

ORIGINAL RESEARCH REPORT

Predicting Search Performance in Heterogeneous Scenes: Quantifying the Impact of Homogeneity Effects in Efficient Search

Alejandro Lleras, Zhiyuan Wang, Anna Madison and Simona Buetti

Recently, Wang, Buetti and Lleras (2017) developed an equation to predict search performance in heterogeneous visual search scenes (i.e., multiple types of non-target objects simultaneously present) based on parameters observed when participants perform search in homogeneous scenes (i.e., when all non-target objects are identical to one another). The equation was based on a computational model where every item in the display is processed with unlimited capacity and independently of one another, with the goal of determining whether the item is likely to be a target or not. The model was tested in two experiments using real-world objects. Here, we extend those findings by testing the predictive power of the equation to simpler objects. Further, we compare the model's performance under two stimulus arrangements: spatially-intermixed (items randomly placed around the scene) and spatially-segregated displays (identical items presented near each other). This comparison allowed us to isolate and quantify the facilitatory effect of processing displays that contain identical items (homogeneity facilitation), a factor that improves performance in visual search above-and-beyond target-distractor dissimilarity. The results suggest that homogeneity facilitation effects in search arise from local item-to-item interaction (rather than by rejecting items as "groups") and that the strength of those interactions might be determined by stimulus complexity (with simpler stimuli producing stronger interactions and thus, stronger homogeneity facilitation effects).

Keywords: visual search; heterogeneity effect; homogeneity facilitation; local interaction; independent processing

Introduction

The literature on visual search has, for the most part, overlooked the variability observed in efficient visual search experiments, that is, under conditions where the target appears to pop-out of the display and is overwhelmingly the first item selected by attention. Thus far, "efficient" visual searches have been characterized through the following rule of thumb: whenever the measured linear slope of the reaction time (RT) by set size function is less than 10 ms/item, one can describe the search as having been efficient, whereas slopes larger than 10 ms/item suggest more inefficient processing (e.g., Haslam, Porter & Rothschild, 2001; Wolfe & Horowitz, 2004). Perhaps the reason for not paying close attention to RT variability in efficient search was the idea that such small magnitude search slopes were not *meaningfully* different from zero. Thus, there was no reason to expect systematic variability in whatever featural processing of the scene was required to detect the pop-out. For example, in Wolfe's (1994)

famous Guided Search model, the time to encode and visually process the display was proposed to be constant across all search conditions.

Recently, Buetti, Cronin, Madison, Wang and Lleras (2016) demonstrated the existence of a key behavioral signature of efficient search: in search tasks where the target item is fixed, RTs increase logarithmically (rather than linearly) with set size. One key factor that determines the magnitude of the logarithmic slope is the similarity relationship between the target template and the non-target elements in the display: the more dissimilar non-target items are to the target, the smaller the corresponding logarithmic slope is. In contrast, in "inefficient" searches where the target item is fixed, RTs increase linearly (rather than logarithmically) with set size. Thus emerged a novel behavioral signature to differentiate efficient from inefficient searches in a more precise manner than just the approximate 10 ms/item rule of thumb.

Buetti et al. (2016) proposed the use of two terms to refer to distractor items in visual search: the term "lures" refers to non-target items that produce logarithmic search efficiency, whereas the term "candidates" refers to non-target items that produce linear search efficiency (Neider & Zelinsky, 2008, coined the term). The difference in terms

helps underline that different types of processing are at play when evaluating whether or not a distractor in a scene is the target. Buetti et al. proposed lures are items that are sufficiently different from the target such that these items can be evaluated and discarded in parallel across the entire scene. Because it is well known that peripheral vision has reduced representational fidelity (e.g., Strasburger, 2011 for a review), the visual differences between lures and the target must be sufficiently large for this efficient, parallel comparison of lures to target to take place with a high degree of success. In contrast, candidates are those objects in the scene that are sufficiently similar to the target that the visual system cannot confidently discard them as non-targets. This implies that either covert or overt spatial attention needs to be deployed to their locations so that a better resolution representation of the item is formed and the comparison to the target template is successful (with low errors, see Hulleman & Olivers, 2017).

The cognitive architecture proposed by Buetti et al. (2016) is straightforward. When observers are asked to look for a specific item in a scene, they construct a target template in their minds. When the scene appears, observers evaluate information at all locations containing an item, in parallel and with unlimited capacity, with the goal of making a binary decision: is this item likely to be a target or not? This parallel evaluation of information is a massive time-saving operation that leverages the vastly parallel architecture of the visual system to quickly reach high-confidence evaluations of items that are so different from the target that they do not require close scrutiny. This process therefore efficiently reduces the set of locations requiring focused, capacity-limited processing. Imagine the case where you are looking for a lawn chair in your garden. When you turn your eyes towards the garden scene, you can quickly arrive at the conclusion, in parallel, that the neither the shed, the trees nor any of the flowers are, in fact, your lawn chair. Locations that might require closer scrutiny might be those containing a wheelbarrow and other various types of lawn furniture.

As mentioned above, one of the key findings in Buetti et al. (2016) was that different lure items may require different amounts of time to be processed in this parallel evaluation stage. More precisely, lures that are relatively more similar to the target template require longer times to be discarded than lure items that are less similar to the target. This difference in average processing time can be measured experimentally. Imagine participants are asked to find a red triangle target. One can run a traditional efficient search experiment in which this target is presented amongst a varying number of (say) orange diamonds. As demonstrated through simulations, Buetti and colleagues found there is a one-to-one correspondence between the observed logarithmic slope of the RT by set size function and the average amount of evidence required to reject those orange diamonds as non-targets. Thus, the observed log slope can be used as an index of the lure-target similarity relationship, with higher log slopes indicating higher similarity (and correspondingly longer evaluation times).

In their initial model, Buetti et al. assumed perfect processing independence. That is, all evidence accumulators accrued evidence independently of one another. This aspect of the model allowed the authors to propose an important extension of their work: the model should predict RTs in novel scenes made up from heterogeneous mixes of various types of lures. More specifically, imagine having evaluated the log slopes corresponding to three different types of lures: blue circles, yellow triangles and orange diamonds (with low, medium and high levels of similarity to the red triangle target, respectively). Those log slopes are estimated using homogeneous search displays (i.e., when all lure items in the display are of one kind). Having done so, then one should be able to predict how long it should take an observer to find a red triangle target in any heterogeneous search display that simultaneously contains varying numbers of blue, yellow and orange lures. This follows because, if processing is truly independent, then when processing, say, a blue lure, the time to reach a decision about it should not be impacted by what other lures are present in the display.

Wang, Buetti, and Lleras (2017) conducted a study to examine this prediction of the Buetti et al.'s model. Wang and colleagues had three specific goals. The first goal was to extend the findings from Buetti et al. (2016) to real-world stimuli. The second goal aimed at finding, using computational simulations of their model, an equation to best predict processing times in heterogeneous displays based on logarithmic slope values estimated during processing of homogeneous displays. In addition to an equation inspired by the Buetti et al.'s model, three other equations were developed (each one associated with a specific cognitive architecture) and the performance of all four equations was compared to simulated processing times. The third goal was to test the prediction of the best equation on humans' data. In order to evaluate the log slopes for three different lure-target pairs, a behavioral experiment using only homogeneous displays was first conducted. Then, in a second experiment on a new group of participants, search performance was examined in heterogeneous displays containing various combinations of two or three different types of simultaneously presented lures. The four equations were then used to predict processing times in heterogeneous scenes in human data (based on the log slopes observed with homogeneous displays). The results indicated that Equation 1 (the equation associated with Buetti et al.'s model) was by far the best among the equations considered. This equation is described below:

$$RT = a + D_1 \ln(N_1 + N_2 + N_3 + 1) + (D_2 - D_1) \ln(N_2 + N_3 + 1) + (D_3 - D_2) \ln(N_3 + 1) \quad (1)$$

To formulate this equation, Wang et al. assumed that: (a) all lures are processed in parallel; (b) that evidence stops accumulating at a location once the decision threshold for that stimulus has been reached; (c) that evidence continues to accumulate at locations where decision thresholds have

not been reached. At the aggregate level, this means that lures with lower lure-target similarity will be rejected sooner than lures with higher lure-target similarity. This reduces the number of “active” accumulators over time. Going back to the example with blue, yellow and orange lures, with a red target, in this model, blue lures would be on average rejected first, then yellow lures, and finally orange lures (on average). In the equation above, there are three different types of lures in the scene, with $D_3 > D_2 > D_1 > 0$ being the three logarithmic slopes for each lure of type i ($i = 1, 2, 3$). And N_i is the number of lures of type i . Finally, a represents the RT to find and respond to a target when it is the only item present in the display. After a , the first term represents average time cost to process all 3 types of lures until the point where all items of type 1 are rejected (set by D_1). Then, evidence for lures of types 2 and 3 continues to accumulate. However, some evidence about these lures has already been accumulated (dictated by D_1). Thus, the second term represents the *additional* average time cost to reject lures of type 2, dictated by (D_2), while also continuing to accumulate evidence for lures of type 3, (hence the term $D_2 - D_1$), and so on. For more details, readers are directed to Wang et al. (2017).

