that were used in Experiments 2 and 3 (see Figure 2)[1]. It is important to properly calibrate the stimuli to avoid potential misinterpretations. For instance, one could create candidates that are either too small or too similar to one another, such that they can only be distinguished from one another via direct foveation. In such scenarios, random scrutiny of candidates would necessarily be observed. Thus, the goal of Experiment 1 was to show that observers can (a) identify these stimuli in the periphery, even if imperfectly, and (b) that they do bear significantly different levels of similarity to the target. Using these well-calibrated stimuli, we then show, in Experiments 2 and 3, that the random scrutiny model is a better description of the data compared to the ideal prioritization model, indicating that serial attentional scrutiny is random and not prioritized by target-candidate similarity.

Experiment 1

The goal of this experiment was to determine whether participants can identify the candidates, even if imperfectly, when they are presented in peripheral locations and surrounded by similar levels of low-level noise (i.e. the presence of lures), similar to the conditions used in Experiments 2 and 3. Displays were flashed briefly to prevent eye movements so that we could assess target-distractor discriminability at three different levels of target eccentricity, encompassing all eccentricities used in Experiments 2 and 3. Participants were asked to report whether the candidate in the display was a T or not. This is the same perceptual discrimination that the visual system must perform to categorize stimuli in the Experiments 2 and 3. This experiment was preregistered on Open Science Framework (https://osf.io/9gktr/). The data and materials can be found at https://osf.io/5n2rt/.

Participants

Participants were recruited from the subject pool from the University of Illinois at Urbana–Champaign. Participants provided informed consent, which was approved by the Institutional Review Board at the University of Illinois at Urbana–Champaign, and were given course credit for taking part in the experiment. The study was run in accordance to the principles expressed in the Declaration of Helsinki. All participants had normal or corrected-to-normal vision and were determined to be noncolorblind using the Ishihara color plates before the start of the experiment. We planned on a sample size of 25 participants, which was determined to be sufficient to detect an effect of η_p^2 .58 at 95% power and α .05. This corresponded to the main effect of the increase in response times as a function of lure set size in search displays with lures and candidates (Experiment 3A in Buetti et al., 2016). Although the required sample size was determined to be 12, we decided to increase it to 25 to reduce measurement noise and to keep the sample size consistent across our many experiments on this topic. In total, 27 participants were recruited (14 Females, mean age 21.3). Data from the first two participants were excluded due to an error in the experimental code that resulted in an incorrect number of experimental trials.

Stimuli and Procedure

There were four kinds of stimuli: three candidates and one lure. The three candidates were: a red T (the target), a letter L, an "offset-L" that was created by shifting the vertical of the letter L by 0.2° to the right, and the lure, which was a thick orange plus sign (see Figure 2). All candidates were randomly presented in one of four possible orientations (rotated in clockwise steps of 90 degrees) except for the letter T, which was rotated either 90 or 180 degrees clockwise. All stimuli subtended .833 of visual angle and were randomly distributed across a 36-point grid. The 36 locations were equally distributed over three concentric rings with varying eccentricities (4.17, 7.73, and 14.3 degrees of visual angle). This concentric display was used to allow for a better estimation of the effect of eccentricity on target discriminability. On each trial, one of the 36 locations contained a candidate (T, L, or offset-L), while the remaining locations contained lures (orange crosses). Participants responded to the identity of the candidate, which was always presented on each trial, by pressing the right arrow key if it was a T or the left arrow key if it was not a T (L or offset-L). Response buttons were counterbalanced across participants. In total, the target T candidate was presented on 50% of the trials, the L candidate was presented on 25% of the trials, and the offset-L candidate was presented on the remaining 25%. There were 720 trials in total.

Each trial began with a fixation cross in the center of the screen for 500 ms. The display was then presented for 100 ms to prevent eye movements. The fixation cross remained visible on the screen during this time. The display then offset to a blank black screen for 2500 ms, during which participants made their response. Upon response, the blank screen continued for another 1500 ms, after which the next trial began. All stimuli were presented against a black background on a 22-in. (400 mm × 300 mm) CRT monitor with a refresh rate of 85Hz and a screen resolution of 1024 × 768

[1] The order in which the experiments were carried out, chronologically, was: Experiment 2, Experiment 3, Experiment 1. We have chosen to present Experiment 1 first in order to highlight the fact that the candidates indeed differed in their similarity to the target.

pixels. Participants viewed the display unrestrained from a distance of approximately 59 cm. All experiments were programmed using Psychopy (Peirce et al., 2019).

Results

All analyses, in this and the following experiments, were conducted in R (R Core Team, 2018). The definition of accuracy depends on the stimulus. For Ts, accuracy refers to the hit rate (responding T). For Ls and offset-Ls, accuracy refers to the correct rejection rate (responding not-T). A one-sample *t* test revealed that overall accuracy was significantly greater than chance (83% vs. 50%), $t(24)$ 23.92, p < .001, 95% CI of mean difference in accuracy: [0.30, 0.36]. A one-way ANOVA revealed that observers differed in their accuracy depending on the candidate type (see Figure 3), $F(2, 48)$ 66.17, p_c < .001, ω_p^2 .72, ε .732 (corrected for sphericity violations using the Greenhouse-Geisser procedure). Follow-up paired-samples *t* tests revealed that Ls were responded to more accurately than Ts, $t(24)$ 8.71, p < .001, 95% CI: [0.10, 0.16] and offset-Ls, $t(24)$ 15.079, p < .001, 95% CI: [0.15, 0.20]. Accuracy did not differ between Ts and offset-Ls (after applying the Bonferroni correction for multiple comparisons), $t(24)$ 2.28, p .0318, 95% CI: [0.0044, 0.088].

We also conducted an exploratory paired-samples *t* test to examine the difference in accuracy between the Ls and offset-Ls. Ls (*M* .93, *SD* .048) were responded to more accurately than offset-Ls (*M* .76, *SD* .069), $t(24)$ 15.08, p < .001, 95% CI: [0.15, 0.20]. This provides further support that the Ls and offset-Ls were not confusable, even at brief presentation times of 100 ms.

Accuracy is potentially an imperfect measure in this experiment since the decision criterion for each candidate is not the same. There is no need for the observer to discriminate between an L or offset-L or to identify it at all (all that is required is to decide that it is not the T). On the other hand, the observer has to identify that the candidate is the letter T in order to respond correctly. We thus present another measure of performance, *d*=. Using the confusion matrix presented in Table 1, we calculated the hit rate and false alarm rate for each individual participant. *D*= was then calculated. A one-sample *t* test revealed that *d* (*M* 1.98, *SD* 0.55) was significantly different from zero, $t(24)$ 17.91, p < .001, 95% CI: [1.75, 2.21].

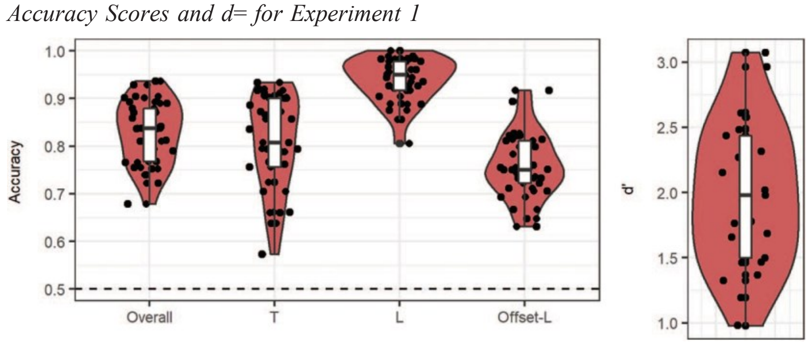
In addition, a one-way ANOVA revealed that *d*= differed depending on candidate eccentricity, $F(2, 48)$ 100.37, p < .001, ω_p^2 .80. Follow-up *t* tests revealed that *d*= was smaller when the candidate was in the furthest eccentricity (14.3°; *M* 1.44, *SD* 0.60) compared to the nearest eccentricity (4.17°; *M* 2.43, *SD* 0.66), $t(24)$ 11.23, p < .001, 95% CI: [0.80, 1.16] as well as the middle eccentricity (7.73°; *M* 2.35, *SD* 0.61), $t(24)$ 12.31, p < .001, 95% CI: [0.76, 1.06]. There was no significant difference between the middle and nearest eccentricities, $t(24)$ 1.03, p .31, 95% CI: [−0.072, 0.21].

