UNIVERSITY of CALIFORNIA PRESS | Collabra

Wang, Z., et al. (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

ORIGINAL RESEARCH REPORT

Predicting Search Performance in Heterogeneous Visual Search Scenes with Real-World Objects

Zhiyuan Wang, Simona Buetti and Alejandro Lleras

Previous work in our lab has demonstrated that efficient visual search with a fixed target has a reaction time by set size function that is best characterized by logarithmic curves. Further, the steepness of these logarithmic curves is determined by the similarity between target and distractor items (Buetti et al., 2016). A theoretical account of these findings was proposed, namely that a parallel, unlimited capacity, exhaustive processing architecture is underlying such data. Here, we conducted two experiments to expand these findings to a set of real-world stimuli, in both homogeneous and heterogeneous search displays. We used computational simulations of this architecture to identify a way to predict RT performance in heterogeneous search using parameters estimated from homogeneous search data. Further, by examining the systematic deviation from our predictions in the observed data, we found evidence that early visual processing for individual items is not independent. Instead, items in homogeneous displays seemed to facilitate each other's processing by a multiplicative factor. These results challenge previous accounts of heterogeneity effects in visual search, and demonstrate the explanatory and predictive power of an approach that combines computational simulations and behavioral data to better understand performance in visual search.

Keywords: visual search; heterogeneity; parallel processing; computational modeling; visual attention; real-world objects

Introduction

Parallel processing in visual search

Starting from the retina, early stages of the human visual system are organized in a parallel architecture, so that low-level information is extracted and represented simultaneously for a wide view of the world (Breitmeyer, 1992). On the other hand, there are several central bottlenecks limiting the amount of information that the mind can actively maintain, process and respond to. Those bottlenecks are exemplified in phenomena like the limited capacity of visual working memory (Sperling, 1960; Luck & Vogel, 1997; Awh, Barton & Vogel, 2007), the psychological refractory period (Pashler, 1992; Sigman & Dehaene, 2008), and the attentional blink (Shapiro, Raymond & Arnell, 1997; Vogel, Luck & Shapiro, 1998). The needs to both have parallel access to basic visual information around us and to focus high-level processing on select information are key constraints in the study of visual attention. Because the mind is almost always motivated by a specific goal, understanding of goal-directed visual processing is thus essential to understanding visual attention. By presenting a target object among various distracting items, visual search provides a convenient method to capture the goal-directed selection process of attention and therefore has been widely used and studied in the visual attention research.

Notably, much of the effort in the visual search literature has been devoted to understanding focused attention, a capacity-limited form of visual attention where items or subsets of items in the display are serially processed. Plenty of empirical research has thus focused on the dependent variable of search slope—how much longer on average it takes for the visual system to process an additional item, and tried to establish relationships between different task settings and corresponding changes in search slopes. These task setting variables include how many features define the target (Treisman & Gelade, 1980), whether the target is defined by known features (Bravo & Nakayama, 1992), the similarity between target and distractors (Duncan & Humphreys, 1989), or what specific features are used to differentiate target from distractors (Wolfe & Horowitz, 2004). When this kind of relationship is successfully mapped out, inferences can be drawn about the nature or function of focused attention. For example, a key question in the history of visual search research had been to examine under what conditions a search slope becomes non-zero, which was thought to reflect the limit of parallel processing in the visual system. The fact that conjunction search produced non-zero search slopes while feature search did not lead to the suggestion that focused attention is necessary for the binding of different features onto an object file (Treisman & Gelade, 1980). This approach assumes a linear relationship between reaction time and

Department of Psychology, University of Illinois at Urbana-Champaign, Champaign, Illinois, US

Corresponding author: Zhiyuan Wang (zwang153@illinois.edu)

number of items (or 'set size'), which is often observed and is consistent with a capacity limitation in high-level processing. As a consequence, many prominent theories of visual search (and by extension, of visual attention) have been essentially accounts of the search slope given a specific search task (e.g., Duncan & Humphreys, 1989; Wolfe, 1994). Although alternative focus on the accuracy of responses has also generated important insights such as the signal detection theory of visual search (Verghese, 2001), the traditional method of studying reaction time as a function of set size has become one of the most popular approaches of attention research.