The generalized equation when L lure types are present in the display is given by equation 2, where N_T is the total number of lures, and D_i parameters are organized from smallest (D_1) to largest (D_L):

$$RT = a + \sum_{j=1}^L (D_j - D_{j-1}) * \ln(N_T - (\sum_{i=1}^{j-1} N_i) + 1) \quad (2)$$

Importantly, using this equation and real-world objects, Wang et al. were able to account for 96.8 percent of the variance in the heterogeneous scenes experiment based on the log slopes measured in the homogeneous scenes experiment.

An unexpected aspect of the Wang et al. results was that predicted RTs systematically underpredicted observed RTs in heterogeneous scenes by a factor of about 1.3, such that $ObservedProcessingTimes = 1.3 \times PredictedProcessingTimes$ (where *Processing Time is the RT in any given condition minus the RT for the target-alone condition*). In other words, heterogeneous displays were processed with a logarithmic efficiency that was about 30% slower than the efficiency with which homogeneous displays were processed. Wang and colleagues proposed that this deviation from a ‘perfect’ prediction reflected a violation of one of the underlying assumptions in their model: processing independence. More specifically, the authors proposed that in homogeneous displays, there is a homogeneity facilitation effect, whereby nearby identical items interact with one another to increase the rate at which those items are processed and discarded (i.e., reducing the time required to discard those items). This inter-item facilitation effect is different from previous distractor-distractor suppression proposals in the literature. Duncan and Humphreys (1989) famously proposed that identical items group together such that the entire set of identical items can be rejected “as a group” rather than as individual

items. However, Equation 2 demonstrates that this form of rejection-by-group mechanism is unlikely because items are discarded on an item-by-item basis (i.e., each item contributes to overall RT). In fact, Equation 2 was formulated on the assumption that there is the same level of lure processing independence in homogeneous and heterogeneous displays. Thus, deviations from predicted RTs based on Equation 2 will allow us to *quantitatively* measure the degree of lure-to-lure interactions in homogeneous scenes. Thus, our approach represents a new contribution to the study of inter-item interactions in visual search by allowing us to quantitatively measure them, irrespective of other factors (like lure-target similarity).

The goal of the present study was two-fold. First, we sought to test again the predictive power of Equation 2 using the same simple geometric stimuli used in Buetti et al.’s Experiment 1A. The idea was to take the log slopes for blue circles, yellow triangles and orange diamonds reported in Experiment 1A from Buetti et al. (2016) and use these slopes in conjunction with Equation 2 to predict RTs in new experiments where new participants only see heterogeneous displays (various combinations of blue circles, yellow triangles and/or orange diamonds). Experiment 1 in the present study was designed with this goal in mind: we tested three separate groups of 20 participants (with each group experiencing 15 different combinations of lure items) and predicted their RTs using Equation 2. If the predictive power of equation 2 in Experiment 1 is confirmed, our second goal was to directly test whether the multiplicative deviations from Equation 2 can indeed be attributed to local lure-to-lure interactions or whether they reflect, more generally, the presence of multiple types of lures simultaneously in the same scene. Experiment 2 was designed to test this idea.

Experiment 1: Heterogeneous Search Using Buetti et al.’s (2016) stimuli

Methods

Participants. Three groups of students from the University of Illinois participated in exchange for course credit in a Psychology class. All participants gave written informed consent before participating. All participants reported normal or normal-to-corrected vision and were tested for color blindness using the Ishihara color plates prior to data collection. Data from the first twenty participants for each group who met the inclusion criteria were analyzed. A total of sixty participants were used for Experiment 1. A participant’s data were replaced if overall performance fell below 90% accuracy or when mean RTs were higher than 2 times the standard deviation of the mean RT for the group. This resulted in a total of three participants being replaced across the three groups of participants. Two participants’ data were replaced in the first group (Group A) due to high error rates (>10%), no participant data was replaced in the second group (Group B), and one participants’ data was replaced in the third group (Group C) due to high error rate (>10% errors). This experiment and the following Experiments 2A and 2B were approved by the University of Illinois Institutional Review Board. Sample size was

determined based on previous investigations in our lab (e.g., Buetti et al., 2016; Madison et al., 2018; Wang et al., 2017) showing that averaging the data of twenty subjects produces stable estimates of the group means for a given search condition.

Apparatus and stimuli. The experiment was programmed in Matlab with the Psychophysics Toolbox 3.0 (Brainard, 1997; Pelli, 1997) and was run on a 64-bit Windows 7 PC. Participants sat in a dimly lit room approximately 49 cm from a 20-inch CRT monitor (20 degrees of visual angle) at 85 Hz refresh rate with 1024 × 768 resolution.

Search displays were always heterogeneous, with a mixture of at least two different lure objects, sometimes three. One exception to this was the target only condition where the target was presented with no lure items. The target was a red triangle, which could face to the left or to the right, and was fixed throughout the experiment session. The lure stimuli were orange diamonds, which were two triangles placed side by side and had high similarity to the target (similar hue and shape), yellow triangles, which could face to the right or left and had moderate similarity to the target (dissimilar hue and similar shape), or blue circles, which had low similarity to the target (dissimilar hue and shape). This set of stimuli was the same set used in Buetti et al. (2016). All search objects subtended 0.833 degrees of visual angle on the display. Search items were randomly assigned positions on an invisible 6-by-6 rectangular grid array with jitter added on a black background. The distance between two adjacent positions on the grid array was on average 3.5 degrees. Example search displays for Experiment 1 are shown in **Figure 1**. As mentioned above, these are the exact same stimuli and stimulus arrangement used by Buetti et al. (2016).

Procedure. Each trial began with a brief presentation of a central fixation cross for duration of 1 second followed by a search array. The search array contained the fixation cross and appeared until the participant gave a response or for 5 seconds with no response and the trial was counted as an error. If the subject made an error, a short beep

followed the trial. There were 1.5 seconds between each trial. Participants were asked to locate the red triangle target and report the orientation (left facing or right facing). Participants responded with the pointer or middle finger of their right hand to report the target orientation by pressing on the left and right arrow keys, respectively.

Design. Reaction time performance was the main dependent variable of interest to compute the cost of display processing time. Varying the number of orange diamonds, yellow triangles and blue circles in the search display created two types of heterogeneous experimental conditions (2 mixed or 3 mixed) as specified in Table A1. To create the different stimulus conditions, for each group, we arbitrarily held constant the number of items of one lure type and then manipulated the number of items of the other two lure types (i.e., 0, 4, 8, 12, 20). See Table A1 for a list of all tested conditions. There were a total of 16 possible experimental conditions per group of participants, including the target only condition. Participants completed 50 trials of each experimental condition completing a total of 800 trials. Trial conditions were randomized for each subject, and breaks were provided every 50 trials.

Results and Discussion

Table A1 lists the correct RTs and average accuracy for each of the 45 conditions tested across the three groups.

Next, we used Equation 2 to predict RTs for each condition in each group. Note that for each group, the value of the variable a was determined by that group's mean RT in the target only condition.¹ **Figure 2A** plots the function of Observed RTs as a function of Predicted RTs, across all 45 heterogeneous conditions. The overall R^2 of the prediction was 0.899, demonstrating that, as in Wang et al. (2017), Equation 2 has great predictive power. In other words, search performance in heterogeneous search displays is very well predicted by performance observed in homogeneous search displays.