Discussion

Experiment 1 demonstrated that, even under very short presentation times (100 ms), the visual system is able to discriminate between Ls, offset-Ls and Ts, albeit with less-than-perfect recognition performance. In addition, Ts were more confusable with offset-Ls, indicating that the two were more similar to each other than to the Ls. Performance, as measured by *d*=, was the worst when the target was in the farthest eccentricity, although there was no difference between the middle and nearest eccentricities. This is not surprising, given that the decrease in resolution as a function of eccentricity is well-known. Importantly, *d*= was still relatively high (1.44) even in the furthest eccentricity, especially considering that the exposure time was only 100 ms. In the following experiments, we show that although this information regarding visual similarity

Figure 3

Accuracy Scores and d= for Experiment 1



Note. *Left*: Each dot (randomly jittered horizontally) represents the mean accuracy of an individual participant. The leftmost plot summarizes overall accuracy, demonstrating that observers are well above chance (50%, indicated by the dashed line) at identifying the target T from candidates L and offset-L. The higher accuracy for Ls than offset-Ls also confirms that offset-Ls were indeed more similar (confusable) with the T than the L stimuli, suggesting that offset-Ls were more likely to be confused with Ts due to their increased visual similarity. *Right:* Average *d*= in Experiment 1. Each dot represents the mean *d* of an individual participant. Average *d*= was relatively high, further suggesting that participants were able to differentiate between the different candidates. See the online article for the color version of this figure.

Table 1

Confusion Matrix Used to Calculate d′ in Experiment 1

Response type	T	L or offset-L
Respond 'T'	Hit	False alarm
Respond 'not T'	Miss	Correct rejection

is available to observers, they seemed reticent to use this information to guide attentional scrutiny.

Experiment 2

If attention scrutinizes items in decreasing order of target-distractor similarity (as proposed by the ideal prioritization model), then lures, which have very low target-distractor similarity, should not contribute to processing times. In addition, attention should visit the high-similarity candidates first and will probably find the target before visiting the low-similarity candidates. We examined the former by comparing response times of displays with and without lures. We then addressed the latter by using candidate-homogeneous displays to predict response times on candidate-heterogeneous displays. If attention is guided by target-distractor similarity, then response times on candidate-heterogenous displays should be equivalent to response times of candidate-homogeneous displays with high-similarity distractors of the same set size.

Participants

A group of new participants were recruited from the same pool of subjects as in Experiment 1. We planned on a sample size of 25 participants, which would be more than sufficient to measure the difference between the two candidate search slopes with 95% power and $\alpha = .05$. This corresponded to the main effect of lure set size ($\eta_p^2 = .58$) in search displays that contain both lures and candidates, which are similar to the stimuli used here (Experiment 3A in Buetti et al., 2016). Although the required sample size was determined to be 12, we decided to increase it to 25 to reduce noise (i.e. to obtain more accurate estimates of each condition mean). Due to the nature of scheduling timeslots, 27 participants took part in this experiment in total (22 Females, mean age 19.1). There were 2 participants with accuracy rates lower than 80% who were excluded from the analyses. The final sample size was thus 25.

Stimuli and Procedure

All stimuli were identical to those used in Experiment 1. The study was designed as a 2 (lure presence: 0 or 24 lures) × 2 (candidate set size: 4 or 8) × 3 (candidate type: homogenous high-similarity, homogenous low-similarity, or heterogenous) within-subjects experiment. The experiment was programmed such that all nontarget candidates were distributed evenly between the four quadrants of the search display; the same was done for the lures. Depending on the condition, there were thus 5, 9, 29, or 33 items in the search display. All stimuli were randomly distributed across a 36-point grid which subtended 20 degrees of visual angle. The smallest distance between two stimuli was about 1.425 degrees of visual angle. In contrast to Experiment 1, in Experiment 2 (and Experiment 3), a square grid was used instead of a concentric grid. This was to maximize comparability between the setup in these experiments and that in Buetti et al. (2016). In Buetti et al. (2016), it was shown that lures increased response times in a logarithmic manner as compared to the linear effect of candidates. It should be noted that Madison et al. (2018) compared performance across these two different grid arrangements and found the difference across grids on RTs to be fairly minimal. In the heterogenous displays, the number of high- and low-similarity candidates were always equal (i.e. 2 or 4 of each depending on the total candidate set size). In total, there were 12 different types of displays. Each participant observed a block of 480 fully randomized trials with 40 trials for each display type.

Each trial began with the presentation of a central fixation cross for 1 s before the display of the search array, which remained on screen until a response was made. Participants responded to the orientation of the target letter T by pressing either the left or right button on a keyboard. Feedback was given in the form of a loud beep whenever an error was made; no feedback was given for correct trials.

Results

Average accuracy was high ($M = 94.6\%$, $SD = 5.32\%$). There was no speed–accuracy trade-off (Table A1 in the Appendix).

Effect of Lures on Candidate-Homogeneous Displays

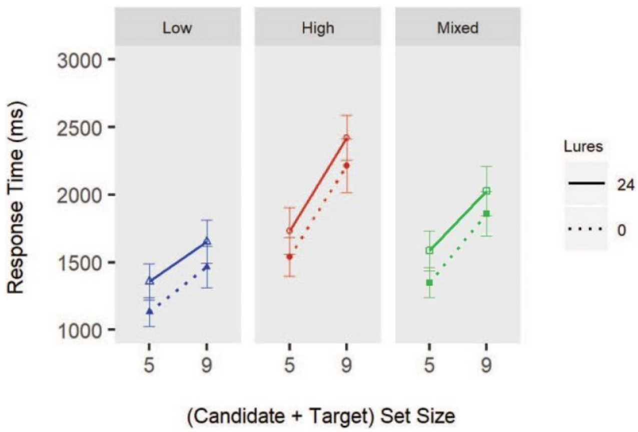
Trials with incorrect or no responses were excluded from analyses. We first conducted a 2 (lure presence) by 2 (candidate type) analysis of variance (ANOVA) on the observed search slopes in the candidate-homogenous displays. Presence of lures did not significantly affect the linear search slopes, $F(1, 24) = 0.057$, $p = .81$, $\eta_p^2 = .0024$. Linear search slopes were higher for high-similarity candidates (170 ms/item) compared to low-similarity candidates (79 ms/item), $F(1, 24) = 70.88$, $p < .001$, $\eta_p^2 = .75$. The large (more than double) difference between the linear search slopes for low- and high- similarity candidates confirmed that the two types of candidates differed greatly in terms of their similarity to the target. Finally, the interaction between candidate type and lure presence was not statistically significant, $F(1, 24) = 0.21$, $p = .65$, $\eta_p^2 = .0088$. The Bayes factor was computed to compare the null hypothesis "No effect of lure presence on search slopes" to the alternative hypothesis (nonzero effect of lure presence), using the BayesFactor package in R (Morey & Rouder, 2018). The data was more likely under the null, with moderate support, $BF_{01} = 6.34$. Thus, the presence of lures did not meaningfully affect the linear search slopes.

The same ANOVA was conducted on intercept values as the dependent variable. Intercept values were significantly increased by the presence of lures (928 vs. 708 ms), $F(1, 24) = 8.65$, $p = .00713$, $\eta_p^2 = .27$, but not candidate type, $F(1, 24) = 0.69$, $p = .41$, $\eta_p^2 = .028$. The interaction was not significant, $F(1, 24) = 0.33$, $p = .57$, $\eta_p^2 = .014$. In other words, there was a constant cost of 120 ms to process displays containing lures that was independent from the number of candidates and candidate-target similarity (see Figure 4), consistent with the predictions of the target contrast signal theory.

Attentional Scrutiny in Candidate-Heterogeneous Displays

Next, we turn to the main question of whether attentional scrutiny is prioritized as a function of target-distractor similarity or

Figure 4
Response Times (in ms) in Experiment 2



Note. The *x*-axis represents the number of candidates plus the target. Two observations are evident here. First, the presence of lures did not significantly affect search slopes (dotted vs. solid lines). Second, search slopes were steeper for high-similarity candidates compared to low-similarity candidates (red [gray] circles vs. blue [dark gray] triangles), while that for the mixed displays were in between (green [light gray] squares). Error bars indicate 95% confidence intervals. See the online article for the color version of this figure.

in fact random. To evaluate the ideal prioritization model and the random scrutiny model, we used response times from the candidate-homogeneous displays to predict response times in the candidate-heterogeneous displays. Four predictions, corresponding to the four different conditions (2 levels of candidate similarity X 2 levels of lure presence), were made. First, mean search slopes were calculated, for each subject and each condition, for the candidate-homogenous displays by fitting the equation:

$$RT_{high} = H \times (number\ of\ candidates + 1) + ch$$

$$RT_{low} = L \times (number\ of\ candidates + 1) + cl$$

H and L represent the search slopes (see Figure 4) for high- and low-similarity displays respectively, and c_h and c_l represent the intercepts for high- and low-similarity displays respectively. The term "+1" denotes that the functional set size is simply the number of candidates plus the target. The ideal prioritization model predicts that response times in heterogeneous displays would be dependent on only the number of high-similarity candidates plus the target. Thus, response times in heterogeneous displays were predicted using the following equation:

$$RT_{heterogenous} = H \times (set\ size_{high} + 1) + ch$$

The random scrutiny model predicts that response times in heterogeneous displays would be independent of candidate-target similarity. Thus, the functional set size in this condition is simply the sum of the number of all high and low candidates, plus the target (in other words, all the candidates that are present in the display, plus the target). Mathematically, this is equivalent to the average response times for high- and low-similarity displays, at any given functional set size:

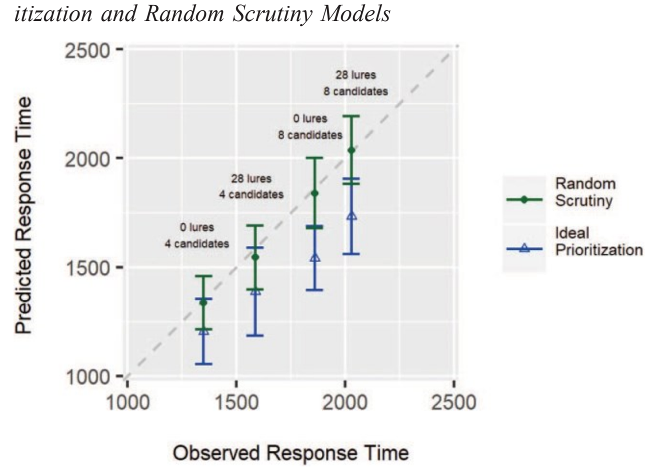
$$RT_{heterogenous} = \frac{RT_{high} + RT_{low}}{2}$$

Figure 5 shows that the ideal prioritization model systematically underpredicts response times (average deviation error 252 ms), while the random scrutiny model makes predictions that show near-perfect correspondence with the observed response times (average deviation error 16 ms). The dashed line (y x) indicates where the points would fall for a model that perfectly predicts response times with zero error.

Individual-level predictions also show the same pattern of results. Figure 6 shows the within-subject residuals (observed–predicted response times). Within-subject residuals from the ideal prioritization model (left panel) show large variability that increases with set size, indicating poor correspondence between the model and the observed data. In addition, the systematic underprediction by the model increases with set size. On the other hand, the residuals from the Random Scrutiny Model (right panel) are centered around zero and show little variability, suggesting good correspondence between the model and the observed data.

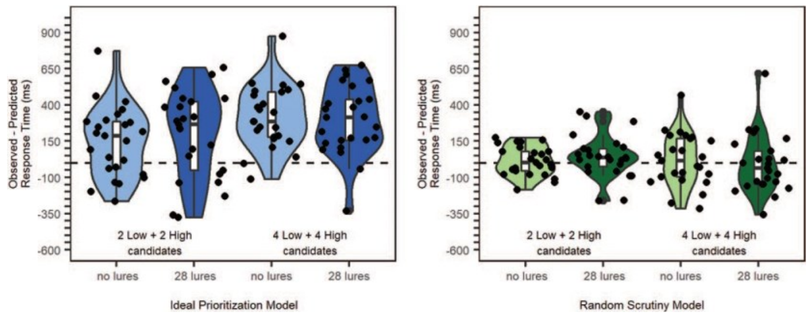
Next, we examined the performance of both models by quantifying the overall prediction error of each model. Traditional null hypothesis significance testing is problematic when the goal is to provide evidence for a null effect (zero prediction error by the Random Scrutiny Model). Thus, we calculated Bayes Factors in lieu of p values. Separate Bayes factors were calculated for each model, for a one-sample t test comparing the residuals (observed—predicted response times, shown in Figure 5) against zero. BF_{01} denotes evidence in favor of the null hypothesis (prediction error is not meaningfully different from zero), while BF_{10} denotes evidence in favor of the alternate hypothesis (prediction error is meaningfully different from zero). The Bayes Factors indicated

Figure 5
Observed Versus Predicted Response Times for the Ideal Prioritization and Random Scrutiny Models



Note. The Random Scrutiny model (green [gray] circles) makes near-perfect predictions, while the Ideal Prioritization model (blue [black] triangles) systematically underpredicts response times. Different conditions are indicated at the top of the figure. The dashed y x line indicates where the predicted values would fall on if predictions were perfect. Error bars indicate 95% confidence intervals of the observed means. See the online article for the color version of this figure.

Figure 6
Within-Subject Residuals Displayed in Violin Plots



Note. Each plot shows the observed minus predicted response times of all participants in the four predicted conditions (from left to right): two low similarity (2L) candidates plus two high-similarity (2H) candidates with no lures, two low plus two high-similarity candidates with 24 lures, four low plus four high-similarity candidates with no lures, and four low plus four high-similarity candidates with 24 lures. In the box-plots, the box indicates the interquartile range and the horizontal marker indicates the median. Each circle represents individual data points, while the shaded area shows the probability density of the data. Dashed lines indicate y 0 (no prediction error). *Left*: residuals from the ideal prioritization model show mean residuals that differ substantially from zero, with large variability, indicating poor correspondence between the predicted and observed data. *Right*: residuals from the random scrutiny model are centered around zero and show little variability, suggesting good correspondence between the random scrutiny model's predicted response times and the observed data. See the online article for the color version of this figure.

moderate evidence that the residuals from the random scrutiny model did not meaningfully differ from zero (BF_{01} 5.46), indicating that predictions from this model were almost perfect. On the other hand, there was strong evidence that the residuals from the ideal prioritization model were meaningfully different from zero, indicating poor correspondence between the model's predictions and the observed data (BF_{10} 2.15×10^{12}).

The ideal prioritization model represents the boundary case whereby the visual system perfectly prioritizes all high-similarity candidates before low-similarity candidates, in terms of what could be expected if prioritization was sufficiently adequate to clearly separate low- and high-similarity candidates. This is likely too extreme, given that the visual system is inherently noisy, but it still provides us with a lower boundary for best RT performance. In contrast, the random scrutiny model can be seen as an upper boundary for how slow RT performance can be expected to be. We can then quantify the degree of prioritization by a ratio of two difference scores: the difference between the observed RT and the predicted RT by the random scrutiny model, divided by the difference between the predicted ideal prioritization RT and the predicted random scrutiny RT. In other words,

$$\text{Prioritization Score} = \frac{RT_{RS} - RT_{Obs}}{RT_{RS} - RT_{IP}}$$

where RT_{Obs} is the observed response time, and RT_{RS} and RT_{IP} are the predicted response times for the random scrutiny and ideal prioritization models respectively. If prioritization were perfect, this ratio would be 1. This is because RT_{Obs} would be equal to RT_{IP} (i.e. the observed response times would be equal to the response times predicted by ideal prioritization model) and therefore, the terms in the numerator and denominator will be identical. If there were no prioritization at all (i.e. scrutiny is completely random), this ratio would be zero. This is because RT_{Obs} would be equal to RT_{RS} (i.e. the observed response times would be equal to the response times predicted by the random scrutiny) and therefore, the numerator would be zero. Finally, if RT_{Obs} is systematically larger than RT_{RS}, this would indicate that participants are taking even longer to respond than they would if they visited all the items (in random order). In other words, systematically negative values would indicate that participants are revisiting previously inspected candidates. The grand means, averaged across conditions and participants, were calculated for RT_{Obs}, RT_{RS}, and RT_{IP} to yield a prioritization score of 0.073, suggesting that there was minimal prioritization, if any, of candidates based on their similarity to the target.

Experiment 3

Experiment 2 revealed two main findings. First, the presence of lures slows down search, but it does so without impacting the search rate through the candidates (corroborating Buetti et al.'s, 2016 findings). More importantly, there was little evidence that attentional scrutiny of candidates was prioritized based on candidate-target similarity despite large differences in terms of candidate-target similarity (as indexed by large differences in search slopes and also supported by the discrimination data from Experiment 1). Instead, there was more evidence for random scrutiny. In this experiment, we sought to replicate these findings using a target detection task.

Method

This experiment was preregistered on Open Science Framework (https://osf.io/8rkny/). The data and materials can be found at https://osf.io/5n2rt/.