Perhaps partly because of this tradition, cognitive experimental research on visual search has become somewhat disinterested in understanding parallel processing in visual search. The specific reason for this disinterest might be because of the assumption that parallel processing is synonymous with 'flat' search functions. This follows from two observations: first, when the linear regression of RT by set size returned a slope coefficient that's close to zero (or smaller than 10ms/item), the search function is typically assumed to be a flat straight line; second, parallel processing with unlimited capacity was assumed to produce no (meaningful) additional time cost as additional items are introduced to a display. Therefore, when search slope is found to be near zero, the usual inference is that items are processed in parallel and that there is no need for attentional selection. Thus, the data producing that pattern were considered to be not informative for understanding visual attention. Yet, as discussed below, recent findings indicate that neither of these assumptions necessarily hold (Buetti, Cronin, Madison, Wang & Lleras, 2016). Nevertheless, major theories of visual search assume that parallel visual processing produces no meaningful variability in reaction time. For example, Guided Search (Wolfe, 1994) used a fixed 400 ms constant for the time cost to process the search display, compute the 'priority map' (and prepare and execute a response to a target, once found). Bundesen's Theory of Visual Attention (1990) also had a mathematically explicit goal to make search time independent of set size in so-called pop-out searches, which were thought to depend entirely on parallel processing.

Our knowledge about the parallel processing of visual scenes has largely come from computationally-oriented approaches to vision, in which the central goal is to predict the series of loci of attention or eye fixations given a specific scene or image. By the success of these computational models of attention in predicting human fixations, one can argue that low level, parallel computations carried out by these models mimic the parallel processing of human vision. For example, the work by Itti and Koch (2000; 2001) suggested that bottom-up, image-based control of attention can be formalized by the computation of a 'saliency map', a sum of various feature contrasts over different spatial scales. This model enjoys some degree of success in free-viewing tasks, particularly with complex scenes. Other saliency models operate on ideas such as self-information (Bruce & Tsotsos, 2009), Bayesian surprise (Itti & Baldi, 2009), and figure-ground segmentation (Zhang & Sclaroff, 2013), which can be seen as alternative accounts of the computational basis of bottom-up attention. One advantage of these models is that they are highly specific and allow for testable predictions regarding human performance, but the majority of these predictions are focused on measures like fixation distribution. The downside of many recent computational models of parallel vision is that their increased complexity does not allow for a clear understanding of the underlying mechanisms. For example, the currently top 2 ranking (using the AUC-Judd metric) saliency models (Kruthiventi, Ayush, & Babu, 2015; Vig, Dorr, & Cox, 2014) on the MIT Saliency Benchmark (Bylinskii, Judd, Durand, Oliva, & Torralba, 2014) are both based on deep neural networks. This approach is based on learning hierarchical features represented by multiple layers of neurons. However, there has been little effort to understand the correspondence between these learned features and actual representations in the human visual system. Thus, while these models describe the computations carried out in early vision, many of them cannot be directly related to visual search behavior. Important exceptions should be noted, such as Zelinsky's Target Acquisition Model (2008), which was developed to predict scan path behavior in target present visual search based on a mechanistic model, and Najemnik & Geisler (2008), whose ideal-observer model reveals important deficiencies in common theories of visual attention. Both provide more specific predictions in terms of sequences of fixations (and thus saccades) rather than simply fixation distributions. Still, none of these models include estimates of visual processing times in humans, which complicates detailed comparisons to human performance. Lastly, Rosenholtz, Huang, Raj, Balas, and Ilie (2012) proposed a texture tiling theory of crowding in peripheral vision that considers search efficiency to be a function of summary visual statistics over peripheral pooling regions that aggregate low level visual information. However, here too, no differentiation was made between visual processing times and focused attention decision times, and an explicit mechanism is lacking to predict actual processing time given those summary statistics.

Consequently, important aspects of the influence of parallel processing in visual search are largely unchartered. There are several reasons why understanding this processing stage is important. *A priori*, selective attention evolved to address the need to optimally bridge the gap of processing capacity between early parallel visual processing and higher level processing, therefore understanding what information the parallel stage can process naturally provides boundaries to what the attentive, limited capacity stage needs to do and/or compute. More importantly, the often-implied assumption that parallel, unlimited capacity processing results on constant processing times simply does not hold: Townsend and Ashby (1983), for instance, provided a precise mathematical formulation of a variety of such processing models, many of which predict non-flat RT by set size functions. The counter-intuitiveness of these results can be perhaps dispelled if one considers that *unlimited* capacity (i.e., the term referring to the fact that information is processed simultaneously at various spatial locations/channels, independently of the number of locations/channels) should not be and is not synonymous with *infinite* capacity (i.e., that there are no limitations to the processing capacity at any one location).