Note that as a comparison, we also plotted in **Figure 2B**, RT predictions from a second equation (also described in

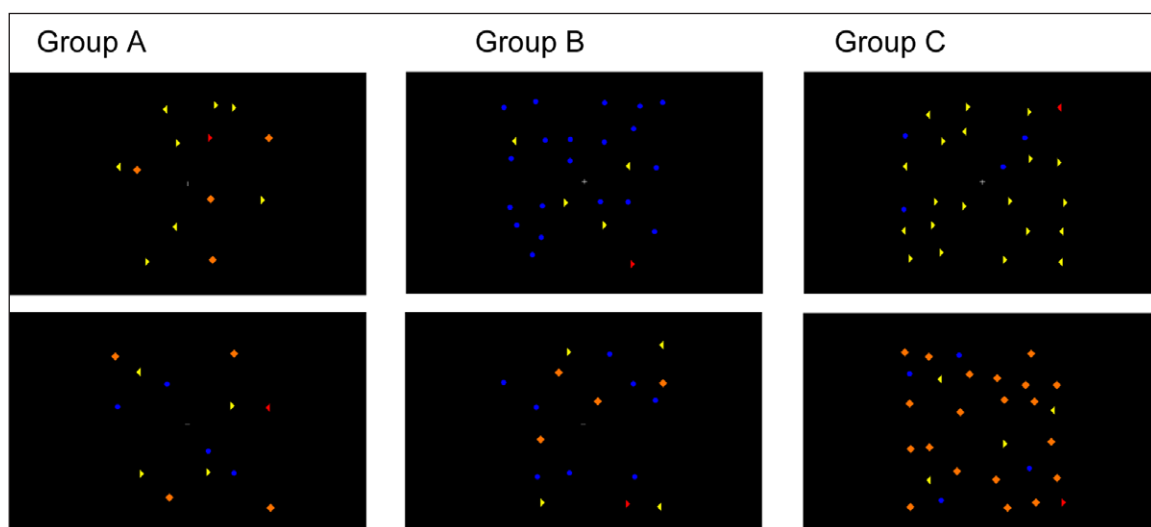


Figure 1: Examples search displays from **Groups A, B, C** are shown. The top row shows examples of the 2-mixed search display conditions and the bottom row shows 3-mixed search display conditions.

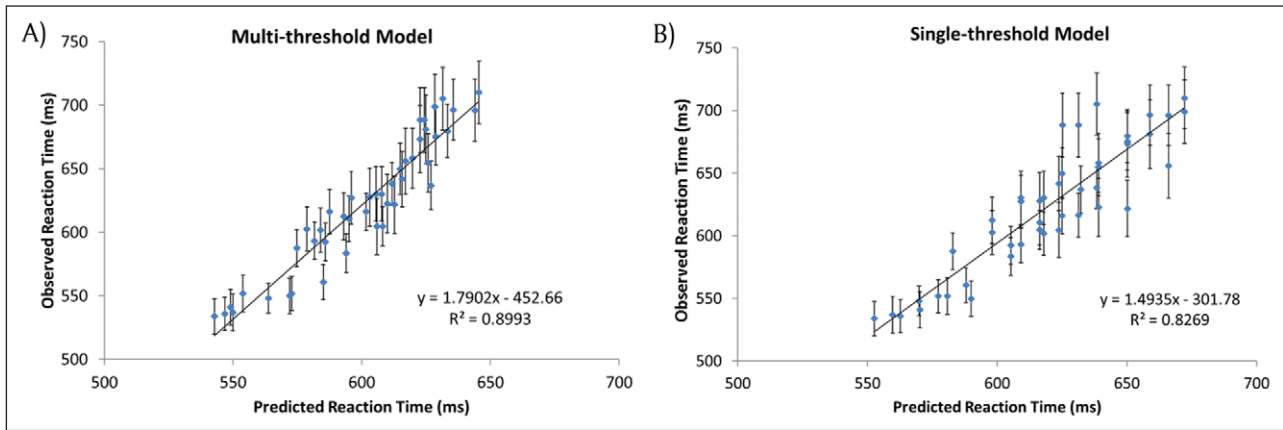


Figure 2: Observed Reaction Times across all 45 conditions as a function of the predicted Reaction Time for each condition. **Panel A** shows the predictions given by Equation 2 (multi-threshold model) and **Panel B** the predictions given by Equation 3 (single-threshold (Max) Model).

Wang, Buetti & Lleras, 2017), where it is assumed that a single rejection threshold is used for all items in a scene (inspired by single-threshold rejection models like Zelinsky's (2008) TAM and other signal detection theory based search models). The idea is that a single decision threshold is utilized across the entire scene to separate likely targets (i.e., signal in signal detection terms) from distractors (i.e., noise items in signal detection terms). In that case, the rejection threshold is set by the items that are the most similar to the target on any given display. Thus, the largest D_i value of the lures currently present in the display determines the efficiency with which the entire display is processed. For instance, if lures of types 1 and 2 are present, D_2 will determine the decision threshold for all items in the scene, whereas if lures of types 2 and 3 are present, D_3 will determine the decision threshold for all items in the scene. If all lures are present, it will be D_3 . Finally, if only lures of type i are present, its corresponding D_i would determine the decision threshold.

The corresponding equation is:

$$RT = a + \max\{D_i\} * \ln(N_T + 1) \quad (3)$$

The corresponding fit was visibly poorer ($R^2 = 0.827$). Using the Akaike Information Criterion (AIC) model comparison metric, we found that the multi-threshold model was 4.3×10^{16} more likely than the single-threshold model given the observed data. This result validates the results of Wang et al. (2017) with a new set of stimuli. A final model that we considered (and ruled out for subsequent analysis) was one where the slope of the log function is not given by the max of the D_i present in the display, but by the average of the D_i . The fit was even poorer ($R^2 = 0.774$, figure not shown).

Going back to Equation 1, the current results once again replicated the findings of Wang et al. (2017) in another manner: Equation 1 multiplicatively *underpredicted* RT performance: the magnitude of the effect was 1.79 (standard error = 0.077). That is, Observed RT = 1.79 * Predicted RTs – 453. Here, the multiplicative factor was numerically quite a bit larger than in Wang et al. (2017), where it was closer to 1.3 (standard error = 0.056).

According to Wang et al.'s proposal, the multiplicative factor reflects the magnitude of inter-item facilitative interactions that the lure stimuli exert on one another when they are used in homogeneous search displays. The fact the multiplicative factor obtained in Experiment 1 was larger than in Wang et al. suggests that this factor might vary as a function of stimulus complexity: relatively simple stimuli with few features like the ones used in Experiment 1 may interact with one another more easily than relatively more complex stimuli, like the real-world stimuli used in Wang et al. (2017).

Experiments 2A and 2B

Experiments 2A and 2B were designed to test the hypothesis that the multiplicative factor by which we underpredicted RTs in heterogeneous search displays reflects the extent of inter-item interactions present in homogeneous displays. If spatial proximity to identical elements is necessary for these inter-item interactions to be observed, then by spatially intermixing the different lure types as we have done so far, then we have (inadvertently) minimized these inter-item interactions in heterogeneous displays, both here and in Wang et al. (2017). It follows then that if we were to create heterogeneous displays where lures are spatially segregated by identity (i.e., all lures of type A would be near each other and all lures of type B would be near each other on a different location in the display), then inter-item facilitative interactions should re-emerge. If so, then Equation 2 should allow us to perfectly predict RTs in spatially segregated but heterogeneous displays (i.e., the multiplicative factor should be close to 1). In other words, heterogeneous spatially-segregated displays ought to be processed with the same efficiency as entirely homogeneous displays.

An alternative hypothesis is that the multiplicative factor represents not the inter-item interactions themselves, but rather, the presence of heterogeneity of the display (or conversely, the absence of homogeneity). That is, it is possible that when multiple types of lure items are present in a display, the displays are processed fundamentally different than displays where all lure items are identical to one another. Put the other way around, it is possible that

homogeneous displays afford a processing advantage that is absent whenever multiple types of lures are present in a scene. If so, whether lures are presented in a spatially intermixed manner or spatially segregated manner should matter little: in both cases, multiple types of lures are present in the display (increasing, in some abstract sense, the degree of scene complexity) and therefore, in both cases, Equation 2 should underpredict RTs, perhaps even to identical degrees as in spatially intermixed displays.

Experiments 2A and 2B differed only with regards to the stimulus used. In Experiment 2A, we used the same stimuli as in Experiment 1 (and in Buetti et al., 2016), simple colored shapes, whereas in Experiment 2B, we used the real-world stimuli used in Wang et al. (2017). Everything else about the two experiments was identical. It is important to note that even if the multiplicative factors observed in heterogeneous displays seemed to differ quite dramatically between stimulus sets (1.8 with colored shapes vs 1.3 with real-world stimuli), whatever magnitude of inter-item facilitation exists in homogeneous displays for a particular stimulus type, the same magnitude interactions ought to be observed in the spatially segregated displays used here. That is, we expect that in both cases, Experiments 2A and 2B, the multiplicative factor ought to be the same (1), in spite of the fact that the multiplicative factor was different in heterogeneous displays. Experiments 2A and 2B were both pre-registered in OSF, including a priori sample size and the predictions detailed above (osf.io/ce9fm and osf.io/dze26, respectively).