Participants

All participants were recruited from the same subject pool and did not take part in Experiments 1 and 2. As described in the preregistration, we planned on a sample size of 25 participants, which would be more than sufficient to measure the difference between the two candidate search slopes with 95% power and α .05. This corresponded to the main effect of candidate similarity in Experiment 2 (η_p^2 .747). Although the required sample size was determined to be 6, we decided to increase it to 25 to reduce noise and to keep the sample size consistent with Experiment 2, allowing better comparability in terms of data precision. In total, 33 participants were recruited (23 Females, mean age 19.4). Due to a computer error, there were no data from 2 participants. Of the remaining 31 participants, 10 had accuracy rates lower than 90%, which was our initial accuracy exclusion criterion. We thus lowered this criterion to 85%, as described in the preregistration, to minimize data loss. We then analyzed the data from the first 25 subjects that met the 85% accuracy inclusion criteria.

Design and Procedure

The stimulus and apparatus were identical to Experiment 2 with the exception that, in Experiment 3, displays only contained candidates and no lures. The task was to report the presence or absence of the target by pressing either the left or right arrow key. The assignment of response buttons was randomized between participants. There were three independent variables: candidate display type (high-similarity, low-similarity, mixed-similarity), total candidate set size (4 or 8), and target presence (present or absent). All other aspects of the design and procedure were identical to Experiment 2.

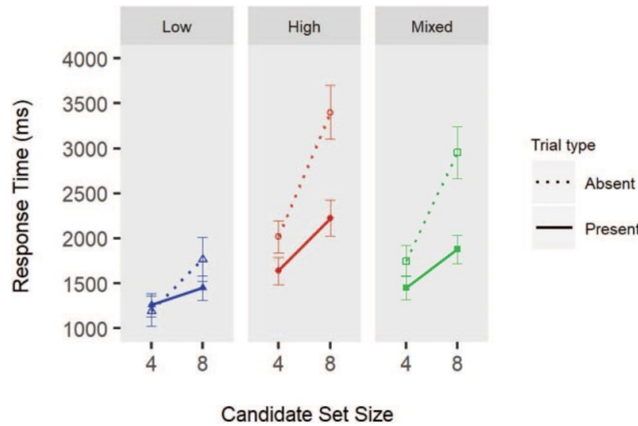
Results

Accuracy was high overall (M 93.5%, SD 0.03%). There was no speed–accuracy trade-off (see Table A2 in the Appendix).

A 2 (candidate similarity) by 2 (target presence) ANOVA was conducted on the search slopes in homogenous displays. Linear search slopes were significantly steeper for high-similarity displays (246 ms/item) compared to low-similarity displays (95 ms/item), $F(1, 24)$ 103.15, $p < .001$, η_p^2 .81. Linear search slopes were also significantly steeper on target-absent trials (244 ms/item) compared to target-present trials (97 ms/item), $F(1, 24)$ 154.70, $p < .001$, η_p^2 .87. The interaction between candidate type and target presence was significant, $F(1, 24)$ 18.65, $p <$ .001, η_p^2 .44. These results are illustrated in Figure 7. Target-absent slopes in high-similarity displays were 2.35 times that of target-present displays while this ratio was 3.02 in the low-similarity displays, suggesting that quitting rules could be influenced by target-distractor similarity. Importantly, the fact that the ratio of the target-absent-to-present search slopes were at least 2:1 indicated that the search processes through the candidate stimuli were inefficient both in this experiment as well as in Experiment 2 where identical stimuli were used and that self-terminating quitting rules halted search on target-present displays (Treisman & Gelade, 1980; Wolfe, 1998; Wolfe et al., 2010).

Figure 7

Mean Response Times (in ms) in Experiment 3



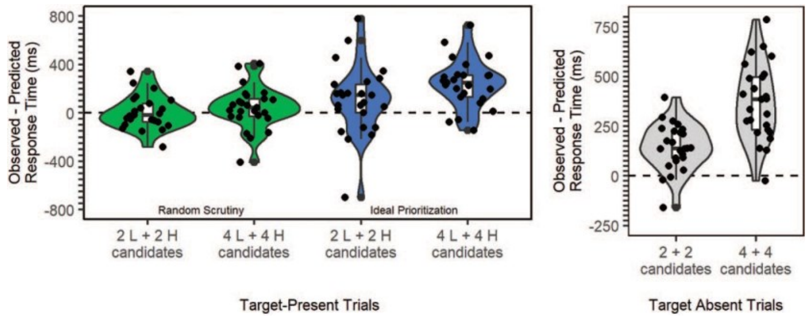
Note. The *x*-axis represents the number of candidates. Two observations are evident here. Search slopes were steeper on target-absent trials (dotted vs. solid lines). Second, search slopes were steeper for high-similarity candidates compared to low-similarity candidates (red [gray] circles vs. blue [black] triangles). The green (light gray) squares represent the observed response times for the heterogeneous displays. Error bars indicate 95% confidence intervals. See the online article for the color version of this figure.

Next, as in Experiment 2, we compared the Random Scrutiny Model with the Ideal Prioritization Model. Predicted response times were calculated with the same method described in Experiment 2. Note that model comparison was done using only the target-present data, since both models make the same prediction for target-absent trials (all items would be scrutinized before the observer decides to quit the search). Figure 8 shows the within-subject residuals (observed–predicted response times). Experiment 3 replicated the results from Experiment 2. There was moderate evidence that the within-subject residuals from the random scrutiny model did not differ meaningfully from zero (BF_{01} 4.02). The residuals also showed little variability. Overall, this model produced near-perfect predictions. In contrast, the within-subject residuals from the ideal prioritization model were meaningfully different from zero (BF_{10} 1504.89) and increased with set size, indicating poor predictive power of the model. The prioritization score was −0.15, again indicating minimal prioritization and perhaps a small tendency for revisiting already inspected candidates.

Finally, we also analyzed target-absent trials. Although these data do not differentiate between the two models (in both cases it is expected that observers will not quit until after having scrutinized all items in the display), they revealed an unexpected finding. Target-absent response times in candidate-heterogeneous displays were much longer than what would be predicted based on target-absent response times observed in candidate-homogeneous displays, and the residuals increased with set size.

Figure 8

Within-Subject Residuals, for Experiment 3, Displayed in Violin Plots



Note. Each violin plot shows the observed minus predicted response times for all participants in the two predicted conditions: 2 low similarity plus 2 high-similarity candidates or 4 of each. In the box-plots, the box indicates the interquartile range and the horizontal marker indicates the median. Each circle represents individual data points, while the shaded area shows the probability density of the data. Dashed lines indicate y 0 (no prediction error). *Left*: residuals from target-present trials. Residuals from the random scrutiny model (green [gray]) showed little variability around 0, suggesting good correspondence with the model. On the other hand, residuals from the Ideal Prioritization Model (blue [dark gray]) were not centered around zero and showed large variability, indicating poor correspondence between the predicted and observed data. *Right*: residuals from target-absent trials. Both models make the same predictions for target-absent trials, since both predict that all items will be scrutinized before the observer quits the search. There was a systematic underprediction, suggesting that candidate heterogeneity lengthened the quitting rule above and beyond what would be predicted based on candidate homogeneous displays. See the online article for the color version of this figure.

Discussion

Overall, Experiments 2 and 3 both provided strong support for the random scrutiny model. In spite of large differences in terms of candidate-target similarity (as indexed by search slopes), there was no evidence that observers prioritized high-similarity candidates during attentional scrutiny. Although the ideal prioritization model represented an "ideal" boundary scenario for perfect prioritization, the prioritization score could have measured any degree of prioritization from 0% to 100%, if any were present. Not even modest amounts of prioritization were observed.

Interestingly, both the random scrutiny and ideal prioritization models underpredicted target-absent response times in candidate-heterogenous displays. There are two potential explanations for this observation. First, it is known that distractor heterogeneity increases response times and changes the rate of evidence accumulation (Duncan & Humphreys, 1989; Lleras et al., 2019). Second, and more likely, it could be that candidate heterogeneity impacted the quitting rule in inefficient search tasks by several hundreds of milliseconds, perhaps as a result of revisiting previously rejected candidates before the observer can be confident of a target-absent response. This second explanation appears to be more likely. Although it is possible that the search slopes measured in the candidate-homogenous displays do not accurately reflect the rate of search in candidate-heterogenous displays, this is unlikely given the data from target-present trials. If search slopes measured in candidate-homogeneous displays were inaccurate or did not reflect the search rate through heterogeneous displays, the random scrutiny model (based on those homogeneous search slopes) would have failed to predict performance in the target-present heterogeneous conditions. It is evident from Figure 8 that the residuals for the target-present predictions are smaller and less variable, while that for the target-absent trials were more variable and increased with set size. Thus, it is unlikely that the underprediction stems from an erroneous measurement of search slopes, but rather a result of revisitations to previously inspected candidates. Thus, the results suggest that candidate heterogeneity impacts a nonvisual process in search. It might be that it increases the noise in the memory representations of locations that have been inspected or that it decreases the amount of locations that are remembered. This phenomenon deserves further study, and these hypotheses could be tested by monitoring eye movements of participants as they complete a candidate heterogeneous search task.