We propose then that developing a better understanding of early parallel processing ought to be very informative to attention research. Empirically, there are various experimental results indicating that the visual system can rapidly access substantial amounts of information without focused attention, such as scene gist (Potter & Levy, 1969; Potter, 1976; Schyns & Oliva, 1994; Oliva, 2005), statistical properties in a scene (e.g., Parkes, Lund, Angelucci, Solomon & Morgan, 2001; Chong & Treisman, 2005a, 2005b; Haberman & Whitney, 2009), and some basic categorical information of objects (Li, VanRullen, Koch, & Perona, 2002; Li, Iyer, Koch & Perona, 2007). Such processing power must be based on this parallel processing stage, of which relatively little has been learned. Additionally, current theories often fail to account for search performance variability in real world scenes (e.g. Itti & Koch, 2000; Wolfe, Alvarez, Rosenholtz, Kuzmova & Sherman, 2011), which could be at least partly due to neglecting the processing variability arising from the parallel processing stage.

Systematic variability in efficient search

Recent work in our lab demonstrated an important reaction time signature of the parallel processing stage in fixed-target, efficient visual search (Buetti, Cronin, Madison, Wang & Lleras, 2016). Our results showed that in addition to a linear increase in reaction time caused by distractor items highly similar to the target, less similar items can produce a logarithmic increase in reaction time as set size increases. This logarithmic function can be easily overlooked if one does not sample the set size conditions appropriately and simply make a linear regression to the data. **Figure 1** illustrates key aspects of our results. These two different signatures in reaction time lead us to propose a distinction between two types of visual distractors: candidates and lures. *Candidates* are items that require focused spatial attention to be distinguished from the target because they share too many visual characteristics with the target (such as color, curvature, line intersection, orientation). As a result, given the known representational limitations of peripheral vision, human observers cannot discriminate candidates from the target in parallel in the peripheral field of view. In contrast, *lures* are items that are sufficiently different from the target along some set of visual features that they do not require close scrutiny. That is, the resolution of peripheral viewing is sufficient for determining that lure items are *not* the target. Take for example the case of looking for a watering can in your garden. Close scrutiny is likely not required to decide that fence, trees, flowers, grass, and large lawn furniture are not a watering can. You can, therefore, discard all such objects as unlikely targets in parallel, and we would refer to them as lures, in this particular example. Other medium sized objects of similar size, color, and material (maybe some children toys) might be confusable with the watering can in peripheral vision. We would refer to those objects as candidates and those candidates would require focused attention to be differentiated from the target.

Returning to lures, lures are sufficiently different from the target that they can be processed in parallel, across the visual scene and, with a high degree of success, they can be ruled out as non-targets. When candidates and lures are both present in a scene, one can dissociate the linear and logarithmic RT contributions to overall RT that each bring about (see **Figure 1A**). Furthermore, we also demonstrated that different types of lures produce logarithmic RT by set size functions of different steepness, depending