Methods

The methods for both experiments were identical except for the stimuli and background used in each experiment.

Participants. In both experiments, we recruited 20 subjects. This sample size was determined a priori (see preregistration) based on effect sizes observed in previous experiments. For both experiments, this sample size was determined to be sufficient to detect differences in the processing for the stimuli used in each experiment with at least 90% power. That is, this sample size was chosen because it allowed sufficient power to be sensitive to different processing time constants (i.e., log slopes) for the different types of stimuli used here. All participants were students at the University of Illinois and received course credit in a Psychology class in exchange for participation. All participants gave written informed consent before participating. All participants reported normal or normal-to-corrected vision and were tested for color blindness using the Ishihara color plates prior to data collection. For Experiment 2A, twenty-six subjects were actually run in this experiment because of scheduling. Three were discarded because they were unable to follow instructions and/or were feeling ill. The data from the first 20 subjects meeting inclusion criteria were analyzed (less than 10% errors, and mean RTs less than 2 standard deviations away from group mean). For Experiment 2B, twenty-three subjects were run, and data from the first 20 subjects meeting the same inclusion criteria were analyzed.

Apparatus and Stimuli. Stimuli were presented on a 20" CRT monitor at 85 Hz refresh rate and 1024*768 pixels resolution. Subjects sat in a dimly lit room at a viewing distance of 49 cm. The experiment was programmed with Psychtoolbox 3.0 on the MATLAB platform, and run on 64 bit Windows 7 OS PCs.

In **Experiment 2A**, the stimuli consisted of one target object (a red triangle pointing to the left or right) and three different types of lures, identical to the stimuli in Buetti et al. (2016), Experiment 1A. The three lure types were simple colored, geometric shapes: orange diamonds, yellow triangles, and blue circles. All of the items subtended 0.833 degrees of visual angle. The items were presented on the display based on an invisible 6 by 6 square grid which spans 20 degrees of visual angle vertically and horizontally. The minimum distance between two items, or two adjacent positions on the grid, is 3.5 degrees of visual angle. A white fixation cross was presented at the center of the screen and was 0.6 degrees of visual angle vertically and horizontally. All displays had a black background.

In **Experiment 2B**, the same real-world stimuli used in Wang et al. (2017) were used. The target was a teddy bear. The three lures were images of a humanoid doll in red dress, a reindeer with white fur, and a grey car. These stimuli have sizes of approximately 1.3 degrees visual angle horizontal and 1.7 degrees visual angle vertical. All items in the search displays are overlaid with a small red dot either on the right or on the left. All displays had a gray background, as in Wang et al. (2017).

Design: This was identical for both experiments. In each search display, there was always only one target, and two types of lure items with varying but equal numbers. The main independent variables are the combination of lure types presented on the display (since there were three different types of lures, there were three different types of pairings, A+B, A+C and B+C) and the number of each type of lure item (3 levels: 4, 8, 16) in the display. There was an equal number of each type of lure in the display, so both sides of the display were always equally populated by lures. We have also included a target-only condition, where the target was the only object presented on the display. We have controlled for other factors that could potentially influence search performance: where each lure type is presented (2 levels: on the left or on the right) and which visual hemifield the target appears in (2 levels), and which direction the target is oriented (2 levels: to the left or to the right). This yields $3*3*2*2*2+2*2 = 76$ different experimental conditions in total. We repeated each of these conditions 10 times to make the total trial number 760, which were randomly intermixed. We did not plan on studying specific lure location or target orientation, and collapsed across those variables in the analysis, yielding a total of 40 repetitions ($10*2*2$) per condition for the analysis. Sample displays are shown in **Figure 3**.

Procedure. Each trial began with a brief presentation of a central fixation cross for one second followed by a search array. The search array contained the fixation cross and appeared until the participant gave a response or for 5 seconds with no response and the trial was counted as an

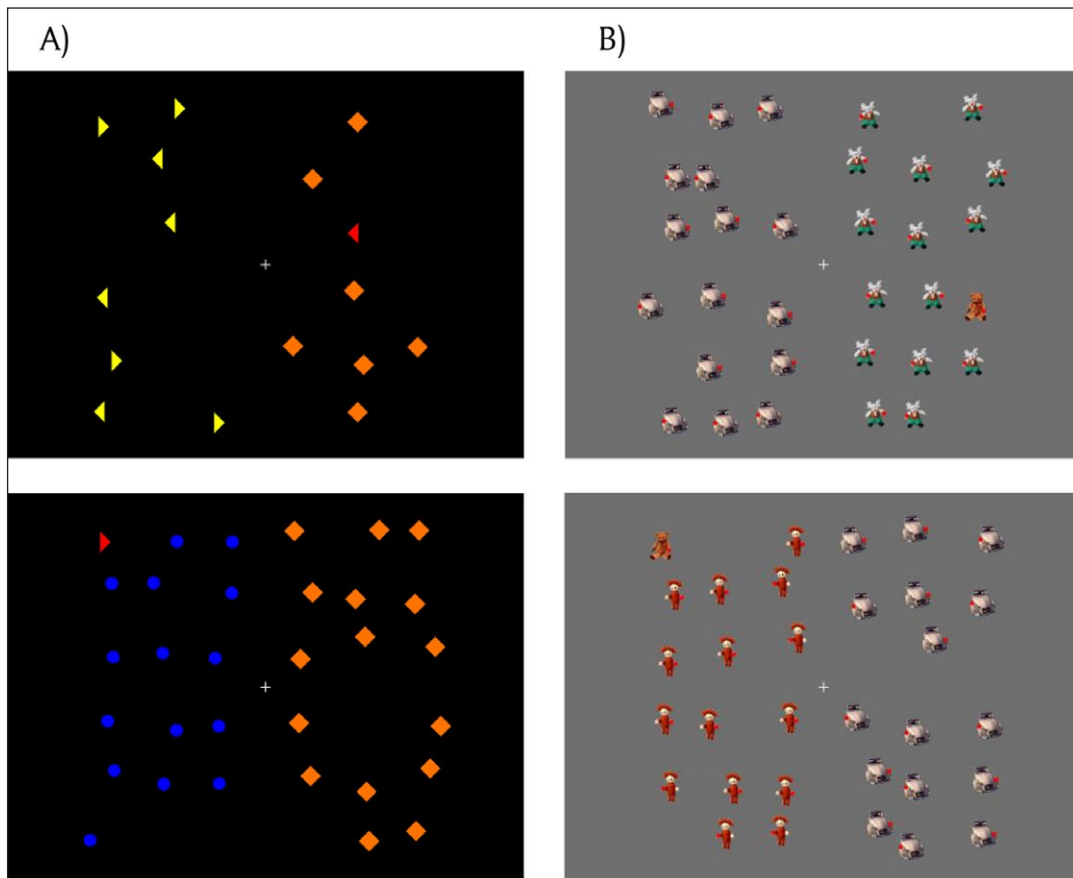


Figure 3: Sample displays used in Experiments 2A (left) and 2B (right).

error. If the subject made an error, a short beep followed the trial. There were 1.5 seconds between each trial. Participants were asked to report the target identity (e.g., the direction the red triangle pointed and the location of the red dot on the teddy bear hand, right or left). Participants responded with the index or middle finger of their right hand to report, using the left and right arrow keys, respectively.

Results

Experiment 2A (simple geometric colored shapes): Performance in the experiment was good overall, with an average accuracy of 98.9%. Table A2 shows the data for each group (lure-pair) condition and for each set size. **Figure 4A and B** plot the observed data as a function of the predicted RTs given by Equations 2 and 3 respectively.

As can be seen in **Figure 4**, Equation 2 produced a series of predictions that came very close to the actual observed data. The average prediction error across the 9 conditions was 8 ms, and the R^2 for the predictions was close to 0.92. Both of these measures were better than the ones obtained for Equation 3, which had a worse average prediction error (21 ms) and worse R^2 (0.85). Model comparison based on AIC indicated that the multi-threshold model (Equation 2) was 41.9 times more likely than the single-threshold model (Equation 3). In addition, a paired two-sample t test showed that prediction errors for Equation 2 were significantly smaller than those for Equation 3: $t(8) = -4.92$, $p = 0.001$. More interestingly, in contrast to Experiment 1, the best-fitting

line for the ObservedRT as a function of PredictedRT now shows a slope close to 1 (0.91, standard error = 0.096) rather than 1.8, much as if this time around, the equation had accurately captured the extent of lure-to-lure interactions present in the displays. In other words, given that the predictions were made based on log slope coefficients obtained from homogeneous displays, the current results suggest that the spatial segregated displays produce just as strong inter-item interactions as do perfectly homogeneous displays, in spite of the added complexity that arises from having two different sets of lures simultaneously present in the display. Experiment 2B therefore consisted of a test of this preliminary conclusion using real-world objects.