General Discussion

Many models of visual search propose that items are scrutinized by attention using some form of similarity-based prioritization, whereby to-be-scrutinized items are grossly ordered in terms of their similarity to the target, from highest to lowest. Attention and/or eye movements then simply visits these items by moving down that list (Ehinger et al., 2009; Najemnik & Geisler, 2005, 2008; Navalpakkam & Itti, 2005, 2007; Rao et al., 2002; Wolfe, 2006; Zelinsky, 2008). These prioritization accounts make two main predictions that the present study demonstrated were incorrect. First, according to these accounts, distractors that are very different from the target (lures) ought to almost never impact search times since attention would not visit these items due to their very low priority. However, the results from Experiment 2 showed that the presence of lures added a cost to overall processing times, in line with results from Buetti et al. (2016). Not only do these items contribute to search times, prior results also showed that the costs to reject lure items from consideration are systematically related to lure-target similarity (Buetti et al., 2016; Lleras et al., 2019; Ng et al., 2018; Wang et al., 2017). Second, similarity-based prioritization accounts propose that distractors that are relatively similar to the target will be scrutinized according to their similarity to the target, with high similarity items having a larger attentional priority. In contrast to these predictions, the results from Experiments 2 and 3 showed that observers inspected high- and low-similarity candidates in a random order, or at least in an order that was not based on similarity-ratings to the target template. This was observed in spite of large differences in discriminability between low- and high-similarity candidates (as indexed by substantial differences in search slopes in candidate-homogeneous displays) and in spite of the fact that observers could reliably differentiate between the two candidates (Experiment 1).

In sum, the current results suggest that attentional prioritization does not work in the manner than many theorists propose. There is an initial stage of distributed attention during which items that are sufficiently dissimilar from the target are discarded via peripheral visual analysis. It takes time to reject these dissimilar items, with more similar items taking longer. Once peripheral analysis has discarded these unlikely targets, focused attention is deployed to the nonrejected locations in an order that is not determined by target-distractor similarity. This makes sense given that peripheral analysis suffers from severe computational limitations, thus, it is difficult for peripheral vision to produce an orderly ranking of candidate items. From this perspective, distractor rejection starts at the bottom of the similarity scale, not at the top, as is often understood. We refer to this orderly rejection of distractors as being bottom-to-top.

The proposal that there is a bottom-to-top attentional prioritization in visual search is consistent with what has been observed previously with displays intermixing lures and candidates in a more systematic fashion (Buetti et al., 2016; Ng et al., 2019). The results from these studies showed that when candidates and lures were both present in a display, there was a time cost associated with rejecting lures that was independent from the time spent searching through the candidates. This was reflected by search times increasing logarithmically with lure set size while search efficiency through candidates remained constant as lure set size varied, indicating that candidate scrutiny was occurring after lure rejection. Bottom-to-top attentional prioritization is also consistent with results on search with lure-heterogeneous displays (Lleras et al., 2019; Wang et al., 2017). When multiple types of lures were present in a display, the lures that were most dissimilar were rejected earlier and the less dissimilar lures took longer to be rejected (see Figure 4), with all lures present in the display contributing to performance.

How Are Candidates Inspected?

The target contrast signal theory proposes that the output of the dissimilarity-based parallel rejection process responsible for rejecting lures is a list of locations of the remaining items (candidates). This list does not contain a precise visual description of these unrejected items precisely because of the resolution limits of peripheral vision during parallel processing (Rosenholtz, Huang, & Ehinger, 2012; Rosenholtz, Huang, Raj, et al., 2012). As such, candidates are neither ordered by contrast values nor indexed by target-distractor similarity. Focused attention will thus visit these locations in a random order without being biased by the similarity relation between the remaining items and the target. Importantly, Experiment 1 demonstrated that participants could reliably differentiate (although not perfectly) between low- and high-similarity candidates. In addition, the search slopes for displays containing high-similarity candidates were much higher than the slopes for displays containing low-similarity candidates, further confirming that the two differed in their similarity to the target. Despite this, Experiments 2 and 3 demonstrated that, in a search display containing multiple types of candidates, the degree of reliability of this candidate evaluation was not sufficiently adequate to be trusted as a source for attentional guidance, or, alternatively, the effort required to use this unreliable information may be too great. Attention was thus deployed in a random manner rather than being guided by target-distractor similarity.

Admittedly, other factors might come into play during attentional scrutiny. For instance, attention (and/or eye movements) might be deployed to whichever target-likely location is closest to current fixation or to a midpoint between target-likely locations (Zelinsky, 2012). Or, participants might use systematic scanning strategies (top-to-bottom, left-to-right, etc.) to visit all nonrejected locations. Aside from non-similarity-based scrutiny strategies, it is possible that there could be some form of imperfect similarity-based prioritization where only some of the high-similarity candidates are prioritized. However, given the near-perfect prediction of the random scrutiny model, this seems unlikely. Furthermore, as indicated by the overall prioritization score, a top-to-bottom prioritization based on candidate-target similarity is infrequent at best.

We should note that attentional prioritization of a subset of candidates is possible under certain circumstances. For instance, memory of previously seen search displays can guide attention during search. Response times are typically faster for targets that appear in search displays that have previously been presented to the observer compared to completely novel displays (Chun, 2000). This phenomenon is known as Contextual Cueing, and has been observed with different repeated contexts, including spatial layout (Chun & Jiang, 1998), identity (Chun & Jiang, 1999; Goujon et al., 2007), as well as the motion trajectory (Chun & Jiang, 1999) of the search items. Memory from the repetition of context leads to a prioritization of attention toward locations where the target is likely to be found (Chun & Jiang, 1998; Goujon et al., 2007, but see Annac et al., 2019). The repetition of scene layouts does seem to guide the deployment of attention (Geyer et al., 2010; Johnson et al., 2007), which raises the possibility that scene-based spatial memories increase the conspicuity (or priority) of the target and its immediate candidate neighbors. Such memory traces can clearly aid attentional prioritization, and they can do so quite quickly, starting 100 ms post display onset (e.g., Chaumon et al., 2008; Conci et al., 2019). This makes it possible that memories of spatial layout boost early display segmentation processes, prioritizing a subset of candidates over another. Interestingly, while the repetition of candidate-stimuli context (either over the entire scene or just over the area immediately surrounding the target) prioritizes attentional deployment to specific regions in the scene, repetition of lure-stimuli context does not. Although the presence of lures slows down RTs, repetition of their spatial context does not produce a prioritization signal to guide attention faster toward the target location (Ng et al., 2019; but see Geyer et al., 2010). Finally, rewards can also influence the prioritization of locations during attentional scrutiny. In studies that examine the influence of reward on search behavior, search items or locations are assigned different rewards for finding the target. Typically, search items that are assigned higher rewards are found more quickly and with greater accuracy (Won & Leber, 2016). In addition, eye movements (and thus overt attention) have been observed to prioritize rewarded locations in visual search tasks (Eckstein et al., 2015; Liston & Stone, 2008). In search with multiple targets, observers make more eye movements to high-reward targets compared to low-reward targets (Navalpakkam et al., 2010).

Why Does Lure Rejection Take Time?

In Experiment 2, the presence of lures slowed down search in a manner that was not affected by the number of candidates. As described above, this provided further support for one of the main proposals of target contrast signal theory: that peripheral vision initially considered all items in the display as potential targets. An alternate possibility is that the presence of lures may slow down performance for low-level reasons. That is, it is possible that the presence of lures might increase the local contrast of candidates and therefore their attentional pull. However, this is unlikely. If anything, a candidate surrounded by lures will have less (not more) contrast than when it sits by itself surrounded by a solid black background. Still, one might argue that the individual attentional pull of any given candidate increases when lures are present because it breaks down the otherwise homogeneity effect produced by a lot of very similar candidates (when no lures are present). Though this is possible, such a mechanism would predict that search through these more "attentionally-sticky" candidates would be slower than search through the candidates that have not been made more salient by lures. Yet, the present data, as well as previous data (Buetti et al., 2016; Ng et al., 2019), have shown that search efficiency through the candidates is identical with or without lures present in the display. For instance, in all search conditions in Experiment 2, candidate search functions with and without lures were parallel (see Figure 4).