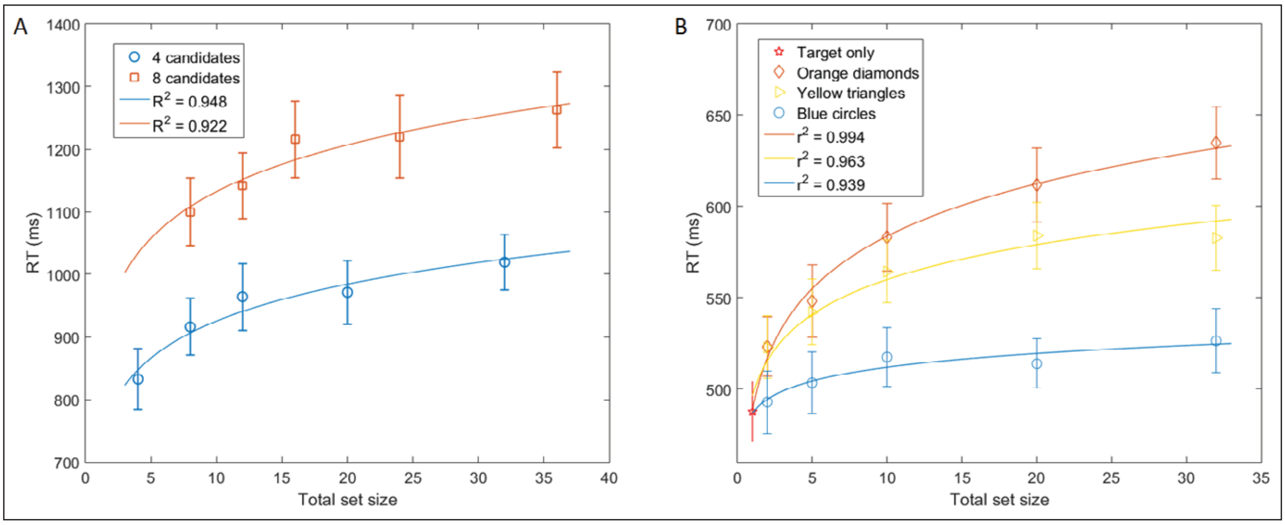

Figure 1: Key findings demonstrating logarithmic RT by set size functions from Buetti et al. (2016). ***Panel A*:** Data from Experiment 3A of Buetti et al. (2016). The task was to find a 'T' target among 'L' candidates and thick orange cross lures. The data are best described as a logarithmic function of total set size when the number of candidates is held constant. Notice that the two curves for two different candidate set sizes are highly parallel, suggesting that candidates introduce a linear increase in reaction time. ***Panel B*:** Data from Experiment 1A of Buetti et al. (2016). Reaction times to find a red triangle target among different types of lures are best fit by logarithmic functions, whose steepness or 'logarithmic slopes' are modulated by the similarity between lure and target.

on their visual similarity to the target, such that lures that are more similar to the target produce steeper logarithmic curves (**Figure 1B**). Notably, if linear regressions were performed on truncated sections of this RT x Set Size functions, most of these data would yield very small linear slopes (in all cases, below the traditional 10ms/item "benchmark" for efficient search).

Given these results, we proposed that lure items are processed in the first parallel stage of vision to the degree that there is sufficient evidence to reject them as possible targets. Naturally, candidates go through this parallel stage because the resolution limitation at this stage of processing means it cannot differentiate them from the target. Locations where information is not differentiated in that manner are passed on for analysis by the second stage of focused spatial attention. Further, the relationship between lure-target similarity and the slope of the logarithmic function indicates that lure-target similarity determines the efficiency of processing for each *individual* lure item.

We developed the following set of hypotheses to construct a theoretical model of stage-one visual processing that allows us to understand variability in stage-one processing times:

(1) Consistent with traditional assumptions of early visual processing (e.g., Treisman & Gelade, 1980; Wolfe, 1994), we proposed that stage-one processing has a parallel architecture and unlimited capacity. Hence, all items in the display are simultaneously processed with a rate that doesn't depend on set size.

(2) During stage-one processing, the visual system is attempting to make a binary decision at each location where there is an item. The question is: is this item sufficiently different from the target? If so, the item is unlikely to be the target and doesn't require further processing. If it is sufficiently similar to the target, given the resolution limitations of peripheral vision, the item will require further processing and its location will be passed on to stage-two processing. An eye movement or a deployment of focused attention will be required to resolve and inspect the item to determine whether or not it is the target.

(3) The amount of evidence required to reach a decision about an item (the "decision threshold") is proportional to its similarity to the target. This follows from the idea that the more visually similar an item is to the target, more information is needed to determine that the item is indeed not the target and will not require further inspection. Given the resolution limitation of peripheral vision, there is a maximum decision threshold. All locations containing items that reach that level (i.e., items too similar to the target and the target itself) will be passed on to the second stage of processing.