Experiment 2B: Performance in the experiment was also good overall, with an average accuracy of 98.5%. Table A3 shows the data for each group (lure-pair) condition and for each set size. **Figure 5A and B** plot the observed data as a function of the predicted RTs given by Equations 2 and 3 respectively.

In this instance, the R^2 produced by both models were not meaningfully different (0.977 and 0.980 for Equations 2 and 3 respectively). Relative likelihood of Equation 2 over Equation 3 based on AIC was found to be 1.98. That said, Equation 2 did produce much more accurate predictions overall: average prediction error was 8 ms compared to 20 ms for Equation 3. A paired two-sample t test showed that prediction errors for Equation 2 were significantly smaller than those for Equation 3: $t(8) = -2.97$, $p = 0.018$. Thus, in the context of these results as well as those of

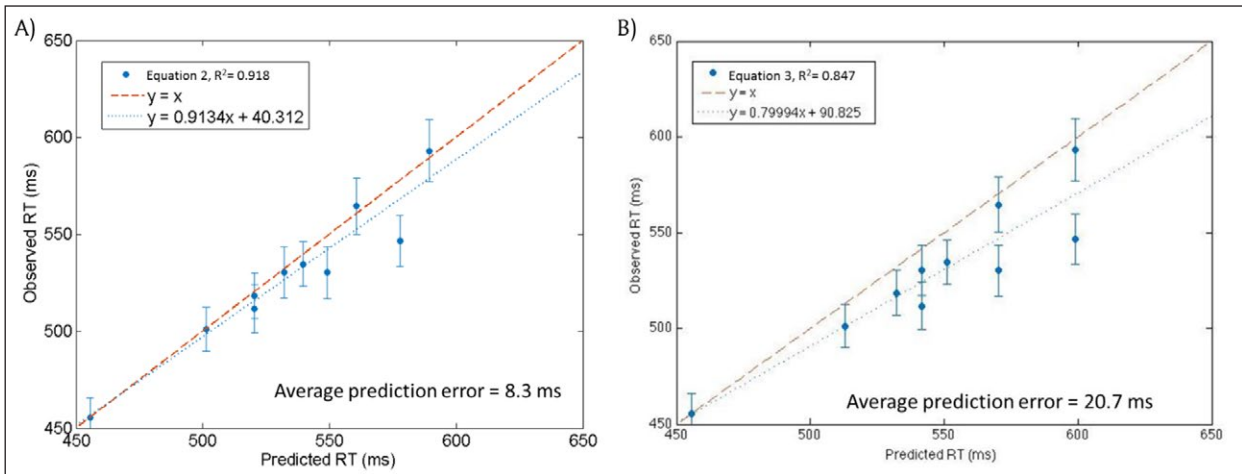


Figure 4: Observed Reaction Times across all 9 conditions of Experiment 2A (plus the target only condition) as a function of the predicted Reaction Time for each condition. **Panel A** (top) shows the predictions given by Equation 2 (multi-threshold model) and **Panel B** (bottom) the predictions given by Equation 3 (single-threshold (Max) Model). Also indicated is the average prediction error (in ms) across conditions for each model.

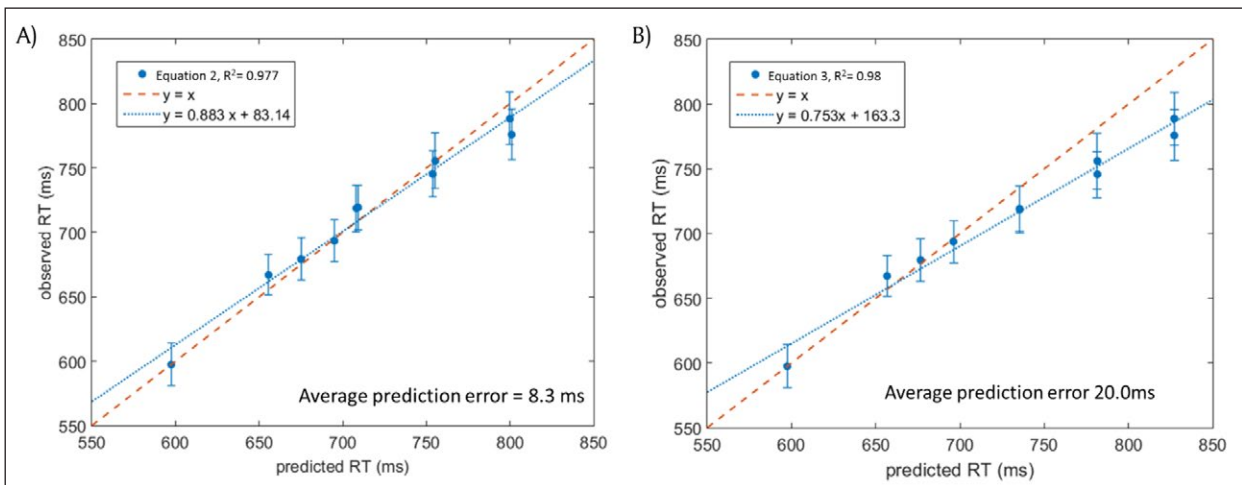


Figure 5: Observed Reaction Times across all 9 conditions (plus the target only condition) in Experiment 2B using real-world objects as a function of the predicted Reaction Time for each condition. **Panel A** (top) shows the predictions given by Equation 2 (multi-threshold model) and **Panel B** (bottom) the predictions given by Equation 3 (single-threshold (Max) Model). Also indicated is the average prediction error (in ms) across conditions for each model.

Experiment 2A and those in Wang et al. (2017), we conclude that Equation 2 outperformed Equation 3. Finally, with regards to Equation 2’s predictions, the best-fitting line for Observed RT as a function of Predicted RT now shows a slope close to 1 (estimate = 0.88, standard error = 0.048) rather than 1.3. This result also confirms the results of Experiment 2A: when heterogeneous, but spatially segregated displays are used, our model can predict RT with a high degree of accuracy both in terms of overall variance explained and in terms of average predicted error.

In sum, we take the results of Experiments 2A and 2B as evidence that lure-to-lure interactions reduce RTs above and beyond lure-target similarity effects and that the magnitude of these interactions depend on the spatial configuration of the lures in the display: the lure-to-lure interactions maximally reduce RTs when identical lures are near each other (as in homogeneous and spatially-segregated displays), and their impact on RT decrease as

they become spatially intermixed with other lure items (spatially intermixed conditions).

General Discussion

The goal of this investigation was two-fold. First, we sought to confirm the findings of Wang, Buetti & Lleras (2017) with simple, colored geometric shapes and second, to test whether we can evaluate the magnitude of inter-item interactions in efficient search. Regarding the first goal, in Wang et al. (2017), we used computational simulations based on a theoretical proposal (Buetti et al.’s 2016) of how stage-one processing ought to unfold, when observers are looking for a known target. The basic assumptions are that there is a parallel evaluation of information at all display locations whereby items at those locations are compared to the target item. The more dissimilar an item is from the target, the sooner a decision is reached about that item *not-being* the target. This parallel evaluation was proposed to be of unlimited capacity and, more critical to the

current paper, evaluations (i.e., evidence accumulation) were assumed to be independent from one another. That is to say, we initially assumed that the comparison process of an item to the target template would be unaffected by what other comparisons are simultaneously taken place and *where*.

There was reason to believe a priori that the independence assumption was in fact too strong. In the cognitive neuroscience literature, there is ample evidence of inter-element interactions. Gilbert and colleagues looking at single-cell responses in V1 in monkeys have shown that the response function of V1 neurons is modulated by surrounding neurons (Kapadia, Westheimer, & Gilbert, 2000). Similarly, it is well established that mechanisms like lateral inhibition and feature-suppression mechanisms (e.g., Li, 1999, 2002) produce interactions between nearby items, and, more generally, the Biased Competition Theory of attention (Desimone & Duncan, 1995) was developed in response to the observation that substantial representational limitations arise when two or more objects are sufficiently close to each other so as to decrease the strength of the neural response to each object. In the cognitive psychology literature, inter-item interactions have been thought to be an integral part of “efficient” search going back as far as Duncan and Humphreys (1989) who proposed that identical elements group together and are subsequently rejected as a perceptual group (a mechanism they termed spreading suppression), facilitating search performance. This distractor-distractor similarity effect was independent of distractor-target similarity effects.