The increase in response times in the presence of lures in Experiment 2 might be consistent with other models. In Itti and Koch's (2000) saliency model, a lure that is surrounded by candidates would have a higher saliency compared to a lure that is surrounded by other lures (or appearing in isolation), because in this model saliency reflects local feature contrasts. The increase in RT could thus be a result of additional shifts of attention triggered to these high-contrast lure locations, rather than by a process of rejecting lures, more generally. But, again, it is well known that the initial Itti and Koch model does poorly at predicting search performance in displays using simple geometric stimuli like ours (see Itti & Koch, 2000). If one considers attention-tuned versions of the saliency model that are meant to perform well in visual search tasks (e.g., Navalpakkam & Itti, 2005, 2007), these models would not predict an RT slow-down in the presence of lures. Indeed, the point of optimally tuning attention to the target features (Navalpakkam & Itti, 2005, 2007) is precisely to cut-off from possible examination items that do not contain features similar to the target.

The FLNN (Farthest-Labeled Nearest Neighbor) model may also be able to provide an explanation for why response times increased in the presence of lures (Avraham et al., 2008). According to the FLNN model, search starts with the random selection of an item in the search display. If this item turns out to not be the target, then attention selects another item that is most dissimilar from the currently selected item. If this item is still not the target, then the next item that is selected will be one that is the most dissimilar from all previously selected items. This process repeats until the target is selected. Thus, when lures are present in a display, it is possible that a lure will be selected at first, which would delay the eventual selection of the target, incurring some delay that would not exist if no lures were present. Furthermore, if a candidate is selected (by chance) after the first attention movement, the farthest neighbor would be a lure stimulus (as opposed to another candidate or the target itself) because lures are much more different from candidates than the target is from the candidates (by definition). Thus, the second attention movement would likely be directed toward a lure. As a result, in candidate-attended-first trials, the presence of lures would also result in longer RTs when lures are present compared to when they are not. It is important, however, to remember that the FLNN model was designed to predict accuracy under limited exposure durations. Some modifications would be needed to translate accuracy predictions into RT predictions in displays that are present until response. It would be interesting to see if the model could be adapted to account for the RT laws that we now know govern efficient search with fixed targets (e.g., logarithmic increases in RT as set size increases when participants are viewing lure-homogeneous displays with a fixed target in mind; the finding that these logarithmic slopes systematically vary as a function of lure-target similarity, see Buetti et al., 2016; Ng et al., 2018; Wang et al., 2018; and the heterogeneity search cost function, see Lleras et al., 2019; Wang et al., 2017). It is entirely possible that it might be able to capture these effects—we just do not know yet. That being said, the more critical theoretical contribution of the FLNN model is that it views selection in a fundamentally different way from similarity-based models, proposing that selection is guided by dissimilarity values instead. This focus on dissimilarity (as opposed to similarity) does make the model more in line with our target contrast signal theory than with more traditional similarity-based models of selection during search.

Does Candidate Heterogeneity Slow Down RTs?

The deviations from predictions observed on target-absent trials could be interpreted as being indicative of interitem interactions that impact how the visual system treats candidates in heterogeneous displays. In other words, because participants took so much longer to terminate target-absent trials, one could argue that this slow down indicates the visual system has a tougher time rejecting candidates when they appear among different candidates than when they appear by themselves in homogeneous conditions. It is indeed true that, generally speaking, distractor heterogeneity increases response times (Duncan & Humphreys, 1989; Lleras et al., 2019; Wang et al., 2017). But the mechanism by which this happens is unclear. For efficient search, our lab has demonstrated that this heterogeneity slowdown is likely the result of local interitem interactions that facilitate parallel rejection of lures when nearby lures are similar to one another (Lleras et al., 2019). Duncan and Humphreys (1989) argued that heterogeneity breaks down grouping effects (and groups can be rejected as a whole), so performance is worse in heterogeneous conditions because there are more items/groups to reject than under homogeneous conditions.

In theory, something along these lines could be happening (i.e. grouping of homogeneous candidates making candidate-homogeneous conditions easier than candidate-heterogeneous conditions), but it is unclear whether it happened in our experiments. Note that the visual processes involved in the rejection of a candidate are agnostic with regard to the presence or absence of the target. Indeed, if the observer already knows that the current trial is a target-present or target-absent trial, then there would be no need to search since the observer could already make their response. Thus, if there are heterogeneity effects in the target absent trials, the same ones ought to be present in the target present trials. However, no evidence of such heterogeneity effects were observed in target present trials: Performance in the candidate-homogeneous conditions (where candidate-heterogeneity effects are impossible) perfectly predicted performance in target-present trials in candidate heterogeneous conditions, across both set sizes, both in Experiment 2 (with and without lures) and in Experiment 3. Furthermore, it is also important to remember the placement of the stimuli in our experiments: Candidates were placed on the search grid such that they would be equally distributed across all four quadrants (so, one per quadrant when 4 were present and two per quadrant when 8 were present). Thus, in set size 5 (1 target, 2 Ls and 2 offset Ls), in three out of four quadrants, the candidates were by themselves (or sometimes accompanied by lures in Experiment 2). Given the size of each grid quadrant (10X10 degrees of visual angle), intercandidate interactions across quadrants and across such large spacing would be quite unlikely. It is also important to note that performance in the set size 9 condition, where such interactions (if they existed) would be possible within each quadrant, was exactly the same as performance in the set size 5 condition: We were equally successful at predicting performance across both set sizes across the two experiments. In sum, it is unlikely that there were candidate– candidate interactions between the two types of candidates in target-present trials, and by extension, in target-absent trials, in our experiments. What is more likely, we believe, is that candidate heterogeneity changes the quitting rule for target absent trials (likely a nonvisual process), inviting more revisitations (e.g., candidate heterogeneity might disrupt the memory representations of what items or locations have been already visited and rejected). Note that quitting rules in target-absent trials are notoriously difficult to understand, let alone predict (Cho & Chong, 2019; Chun & Wolfe, 1996; Fleck et al., 2010; Mitroff et al., 2015; Wolfe & Van Wert, 2010).

As discussed in the introduction, the ideal prioritization model represented (from the start) a lower boundary condition: how fast search could unfold in heterogeneous displays if participants were able to perfectly prioritize high-similarity candidates. Yet, as mentioned above, there are both empirical and theoretical reasons to have expected candidate heterogeneity to slow down performance, such that even if items were perfectly prioritized by similarity, actual performance on heterogeneous displays would have been slower than what would have been predicted based on high-similarity candidate homogeneous performance. The point of this ideal prioritization model was to serve as a boundary condition regarding how good performance could be in the heterogeneous condition. On the other extreme, there was the random scrutiny model: Performance should not be worse than this model because it is a model that does not care about candidates' similarity to the target. The experiments could have shown RTs somewhere in the middle between these two extremes: neither perfectly prioritized nor perfectly random. In that case, the RTs could have been the result of either some form of poor prioritization or (simply) a slowdown due to candidate heterogeneity. What we found instead, in six separate conditions (four predictions in Experiment 2 and two predictions in Experiment 3), is that RTs in the heterogeneous condition perfectly matched the RTs predicted by the random scrutiny model. So, although it is theoretically possible that there was a slow-down due to heterogeneity, it would be quite a coincidence that the heterogeneity slowdown was exactly of the correct magnitude to match the RTs predicted by the random scrutiny model across six separate conditions (and two separate groups of subjects). In our opinion, that is highly unlikely, but it is nonetheless possible.

Limitations and Future Directions

Eye Movements

In this study, eye movements were not measured. Thus, the observed results most likely arose from a combination of both overt and covert attentional processes. It is highly likely that overt attention, as a result of eye movements, were being measured here. Even in efficient search tasks, observers overwhelmingly tend to choose to move their eyes, even when the task can be completed more quickly and efficiently without eye movements (Ng et al., 2018; Zelinsky, 2008). Currently, target contrast signal theory does not differentiate between overt and covert attention at this time, and it is worth noting that voluntary deployment of covert attention takes about the same time as voluntary eye movements (~200 ms, see Wolfe et al., 2000). Nevertheless, this is an important avenue for future work since much of the theory relies on the differences between foveal versus peripheral processing, especially when determining which stimuli are candidates and which are lures.

Candidate Discriminability and Crowding

In Experiment 1, there was only one candidate among 35 lures. There was thus no visual crowding of the candidate by other candidates. This raises the concern that discriminability of candidates in Experiments 2 and 3 might thus have been poorer due to crowding by candidates because in those experiments, there were always several candidates present in the display at the same time. That said, we have some confidence in our results because performance was relatively high in Experiment 1 even under brief exposure times of 100 ms and also because in Experiments 2 and 3, the search display was constrained such that the nontarget candidates were distributed evenly between the four quadrants of the search display. Thus, at set size 5 (1 target and four candidates), the target was either by itself in its own quadrant (75% of trials) or, at most, with one additional candidate nearby. Overall, three out of the four quadrants only contained 1 candidate on each trial. At set size 9, most of the time, the target appeared alongside one additional candidate in its own quadrant (and in 25% of trials with just two candidates in the same quadrant). Three out of four quadrants contained just two candidates on every trial. Given the size of our displays (quadrants were about 10 degrees of visual angle in width height), the concerns about candidate– candidate crowding are therefore relatively minor. Thus, while candidate–candidate crowding was possible, it probably did not play too much of a role in terms of substantially lowering the discriminability of the target in the periphery in Experiments 2 and 3 compared to what was measured in Experiment 1.