In order to make explicit predictions with this theory, we specified the following assumption to model individual item processing times:

(4) Processing of individual items is modeled by noisy accumulators. The rate of information accumulation at each instant is drawn from a Gaussian distribution with a positive mean value. Processing is complete when accumulated evidence reaches a decision threshold. As proposed above, the decision threshold is proportional to the item's similarity to the target. This process is thus mathematically equivalent to a Brownian motion with a constant drift rate towards a given threshold. Completion time t of this process follows the Inverse Gaussian distribution (Chhikara, 1988):

$$f(t \mid A, k, \sigma) = \frac{A}{\sqrt{2\pi\sigma^2 t^3}} e^{-\frac{(A-kt)^2}{2\sigma^2 t}} \tag{I}$$

where A is the accumulator's threshold, k is the constant drift rate (or mean accumulation speed), and σ is the standard deviation of accumulated information at each instant.

These assumptions enabled us to numerically simulate different implementations of parallel, unlimited capacity processing system, and derive the expected time cost as a function of number of items to be processed, modulated by the similarity of items to the target. Specifically, following the pioneering work by Townsend and Ashby (1983), we implemented different termination rules (self-terminating vs. exhaustive) in systems with or without resource reallocation in the case of efficient search (see Buetti et al. (2016) for results and detailed methods of these simulations in their **Figure 3** and Appendix A.). Our simulation results indicated that only a system with an exhaustive termination rule (i.e., the stage is complete once all items are fully processed) and no reallocation of resources produces logarithmic curves. Further, we demonstrated that in such cases, the steepness of these logarithmic curves are modulated by similarity of lure items to the target just as observed in our experiments (see **Figure 1B**). In other words, we demonstrated a one-to-one correspondence between decision thresholds in our accumulator models and the slopes of the logarithmic completion times, such that smaller decision thresholds produce flatter logarithmic slopes and larger decision thresholds produce steeper logarithmic slopes. In sum, we found evidence (based on empirical data and a set of reasonable assumptions) that stage one in visual search functions as a parallel, unlimited capacity, exhaustive processing system. When there are no candidate items in the display (other than the target), this model can account for all the systematic reaction time variation caused by changes in the number of lure items. Our simulations combined with our behavioral results suggest that the coefficient of the logarithmic slope observed in behavioral experiments can be interpreted as an index of lure-target similarity, as it reflects the amount of evidence required to reject the lure as a possible target (Buetti et al., 2016, Appendix A).

Predicting performance in lure-heterogeneous displays

Given our current theory of stage-one processing in visual search, one intriguing application is to the understanding of performance in search tasks with multiple types of distractors presented simultaneously. Many laboratory experiments on visual search use highly homogeneous displays, i.e. the distractor items are either completely identical, or composed of groups that differ from each other in only one feature dimension. In the real world, however, an arbitrary scene often consists of mostly non-repeating objects. When a specific target is defined, it is also usually the case that most non-target objects are highly dissimilar to the target, so that very few of them need to be actually examined (Neider & Zelinsky, 2008; Wolfe et al., 2011). Thus, it seems that many visual search-like tasks performed in the real world will be best conceptualized as search for the target amongst a heterogeneous set of lure items.

Notice that many conclusions drawn from homogeneous search tasks cannot be easily extended to a heterogeneous search scenario simply. Duncan and Humphreys (1989) already pointed out that distractor-distractor similarity (or heterogeneity in the distractor set) has an effect independent of target-distractor similarity. Guided Search theory (Wolfe, 1994) proposed that top-down attention could 'guide' parallel processing by prioritizing items with specific feature values of the target. Yet in the real world, objects are defined as conjunctions of many different feature dimensions so that groups of objects can share a few features, while still being the case that each object is sufficiently dissimilar to every other one along several feature dimensions. Nordfang and Wolfe (2014) found that in the case of high feature dimensionality, the effect of heterogeneity in visual search could not be explained by a linear summation of the 'guidance' afforded by each feature dimension. Therefore, the difference or relationship between homogeneous and heterogeneous search is still relatively unclear.

One prominent aspect of our current theory is that it emphasizes the role of visual similarity in the parallel stage of processing, and it makes a more specific formulation of the effect of target-distractor similarity in comparison to Duncan and Humphreys (1989) and other previous theories. Further, the concept of visual similarity is abstract enough to be applied to both artificial and naturalistic stimuli alike. Hence, we expect our previous results to extend to tasks using natural images as search items. Specifically, efficient search should always be modulated by lure-target similarity, and should produce logarithmic RT by set size functions, when observers are looking for a specific target. For example, search for a teddy bear target among an array of toy pandas and model cars should both produce logarithmic RT by set size functions because both toy pandas and model cars look sufficiently dissimilar to the target. And the function for toy panda lures should be steeper than the log curve produced by a search for a teddy bear among model cars, as long as the toy panda is visually more similar to the teddy bear than the model car.