In sum, it is not surprising that we find a discrepancy between parallel search performance in homogeneous displays (where inter-item interactions ought to be maximized) and parallel search performance in heterogeneous displays (when those interactions ought to be weakened by interleaving different types of items in the display). That said, what is remarkable is that search performance measured on lure-homogeneous displays predicted almost all the variability of performance observed on lure-heterogeneous displays (with inter-mixed lures). This was initially observed in Experiment 2 of Wang et al. (2017) using real-world images as stimuli ($R^2 = 97\%$) and again here in Experiment 1 ($R^2 = 90\%$). This predictive success puts constraints on how we understand lure-to-lure interactions: *lure-to-lure interactions produce multiplicative effects in a logarithmic efficiency scale*.

From the perspective of the architecture proposed by Buetti et al. (2016) and Wang et al. (2017), there are at least two possible mechanisms to implement these lure-to-lure interactions, as proposed by Wang et al. (2017). First, it is possible that adjacent identical lures facilitate each other’s processing by lowering the amount of evidence needed for each lures’ accumulator to reach a decision. This mutual lowering of evidence thresholds would be maximized when identical lures are near each other and would be weakened when different types of lures are spatially intermixed in the display. Our results are consistent with this hypothesis. A second hypothesis is that the lowering of thresholds does not result from lure-level inter-item interactions but rather

results from a higher-up analysis of the scene. An *evidence monitoring* mechanism would be one that observes how evidence is accumulating across larger regions of the display and sums up (or averages) the collected evidence, much like global motion detectors sum over local motion signals to extract a global motion direction. Lure homogeneous regions of the display could be rejected sooner than lure-heterogeneous regions because it would become clearer sooner that there is an absence of evidence for “targetness” over the homogeneous region. Our results are also consistent with this evidence monitoring hypothesis. Therefore, we cannot adjudicate between the two accounts with the current data and more research is needed to better understand these inter-item interactions. That said, we should also note that these two alternatives are not mutually exclusive and they may very well be both contributing to performance to different extents.

In contrast, we believe our results are evidence against Duncan and Humphreys’ (1989) initial hypothesis that identical distractors tend to group together and be rejected as groups. To be clear, we cannot rule out that similar-looking lures may perceptually “group” together to some extent. However, what we can argue against is the idea that the group is the basis for the rejection of those lures. This follows because our equations demonstrate that each individual item in the display contributes equally to performance. That is, in efficient search, each lure is assessed and compared to the target in parallel, and therefore (partly because of the stochastic nature of parallel processing), the resulting RT by set size function increases logarithmically as a function of set size (Buetti et al., 2016; Wang et al., 2017). Moreover, if lure items grouped together and were rejected in a *qualitatively* different manner than when they were spatially segregated than when they were spatially inter-mixed, it should not be possible to predict variability in performance for the two different arrangements of items based solely on observations from entirely homogeneous displays. Yet, the current data (as well as the data from Wang et al., 2017) confirm that it is possible, to a surprisingly high degree of success. In sum, we believe that there is nothing qualitatively different about rejecting lures in entirely homogeneous lure displays, in heterogeneous but spatially segregated displays and in heterogeneous and spatially intermixed displays. The only thing that varies seems to be a continuous function determining the extent (and strength) of nearby inter-item interactions.

Recent results reported in Utochkin and Yurevich (2016) offer an additional perspective on the nature of heterogeneity effect in efficient visual search. Utochkin and Yurevich manipulated the distribution of a heterogeneous nontarget feature (e.g., bar orientation), so that two heterogeneous conditions were equal in terms of average target-nontarget similarity, but differ in heterogeneity within nontarget items. There were two distinct nontarget types in one condition (e.g., 135, 90 degrees orientation), while the other condition had 6 different types of nontargets (e.g., 135, 126, 117, 108, 99, 90 degrees). They showed that the condition in which nontarget features transition smoothly (e.g., 6 different

orientation values) had faster RT than the condition where nontargets were more distinct from each other. The authors proposed the concept of ‘segmentability’ to account for this effect: nontargets are more segmentable in the distinct nontarget condition than the smooth-transition nontarget condition. When nontargets are easily segmented into spatially overlapping subsets, search is conducted as a serial inspection of each item; however, when nontargets form smooth transitions, they are no longer segmented, but perceived as a single group, thus facilitating the search. We believe our adjacent item interaction hypothesis stated above can also account for their results: in the smooth transition condition, adjacent nontargets are likely to be sufficiently similar to each other, so that facilitative interaction (lowering of accumulation threshold) takes place for all adjacent item pairs and speeds up perceptual decision making, while in the distinct condition, by chance, only half of the adjacent item pairs would have the same facilitative interaction. In other words, adjacent item interactions may be the underlying mechanism for the holistic construct of ‘segmentability’.

It is also worth noting that the proposed nature of inter-item interactions in our experiments is probably not fundamentally different from those local interactions that are believed to be useful in the detection of textures and boundaries (e.g., Knierim & Van Essen, 1992). These local interactions are common place in many theories of attention and visual search (e.g., Bacon & Egeth, 1991; Bravo & Nakayama, 1992; Duncan & Humphreys, 1989, 1992; Itti & Koch, 2001; Nothdurft, 1992; Rubenstein & Sagi, 1993; Wolfe, 1994). What is unique to our proposal is the hypothesis that in spite of objects interactions with nearby similar objects, the fundamental basis for scene analysis remains the objects themselves rather than groups of objects (e.g., Bacon & Egeth, 1991; Duncan & Humphreys, 1989, 1992). Thus, distractor-homogeneous displays are processed faster by the visual system not because the distractors group and are rejected as a group, but rather because the rejection of each individual object is facilitated by the presence of nearby identical items.

The current results suggest that there are two factors impacting the strength of inter-item interactions. The first, and most obvious, is the spatial arrangement of stimuli: when items are spatially segregated, the inter-item interaction observed between items is identical to the one measured in homogeneous displays. Evidence supporting this conclusion comes from the finding that the multiplicative constant in the function Observed RTs as a function of Predicted RTs is close to 1, when predicting performance in spatially-segregated displays (Experiments 2A and 2B, slope = 0.91 and 0.88, respectively).

Our results also suggest a second factor that might determine the strength of inter-item interactions: stimulus complexity. Though it may be premature to conclude based on tests from just two different stimulus sets (colored geometric shapes and real-world stimuli), there appears to be the intriguing possibility that stimulus complexity affects the extent of inter-item interactions. Indeed, we propose that the slope of

the function Observed RT as a function of Predicted RT when testing spatially-intermixed lure displays can be taken as a quantitative index of the strength of these interactions. The rationale is as follows. The predicted RTs are based on logarithmic slope parameters obtained from entirely homogeneous lure displays. In those displays, two factors determine the efficiency (i.e., the log slope) of processing: lure-target similarity and inter-item interactions. Because the displays are homogeneous, those interactions are maximal. When the same lure items are then spatially intermixed in a subsequent experiment, the second factor (inter-item interactions) is minimized, leaving only lure-target similarity as a contributing factor to processing efficiency. As a result, search performance in heterogeneous displays will always be *consistently underestimated*. That is to say, search performance will always be relatively slower in heterogeneous search displays than in homogeneous search displays. But this difference in performance should not be interpreted as a change in the processing architecture needed to process heterogeneous displays, nor does it reflect some fundamentally different type of processing across the two types of displays (homogeneous and heterogeneous). This follows because if there was a completely different factor at play contributing to performance in one but not the other display type, one should not be able to predict performance in one case (heterogeneous displays) based solely on parameters measured in the other (homogeneous displays). Furthermore, our results suggest that the difference in performance between homogeneous and heterogeneous displays does not arise from the increased complexity of heterogeneous displays (i.e., not because of the mere presence of multiple types of items) but rather because of the spatial arrangement of those items in the scene. Indeed, we can almost perfectly predict RTs in heterogeneous displays based on parameters from homogeneous search displays just as long as each type of lure is spatially segregated from other types.