Stimulus Dependency

Experiment 1 showed that the L and offset-L candidates could be reliably differentiated even at exposure times of 100 ms. We believe this indicates that this information was available to the visual system although it was not used to prioritize the deployment of attention in Experiments 2 and 3. However, it is possible that our results might be overly dependent on the specific types of stimuli we used. Therefore, it might be important to continue to test this hypothesis with more varied stimuli. It is indeed possible that there exist some sets of stimuli that are similar enough to the target template to be deemed as candidates, yet with sufficiently large differences in terms of their similarity to the target that the attentional system might be able to use differences in target-candidate similarity to prioritize candidates by similarity. It would be challenging, however, to create such a set of stimuli. Whether an item is a candidate or a lure depends on whether parallel processing can reject that item as a nontarget given the processing limitations of peripheral vision. Although this is mainly determined by the features of the stimulus, an important factor is also the location of the stimulus in the visual field. A stimulus could act like a candidate in the far periphery but act like a lure in the near periphery, where the resolution of parallel processing might be sufficient to reject it as a distractor. Thus, in such sets of stimuli, any observed prioritization would have to be carefully assessed so that it might not be confused with eccentricity effects (leading to better rejection of candidates near fixation).

It is also a limitation that candidate similarity was defined only in the shape dimension (candidates had the same color as the target). We could have run the same experiments defining candidate similarity in the color dimension (and keeping target and candidates shapes identical). There is often a sense that color is "special" in terms of its ability to guide attention. Therefore, one might be concerned that the results are unique to the way we designed our stimuli, perhaps because participants could have first tuned to color to reject lures and then tune to shape to try to tackle the candidates. This sequential tuning of attention to different feature dimensions might somehow impact how well prioritization can be achieved. However, recent work from our lab helps assuage these concerns. In Buetti et al. (2019), we demonstrated that when a target differs from lures across both color and shape, search is not "guided" first by color then by shape. Using stimuli very similar to the ones used here, Buetti et al. (2019) demonstrated that search for a target defined by a given color and shape is perfectly predicted by the degree of its color distinctiveness (evaluated when all shapes are identical, on a separate group of participants) and shape distinctiveness (evaluated with all colors are identical, also on a separate group of participants), simultaneously. This simultaneous and independent computation of color and shape distinctiveness occurs even when one feature dimension (say color) carries a much more distinctive signal than the other (shape). Performance on 90 different conditions (mixing different types of colors, shapes and set sizes) was almost perfectly predicted by this simultaneous color + shape guidance account. Therefore, in spite of intuitions that might suggest that in tasks such as the one we used here one feature (color) is prioritized more or before the other (shape), what actually happens is that the visual system is computing both color and shape differences simultaneously (maybe over different brain regions) and using a combined distinctiveness signal to "guide" attention (rejecting lures, direct attention toward candidates). Thus, given that both color and shape distinctiveness signals are computed simultaneously and combined together to guide attention irrespective of which feature dimensions carries a larger distinctiveness signal, we feel confident that the results would have been similar had we used color to define candidates (rather than shape).

In terms of generalizability, it would also be important to test our results with images that are more complex (e.g., photos of real-world stimuli) and also to explore presentation of these stimuli in more complex, realistic backgrounds (as opposed to black backgrounds). While we believe there is no clear reason why our results would not generalize well to those stimuli, the increased visual complexity of these images would move us closer to ecologically valid vision.

Conclusion

In a series of experiments, we demonstrated that the idea that attentional scrutiny prioritizes items in terms of decreasing target-distractor similarity is incorrect. Attention does not prioritize items in a top-to-bottom manner. For items that are potential targets (i.e. *candidates* in our terminology), attentional scrutiny occurs at random (or at least in a manner that is not ordered by target-distractor similarity). Furthermore, counter to the standard top-to-bottom prioritization account, processing items that are quite dissimilar to the target (i.e. lures) and therefore ought to never impact performance are in fact processed by (distributed) attention, resulting in systematic time costs involved in rejecting those items. These results are in line with the target contrast signal theory: Items are instead rejected in a bottom-to-top manner, in reverse order of their similarity to the target. This orderly rejection process continues up to the point where the visual system is unable to reject target-similar items with sufficiently high confidence because of limitations in peripheral processing. That being said, there are certainly other sources of attentional guidance that may play a role in directing attention to likely target locations that are not similarity-based. For example, contextual cueing, rewards, and top-down strategies have been shown to reduce search times (Chun & Jiang, 1998; Kristjánsson et al., 2010; Smilek et al., 2006). We propose that in the absence of such sources of information, attentional scrutiny is best characterized by a random process rather than one that involves a top-down similarity-based prioritization that starts with the most target-similar distractors and moves down the similarity scale.

References

Annac, E., Pointner, M., Khader, P. H., Müller, H. J., Zang, X., & Geyer, T. (2019). Recognition of incidentally learned visual search arrays is supported by fixational eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(12), 2147–2164. https://doi.org/10.1037/xlm0000702

Avraham, T., Yeshurun, Y., & Lindenbaum, M. (2008). Predicting visual search performance by quantifying stimuli similarities. *Journal of Vision*, *8*(4), Article 9. https://doi.org/10.1167/8.4.9

Balas, B., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision*, *9*(12), 13. https://doi.org/10.1167/9.12.13

Buetti, S., Cronin, D. A., Madison, A. M., Wang, Z., & Lleras, A. (2016). Towards a better understanding of parallel visual processing in human vision: Evidence for exhaustive analysis of visual information. *Journal of Experimental Psychology: General*, *145*(6), 672–707. https://doi.org/10.1037/xge0000163

Buetti, S., Xu, J., & Lleras, A. (2019). Predicting how color and shape combine in the human visual system to direct attention. *Scientific Reports*, *9*(1), 1–11. https://doi.org/10.1038/s41598-019-56238-9

Chaumon, M., Drouet, V., & Tallon-Baudry, C. (2008). Unconscious associative memory affects visual processing before 100 ms. *Journal of Vision*, *8*(3), 10. https://doi.org/10.1167/8.3.10

Cho, J., & Chong, S. C. (2019). Search termination when the target is absent: The prevalence of coarse processing and its intertrial influence. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(11), 1455–1469. https://doi.org/10.1037/xhp0000686

Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences*, *4*(5), 170–178. https://doi.org/10.1016/S1364-6613(00)01476-5

Chun, M. M., & Jiang, Y. V. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*(1), 28–71. https://doi.org/10.1006/cogp.1998.0681

Chun, M. M., & Jiang, Y. V. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science*, *10*(4), 360–365. https://doi.org/10.1111/1467-9280.00168

Chun, M. M., & Wolfe, J. M. (1996). Just say no: How are visual searches terminated when there is no target present? *Cognitive Psychology*, *30*(1), 39–78. https://doi.org/10.1006/cogp.1996.0002

Conci, M., Zinchenko, A., Töllner, T., Müller, H. J., & Geyer, T. (2019). Attentional (mis) guidance by a contextual memory template in early vision. *Journal of Vision*, *19*(10), 214a. https://doi.org/10.1167/19.10.214a

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

Eckstein, M. P., Schoonveld, W., Zhang, S., Mack, S. C., & Akbas, E. (2015). Optimal and human eye movements to clustered low value cues to increase decision rewards during search. *Vision Research*, *113*(Part B), 137–154. https://doi.org/10.1016/j.visres.2015.05.016

Ehinger, K. A., Hidalgo-Sotelo, B., Torralba, A., & Oliva, A. (2009). Modelling search for people in 900 scenes: A combined source model of eye guidance. *Visual Cognition*, *17*(6–7), 945–978. https://doi.org/10.1080/13506280902834720

Fleck, M. S., Samei, E., & Mitroff, S. R. (2010). Generalized "satisfaction of search": Adverse influences on dual-target search accuracy. *Journal of Experimental Psychology: Applied*, *16*(1), 60–71. https://doi.org/10.1037/a0018629

Geyer, T., Zehetleitner, M., & Müller, H. J. (2010). Contextual cueing of pop-out visual search: When context guides the deployment of attention. *Journal of Vision*, *10*(5), 20. https://doi.org/10.1167/10.5.20

Goujon, A., Didierjean, A., & Marmèche, E. (2007). Contextual cueing based on specific and categorical properties of the environment. *Visual Cognition*, *15*(3), 257–275. https://doi.org/10.1080/13506280600677744