More importantly, the degree of similarity between one distractor item and the target item *should not* depend on what other objects are present in the scene, or whether the distractor set is homogeneous or heterogeneous. Moreover, as mentioned above, our theory and results suggest that the 'slope' of the logarithmic function measured in homogeneous search can be a valid behavioral index of lure-target similarity. Hence, in principle, we should be able to predict search times in *lure-heterogeneous* displays based on participant's performance on *lure-homogeneous* displays. This follows because, as we just mentioned, our model proposes that there is a one-to-one correspondence between accumulation thresholds and the log slope coefficients of homogeneous search. If this is correct, then we should be able to derive accumulation thresholds for each lure type from the observed log slopes of lure-homogeneous search data. Then, we should be able to use these thresholds to predict search RT in novel, heterogeneous scenes.[1] This fact illustrates the generalizability and specificity of our theory: it makes specific RT predictions for performance in novel, untested experimental scenarios. We can then compare the RT predictions to observed experimental data to test the accuracy of the model. Further, systematic deviations from the model's predictions can be used to infer undiscovered properties of human parallel processing, as we demonstrate below.

There are two obstacles that need to be resolved before we can take on this approach. The first issue is that an analytical solution for stage-one processing time based on our current model is not readily available, which means that given observed log slope values, we cannot directly compute the corresponding accumulator thresholds. This is because even though the individual accumulator's completion time is well understood (formula 1), in the case of heterogeneous displays (where individual completion times are sampled from multiple groups of different Inverse Gaussian distributions), the maximum of all items' completion times (since our model assumes an exhaustive termination rule) requires an integral that seems to be analytically unsolvable.[2] To circumvent this issue, in Buetti et al. (2016) we used a computational simulation approach to find numerical mappings between thresholds and log slopes. This is the same approach that we will use here to make numerical predictions of heterogeneous search performance. Specifically, we developed several equations predicting heterogeneous search time based on different theoretical assumptions, and compared their predictions to simulated heterogeneous search results. The best-performing equation was taken as the prediction of our theory, in lieu of the exact analytical solution.

A second issue lies in the fact that our model assumed individual items' processing are independent of each other, and this assumption was not directly backed by evidence. In Buetti et al. (2016), we rejected one type of processing interaction: the resource-reallocating model. We were able to do so because this family of models produces a qualitatively different RT by set size function than non-reallocating models (i.e., a monotonically decreasing function). However other types of processing interactions are possible. In particular, models where lure-to-lure interaction effects are additive or multiplicative and constant over the time course of stage-one processing could not be ruled out. This is because in Buetti et al. (2016) we made a qualitative comparison between the various simulation results and the shape of the observed RT by set size functions in human participants. Thus, if homogeneity or heterogeneity in the search scene were to introduce a constant (additive or multiplicative) change in the processing of items, the overall shape of the RT functions would still be logarithmic and our model's predictions would be inaccurate by either an additive or multiplicative factor. For instance, one might expect that display heterogeneity slows down overall processing, or the reverse, that display homogeneity facilitates processing via a mechanism like the 'spreading suppression' originally suggested by Duncan and Humphreys (1989). This issue is therefore an empirical question, and we resolved it by designing empirical tests of our theory's predictions. To anticipate, by examining the deviations of observed heterogeneous search performance from the predictions based on the independence assumption, we gained insight to what type of interaction might be taking place in homogeneous displays to facilitate homogeneous search performance.

Strategy adopted in our computational approach and predictions

First, we began by simulating homogeneous search completion times for three types of lures by using a different accumulation threshold for each lure type. We then *estimated* the log slope coefficients for each of these lure types by finding the best-fitting logarithmic slope coefficient for the function relating number of lures to completion times. We refer to these slopes as D values.

Second, we ran simulations of completion times for heterogeneous search scenes. Each scene was composed of a varying number of each of the three types of lures. Processing of every lure was modeled by *the same type of accumulators,* with the same accumulation thresholds as used in homogeneous displays. In other words, here we assumed that lure-target similarity is context-independent, that is, the degree of similarity between one type of lure and the target should not depend on what other objects are present in the scene, or whether the distractor set is homogeneous or heterogeneous.