Following this rationale then, we propose that we can (and should) meaningfully interpret the slope of the function Observed RTs as a function of Predicted RTs. Doing so reveals that the slope of this function is substantially larger with simpler stimuli (1.8, Experiment 1) than with real-world stimuli (1.3, Wang et al., 2017). In other words, simpler visual stimuli inter-act strongly with each other than more complex visual stimuli. Granted, we are only offering a qualitative description of what counts as simple or complex stimulus, mostly because at this time we have not systematically investigated this issue, but also because there appears to be a certain degree of face validity to this idea, because it resonates with similar differentiations proposed in the literature (e.g., Reverse Hierarchy Theory, Ahissar & Hochstein, 2004) and because it mirrors the processing complexity of the visual system. V1 through V4 neurons have narrow receptive fields and code for the sort of simple visual features required to process simple geometric shapes like the ones used in Experiment 1, whereas more complex objects (like cars and teddy bears) require processing in more anterior regions of the visual hierarchy (such as IT), with larger

receptive fields. Thus, whereas simple geometric stimuli that are visually similar are likely to generate inter-item interactions at all levels of the visual system (from V1 to IT), more complex stimuli might generate such facilitatory interactions only at relatively higher visual areas (IT). In sum, we argue that the current results suggest that stimulus complexity determines the strength of inter-item interactions that take place between identical items (and that end up facilitating visual processing efficiency in parallel search), with simpler visual stimuli producing stronger inter-item interaction effects. This tentative conclusion deserves further testing, perhaps using converging methodologies.

Limitations

The current results are limited on several fronts. First, it should be noted that our predictions are based on group-aggregated data. From one perspective, this is strong evidence of the reliability and validity of our model and equations because it means that we can make predictions across independent groups of subjects. However, from another perspective, it would be an even stronger demonstration if we had used a psychophysical approach to do this testing: evaluate for each subject their own logarithmic slope for various lure-target combinations and different stimulus types (geometric shapes vs. real-world stimuli) and then use those estimates to predict that same individual's performance in heterogeneous search displays. This experimental design is complicated by the need of multiple sessions per participant. Indeed, a single one hour session is not enough to get a stable estimate of, say, three logarithmic slopes (corresponding to three different lures for a given target) at the subject level. After one session, log fits are much noisier at the individual level than at the group level (and linear fits are even noisier). We believe this is the case because of the stochastic nature of processing and the many different processing configurations and factors that impact any one estimate of a log slope. Indeed, we know that the log slope is impacted by factors such as crowding (Madison, Lleras & Buetti, in press), eccentricity (Buetti et al., 2016) and cortical magnification (Wang, Buetti & Lleras, 2018). That is to say, low level visual factors impact how well evidence accumulates at a given location and when we estimate a single log slope for a given lure-target stimulus pair, we are averaging across all the different lure locations, target locations, size, eccentricity, crowding factors to get a single measure that therefore represents average processing efficiency across *family* of display and stimulus arrangements. In the future, it might be worth attempting a fuller within-subject characterization of the results presented here.

A second limitation of the present study relates to the findings of Experiment 2B. Although we believe our equation does overall a better job at predicting performance (smaller average RT errors), in terms of variability (R^2), the two equations performed relatively similarly. This may be related to the stimuli set itself used in Experiment 2B: two of the three lures have almost identical log slopes (the car and the reindeer).

Thus, it is possible that lacking variability in the log slope parameters made the differentiation between the two models more difficult. However, we knew that going into this project and we took that risk because we wanted to use log slopes of published homogeneous search data to predict performance in novel scenarios (heterogeneous scenes). Future work could do a more thorough test of our hypothesis by using complex stimulus with more differentiated log slopes.

Conclusion

We validated an equation (Equation 2) used to predict search performance in heterogeneous search scenes based on performance observed in homogeneous search scenes. The equation arises directly from our computational model of parallel visual processing in a search task where the target is known ahead of time and all the distractor stimuli are lures. We were able to predict the variability in performance quite well both when lure stimuli were arranged haphazardly around the display (spatially intermixed condition) and when lures were arranged in spatially segregated manner. We conclude that the processing required to complete a parallel search task in lure-heterogeneous scenes is fundamentally the same as in lure-homogeneous scenes, even if at first, lure heterogeneous scenes (**Figure 1**) might appear to be somewhat more complex than the traditional lure-homogeneous scenes used to study efficient search (i.e., when all non-target items in the display are identical to one another). However, we did notice a difference between spatially-intermixed and spatially-segregated scenes, with spatially-segregated scenes producing search performance as efficient as completely homogeneous scenes. This result was obtained both with simple and complex visual stimuli. We take this as evidence that lure-to-lure stimuli interactions facilitate processing in parallel search and that the strength of those interactions is maximal when identical lures are near each other. Finally, we proposed a method that allowed us to quantitatively measure the strength of those lure-to-lure interactions. This method suggests simpler stimuli produce stronger interactions than more complex stimuli.

Data Accessibility Statement

The data reported in this paper is publicly available in OpenScienceFramework: <https://osf.io/v9f3e/>.

Note

¹ Doing this allows us to correct for overall RT differences across groups of participants. As a reminder, the values of the D parameters used for the predictions here come from observations published in Buetti et al. (2016), with entirely different participants. It is also worth noting that because our main measure of success here is a model's R^2 (a measure of variability), the value of the intercept of the linear regression does not change the value of the R^2 .

Competing Interests

The authors have no competing interests to declare.

Appendix

Table A1: Search display conditions from Experiment 1 for Group A, B, and C are shown by set size of lure object. An asterisk (*) indicates the lure item was held constant throughout the experimental session for that group of participants.

Group A:

	Orange Diamonds*	Yellow Triangles	Blue Circles	Total Set Size	Mean RT (SEM)	Mean Error Rate
Target only	0	0	0	1	533 (22)	1.2%
Subset 1 (2 mixed)	4	0	4	9	605 (22)	1.2%
	4	0	8	13	623 (23)	1.8%
	4	0	12	17	622 (22)	1.2%
	4	0	20	25	656 (26)	1.2%
Subset 2 (2 mixed)	4	4	0	9	642 (22)	1.5%
	4	8	0	13	655 (23)	2.2%
	4	12	0	17	680 (21)	1.3%
	4	20	0	25	696 (24)	2.1%
Subset 3 (3 mixed)	4	4	4	13	658 (24)	2%
Subset 4 (3 mixed)	4	4	8	17	673 (26)	2.5%
	4	4	12	21	681 (27)	1.9%
	4	4	20	29	699 (25)	1.4%
Subset 5 (3 mixed)	4	8	4	17	675 (23)	1.1%
	4	12	4	21	697 (24)	1.7%
	4	20	4	29	710 (25)	1.5%

Group B:

	Orange Diamonds	Yellow Triangles*	Blue Circles	Total Set Size	Mean RT (SEM)	Mean Error Rate
Target only	0	0	0	1	490 (15)	0.6%
Subset 1 (2 mixed)	0	4	4	9	535 (13)	1.5%
	0	4	8	13	538 (12)	1.3%
	0	4	12	17	543 (13)	0.9%
	0	4	20	25	554 (13)	1.3%
Subset 2 (2 mixed)	4	4	0	9	592 (14)	1.8%
	8	4	0	13	622 (18)	1.5%
	12	4	0	17	637 (21)	1.1%
	20	4	0	25	694 (24)	2%
Subset 3 (3 mixed)	4	4	4	13	607 (16)	1.1%
Subset 4 (3 mixed)	8	4	4	17	633 (19)	1.2%

(contd.)

	Orange Diamonds	Yellow Triangles*	Blue Circles	Total Set Size	Mean RT (SEM)	Mean Error Rate
	12	4	4	21	637 (21)	1.5%
	20	4	4	29	694 (24)	1.6%
Subset 5 (3 mixed)	4	4	8	17	601 (15)	1.2%
	4	4	12	21	607 (17)	0.7%
	4	4	20	29	622 (17)	0.9%

Group C:

	Orange Diamonds	Yellow Triangles	Blue Circles*	Total Set Size	Mean RT (SEM)	Mean Error Rate
Target only	0	0	0	1	499 (13)	0.6%
Subset 1 (2 mixed)	4	0	4	9	550 (14)	1.5%
	8	0	4	13	584 (15)	1.7%
	12	0	4	17	605 (16)	1.1%
	20	0	4	25	637 (19)	1.4%
Subset 2 (2 mixed)	0	4	4	9	537 (15)	3.1%
	0	8	4	13	548 (12)	3%
	0	12	4	17	552 (13)	2.5%
	0	20	4	25	561 (14)	3.3%
Subset 3 (3 mixed)	4	4	4	13	592 (15)	2.6%
Subset 4 (3 mixed)	8	4	4	17	628 (23)	1.9%
	12	4	4	21	650 (20)	2.3%
	20	4	4	29	705 (25)	2.1%
Subset 5 (3 mixed)	4	8	4	17	611 (18)	2.6%
	4	12	4	21	616 (15)	2.3%
	4	20	4	29	638 (17)	1.9%

Table A2: Search display conditions in Experiment 2A considered in regression analysis and corresponding RT and error rate data.