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

Johnson, J. S., Woodman, G. F., Braun, E., & Luck, S. J. (2007). Implicit memory influences the allocation of attention in visual cortex. *Psychonomic Bulletin & Review*, *14*(5), 834 – 839. https://doi.org/10.3758/BF03194108

Kristjánsson, Á., Sigurjónsdóttir, Ó., & Driver, J. (2010). Fortune and reversals of fortune in visual search: Reward contingencies for pop-out targets affect search efficiency and target repetition effects. *Attention, Perception, & Psychophysics*, *72*(5), 1229–1236. https://doi.org/10.3758/APP.72.5.1229

Liston, D. B., & Stone, L. S. (2008). Effects of prior information and reward on oculomotor and perceptual choices. *The Journal of Neuroscience*, *28*(51), 13866–13875. https://doi.org/10.1523/JNEUROSCI.3120-08.2008

Lleras, A., Wang, Z., Madison, A., & Buetti, S. (2019). Predicting search performance in heterogeneous scenes: Quantifying the impact of homogeneity effects in efficient search. *Collabra*, *5*, 1–15. https://doi.org/10.1525/collabra.151

Lleras, A., Wang, Z., Ng, G. J. P., Ballew, K., Xu, J., & Buetti, S. (2020). A target contrast signal theory of parallel processing in goal-directed search. *Attention, Perception & Psychophysics*, *82*, 394–425. https://doi.org/10.3758/s13414-019-01928-9

Madison, A., Lleras, A., & Buetti, S. (2018). The role of crowding in parallel search: Peripheral pooling is not responsible for logarithmic efficiency in parallel search. *Attention, Perception, & Psychophysics*, *80*, 352–373. https://doi.org/10.3758/s13414-017-1441-3

McCarley, J. S., Wang, R. F., Kramer, A. F., Irwin, D. E., & Peterson, M. S. (2003). How much memory does oculomotor search have? *Psychological Science*, *14*(5), 422–426. https://doi.org/10.1111/1467-9280.01457

Mitroff, S. R., Biggs, A. T., Adamo, S. H., Dowd, E. W., Winkle, J., & Clark, K. (2015). What can 1 billion trials tell us about visual search? *Journal of Experimental Psychology: Human Perception and Performance*, *41*(1), 1–5. https://doi.org/10.1037/xhp0000012

Morey, R. D., & Rouder, R. N. (2018). BayesFactor: Computation of Bayes Factors for common designs (R package version 0.9.12–4.2) [Computer software]. https://CRAN.R-project.org/package BayesFactor

Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature*, *434*(7031), 387–391. https://doi.org/10.1038/nature03390

Najemnik, J., & Geisler, W. S. (2008). Eye movement statistics in humans are consistent with an optimal search strategy. *Journal of Vision*, *8*(3), 4. https://doi.org/10.1167/8.3.4

Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research*, *45*(2), 205–231. https://doi.org/10.1016/j.visres.2004.07.042

Navalpakkam, V., & Itti, L. (2007). Search goal tunes visual features optimally. *Neuron*, *53*(4), 605–617. https://doi.org/10.1016/j.neuron.2007.01.018

Navalpakkam, V., Koch, C., Rangel, A., & Perona, P. (2010). Optimal reward harvesting in complex perceptual environments. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(11), 5232–5237. https://doi.org/10.1073/pnas.0911972107

Neider, M. B., & Zelinsky, G. J. (2008). Exploring set size effects in scenes: Identifying the objects of search. *Visual Cognition*, *16*(1), 1–10. https://doi.org/10.1080/13506280701381691

Ng, G. J. P., Buetti, S., Dolcos, S., Dolcos, F., & Lleras, A. (2019). Distractor rejection in parallel search tasks takes time but does not benefit from context repetition. *Visual Cognition*, *27*(5-8), 609 –925. https://doi.org/10.1080/13506285.2019.1676353

Ng, G. J. P., Lleras, A., & Buetti, S. (2018). Fixed-target efficient search has logarithmic efficiency with and without eye movements. *Attention,* *Perception, & Psychophysics*, *80*(7), 1752–1762. https://doi.org/10.3758/s13414-018-1561-4

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Rao, R. P. N., Zelinsky, G. J., Hayhoe, M. M., & Ballard, D. H. (2002). Eye movements in iconic visual search. *Vision Research*, *42*(11), 1447–1463. https://doi.org/10.1016/S0042-6989(02)00040-8

R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rosenholtz, R., Huang, J., & Ehinger, K. A. (2012). Rethinking the role of top-down attention in vision: Effects attributable to a lossy representation in peripheral vision. *Frontiers in Psychology*, *3*, Article 13. https://doi.org/10.3389/fpsyg.2012.00013

Rosenholtz, R., Huang, J., Raj, A., Balas, B. J., & Ilie, L. (2012). A summary statistic representation in peripheral vision explains visual search. *Journal of Vision*, *12*(4), 14. https://doi.org/10.1167/12.4.14

Smilek, D., Enns, J. T., Eastwood, J., & Merikle, P. M. (2006). Relax! Cognitive strategy influences visual search. *Visual Cognition*, *14*(4 – 8), 543–564. https://doi.org/10.1080/13506280500193487

Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. CUP Archive.

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

Wang, Z., Buetti, S., & Lleras, A. (2017). Predicting search performance in heterogeneous visual search scenes with real-world objects. *Collabra*, *3*(1), Article 6. https://doi.org/10.1525/collabra.53

Wang, Z., Lleras, A., & Buetti, S. (2018). Parallel, exhaustive processing underlies logarithmic search functions: Visual search with cortical magnification. *Psychonomic Bulletin & Review*, *25*, 1343–1350. https://doi.org/10.3758/s13423-018-1466-1

Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, *1*(2), 202–238. https://doi.org/10.3758/BF03200774

Wolfe, J. M. (1998). What can 1 million trials tell us about visual search? *Psychological Science*, *9*(1), 33–39. https://doi.org/10.1111/1467-9280.00006

Wolfe, J. M. (2006). Guided search 4.0. *Integrated Models of Cognitive Systems*, *3*, 99–120. https://doi.org/10.1007/978-94-011-5698-1_30

Wolfe, J. M., Alvarez, G. A., & Horowitz, T. S. (2000). Attention is fast but volition is slow. *Nature*, *406*(6797), Article 691. https://doi.org/10.1038/35021132

Wolfe, J. M., Palmer, E. M., & Horowitz, T. S. (2010). Reaction time distributions constrain models of visual search. *Vision Research*, *50*(14), 1304–1311. https://doi.org/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*(2), 121–124. https://doi.org/10.1016/j.cub.2009.11.066

Won, B. Y., & Leber, A. B. (2016). How do magnitude and frequency of monetary reward guide visual search? *Attention, Perception, & Psychophysics*, *78*(5), 1221–1231. https://doi.org/10.3758/s13414-016-1154-z

Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological Review*, *115*(4), 787– 835. https://doi.org/10.1037/a0013118

Zelinsky, G. J. (2012). TAM: Explaining off-object fixations and central fixation tendencies as effects of population averaging during search. *Visual Cognition*, *20*(4–5), 515–545. https://doi.org/10.1080/13506285.2012.666577

(*Appendix follows*)

Appendix

Accuracy for Experiments 2 and 3

Table A1
Accuracy Broken Down by Condition for Experiment 2

Candidate similarity	Lures	Candidates	Mean accuracy (%)	Mean RT (ms)
Low	0	4	97.8	1132
		8	98.0	1465
	24	4	97.5	1357
		8	97.0	1652
High	0	4	94.6	1541
		8	89.9	2214
	24	4	94.3	1732
		8	90.7	2422
Mixed	0	4	95.3	1350
		8	93.5	1859
	24	4	94.7	1586
		8	92.5	2028

Note. Slower conditions also had lower accuracy, indicating that there were no speed–accuracy tradeoffs in the experiment.

Table A2
Accuracy Broken Down by Condition for Experiment 3

Candidate similarity	Trial type (target presence)	Candidates	Mean accuracy (%)	Mean RT (ms)
Low	Absent	4	98.9	1192
		8	99.4	1766
	Present	4	96.1	1260
		8	93.3	1450
High	Absent	4	95.6	2018
		8	91.9	3400
	Present	4	87.2	1638
		8	83.6	2226
Mixed	Absent	4	95.7	1750
		8	94.6	2954
	Present	4	91.6	1452
		8	89.8	1880

Note. Slower conditions also had lower accuracy, indicating that there were no speed–accuracy tradeoffs in the experiment.

Received December 10, 2019
Revision received October 2, 2020
Accepted October 6, 2020 ■