	Case 1	Case 2	Total Set Size	Mean RT (SEM)	Mean Error Rate
Target only	1T	1T	1	455 (10)	0.9%
Orange Diamonds (OD) + Yellow Triangles (YT)	4OD + 3YT + 1T	4YT + 3OD + 1T	8	530 (13)	1.4%
	8OD + 7YT + 1T	8YT + 7OD + 1T	16	565 (15)	1.3%
	16OD + 15YT + 1T	16YT + 15OD + 1T	32	593 (16)	1.8%
Orange Diamonds (OD) + Blue Circles (BC)	4OD + 3BC + 1T	4BC + 3OD + 1T	8	512 (12)	0.5%
	8OD + 7BC + 1T	8BC + 7OD + 1T	16	530 (13)	0.9%
	16OD + 15BC + 1T	16BC + 15OD + 1T	32	547 (13)	0.8%
Yellow Triangles (YT) + Blue Circles (BC)	4BC + 3YT + 1T	4YT + 3BC + 1T	8	501 (11)	1.2%
	8BC + 7YT + 1T	8YT + 7BC + 1T	16	518 (12)	1.6%
	16BC + 15YT + 1T	16YT + 15BC + 1T	32	535 (11)	1.2%

Table A3: Search display conditions in Experiment 2B considered in regression analysis and corresponding RT and error rate data.

	Case 1	Case 2	Total Set Size	Mean RT (SEM)	Mean Error Rate
Target (T) only	1T		1	598 (17)	1.5%
Grey Car (GC) + White Reindeer (WR)	4GC + 3WR + 1T	4WR + 3GC + 1T	8	667 (16)	1.9%
	8GC + 7WR + 1T	8WR + 7GC + 1T	16	679 (16)	1.4%
	16GC + 15WR + 1T	16WR + 15GC + 1T	32	694 (16)	0.9%
Grey Car (GC) + Red Doll (RD)	4GC + 3RD + 1T	4RD + 3GC + 1T	8	719 (18)	1.4%
	8GC + 7RD + 1T	8RD + 7GC + 1T	16	746 (18)	1.6%
	16GC + 15RD + 1T	16RD + 15GC + 1T	32	789 (20)	1.4%
White Reindeer (WR) + Red Doll (RD)	4RD + 3WR + 1T	4WR + 3RD + 1T	8	719 (17)	1.7%
	8RD + 7WR + 1T	8WR + 7RD + 1T	16	756 (22)	1.3%
	16RD + 15WR + 1T	16WR + 15RD + 1T	32	776 (20)	1.2%

Author Contributions

SB, ZW and AL designed the experiments. ZW and AM programmed the experiments. ZW and AM analyzed the data. AL and SB wrote the paper.

References

- Ahissar, M., & Hochstein, S.** (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, *8*, 457–464. DOI: <https://doi.org/10.1016/j.tics.2004.08.011>
- Bacon, W. F., & Egeth, H. E.** (1991). Local processes in preattentive feature detection. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 77–90. DOI: <https://doi.org/10.1037/0096-1523.17.1.77>
- Brainard, D. H.** (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436. DOI: <https://doi.org/10.1163/156856897X00357>
- Bravo, M. J., & Nakayama, K.** (1992). The role of attention in different visual-search tasks. *Perception & Psychophysics*, *51*, 465–472. DOI: <https://doi.org/10.3758/BF03211642>
- 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. DOI: <https://doi.org/10.1037/xge0000163>
- Desimone, R., & Duncan, J.** (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*(1), 193–222. DOI: <https://doi.org/10.1146/annurev.ne.18.030195.001205>
- Duncan, J., & Humphreys, G.** (1992). Beyond the search surface: Visual search and attentional engagement. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 578–588. DOI: <https://doi.org/10.1037/0096-1523.18.2.578>
- Duncan, J., & Humphreys, G. W.** (1989). Visual search and stimulus similarity. *Psychological Review*, *96*, 433–458. DOI: <https://doi.org/10.1037/0033-295X.96.3.433>
- Haslam, N., Porter, M., & Rothschild, L.** (2001). Visual search: Efficiency continuum or distinct processes? *Psychonomic Bulletin & Review*, *8*, 742–746. DOI: <https://doi.org/10.3758/BF03196212>
- Hulleman, J., & Olivers, C.** (2017). On the brink: The demise of the item in visual search moves closer. *Behavioral and Brain Sciences*, *40*. DOI: <https://doi.org/10.1017/S0140525X16000364>
- Itti, L., & Koch, C.** (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, *2*, 194–203. DOI: <https://doi.org/10.1038/35058500>
- Kapadia, M. K., Westheimer, G., & Gilbert, C. D.** (2000). Spatial distribution of contextual interactions in primary visual cortex and in visual perception. *Journal of Neurophysiology*, *84*(4), 2048–2062. DOI: <https://doi.org/10.1152/jn.2000.84.4.2048>
- Knierim, J. J., & van Essen, D. C.** (1992). Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *Journal of Neurophysiology*, *67*, 961–980. DOI: <https://doi.org/10.1152/jn.1992.67.4.961>
- Li, Z.** (1999). Visual segmentation by contextual influences via intracortical interactions in the primary visual cortex. *Network*, *10*, 187–212. DOI: https://doi.org/10.1088/0954-898X_10_2_305
- Li, Z.** (2002). A saliency map in primary visual cortex. *Trends in Cognitive Sciences*, *6*, 9–16. DOI: [https://doi.org/10.1016/S1364-6613\(00\)01817-9](https://doi.org/10.1016/S1364-6613(00)01817-9)
- Neider, M. B., & Zelinsky, G. J.** (2008). Exploring set size effects in scenes: Identifying the objects of search. *Visual Cognition*, *16*(1), 1–10. DOI: <https://doi.org/10.1080/13506280701381691>
- Nothdurft, H. C.** (1992). Feature analysis and the role of similarity in preattentive vision. *Perception & Psychophysics*, *52*, 355–375. DOI: <https://doi.org/10.3758/BF03206697>
- Pelli, D. G.** (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442. DOI: <https://doi.org/10.1163/156856897X00366>

- Rubenstein, B. S., & Sagi, D.** (1993). Effects of foreground scale in texture discrimination tasks: Performance is size, shape, and content specific. *Spatial Vision*, 7, 293–310. DOI: <https://doi.org/10.1163/156856893X00450>
- Strasburger, H., Rentschler, I., & Jüttner, M.** (2011). Peripheral vision and pattern recognition: A review. *Journal of Vision*, 11(5), 13, 1–82. DOI: <https://doi.org/10.1167/11.5.13>
- Utochkin, I. S., & Yurevich, M. A.** (2016). Similarity and heterogeneity effects in visual search are mediated by “segmentability”. *Journal of Experimental Psychology: Human Perception and Performance*, 42(7), 995–1007. DOI: <https://doi.org/10.1037/xhp0000203>
- Wang, Z., Buetti, S., & Lleras, A.** (2017). Predicting Search Performance in Heterogeneous Visual Search Scenes with Real-World Objects. *Collabra: Psychology*, 3(1), 6. DOI: <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*, 1–8. DOI: <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, 202–238. DOI: <https://doi.org/10.3758/BF03200774>
- Wolfe, J. M., & Horowitz, T. S.** (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, 5, 495–501. DOI: <https://doi.org/10.1038/nrn1411>
- Zelinsky, G. J.** (2008). A theory of eye movements during target acquisition. *Psychological Review*, 115(4), 787–835. DOI: <https://doi.org/10.1037/a0013118>

Peer review comments

The author(s) of this paper chose the Open Review option, and the peer review comments are available at: <http://doi.org/10.1525/collabra.151.pr>

How to cite this article: 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: Psychology*, 5(1): 2. DOI: <https://doi.org/10.1525/collabra.151>

Senior Editor: Rolf Zwaan

Editor: John Henderson

Submitted: 26 February 2018

Accepted: 27 November 2018

Published: 08 January 2019

Copyright: © 2019 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.