Third, based on different assumptions about how the processing of heterogeneous search scenes might unfold, we developed four different theoretical models of stage-one processing. For each of these models, we derived a hypothetical equation that approximates the completion time as a function of the number of lures of each type present on the display and each lure type's logarithmic slope coefficient (i.e., the D values extracted from lure-homogeneous simulations). The models and their corresponding equations were pre-registered on the OpenScienceFramework website, in the context of the pre-registration for Experiment 2 (osf.io/2wa8f).

Fourth, we compared the *simulated* completion time for each display with the completion time *predicted* by each of the four processing models. This comparison allowed us to select the best-performing equation as the optimal formulation of the relation between homogeneous and heterogeneous search performance. As a reminder, this series of steps was necessary because we do not know of an exact analytical solution for completion times in heterogeneous displays using our accumulators (see Note 2). The detailed methods and results of this simulation procedure is presented below in the Predictive Simulation section.

Finally, we did an *empirical* test of our computational model by using the best-performing equation to predict behavioral data. That is, we used performance observed in a homogeneous search experiment (Experiment 1) to predict performance in a separate experiment with heterogeneous displays (Experiment 2). This amounts to a rigorous empirical test of our computational model and of its underlying assumptions. Using Experiment 1 data, we can estimate best-fitting log slopes for each type of lure type. We can then use those estimated D values in conjunction with the best-performing equation to predict what RTs ought to be in the various heterogeneous conditions tested in Experiment 2. We can then compare predicted RTs with *observed RTs* to test whether the equation favored by our theory-based simulation also outperforms all other alternative equations in behavioral data. This comparison allowed us to assess the predictive power of our theory.

This predicted-to-observed RT comparison will also be used to investigate whether individual items are processed independently of one another or whether there are inter-item interactions that produce systematic deviations from the model. Because our simulation was conducted under the assumption of between-item independence, the predictions of this equation naturally carry that assumption along. Hence, any systematic deviation of predicted stage-one processing times from observed heterogeneous search data can be interpreted as effects of homogeneity (or heterogeneity, depending on the viewpoint) on processing times. To estimate the deviations from our model, we fit predicted stage-one processing times to observed data via a linear regression. On the one hand, if the model allowed for a perfect prediction, then observed and predicted stage-one time costs should line up precisely along the $y = x$ line, with y being observed stage-one time costs on every heterogeneous condition tested, and x being the predicted stage-one time costs for each condition, based on the model and the parameters obtained from homogeneous displays. We would then conclude that items are always processed independently of context. On the other hand, if there are systematic deviations from the predicted stage-one time costs in observed data, these indicate that lures are not processed in a context-independent fashion. Two types of systematic deviations can be therefore obtained: a multiplicative deviation and an additive deviation. If the estimated slope coefficient between observed and predicted stage-one time costs is different from 1, then this would indicate a multiplicative effect of homogeneity on single-item processing. If the intercept coefficient of the observed-by-predicted time cost function is different from 0, this would indicate an additive effect of homogeneity on stage-one processing times (the observed time costs would then be shifted below the $y = x$ line).

As a summary, the current study consists of a predictive simulation and two experiments. The simulation provides a best-performing equation for predicting the completion times for stage-one processing in heterogeneous visual search, using parameters obtained from homogeneous visual search displays. Next, we used a set of real-world object images to construct homogeneous search displays in Experiment 1 and heterogeneous search displays in Experiment 2. Experiment 1 served both as an extension of our previous 'feature search' results (as in Experiment 1 of Buetti et al., 2016) to real-world objects and as the data source for estimating log slope (D) coefficients for predicting heterogeneous search data in Experiment 2. In Experiment 2, we collected behavioral data of visual search with heterogeneous displays each containing a mix of the objects used in Experiment 1, using a different group of subjects. By comparing different equations' predictions of RT to observed RT, we were able to (1) test whether the best-performing equation from our theory-based simulation works best in reality, and (2) examine whether there is a multiplicative or additive effect of homogeneity on stage-one processing and if so, estimating the magnitude of that effect in our data.

Predictive Simulation

Methods

Approximating equations for heterogeneous search

We developed the following set of equations in the hope that some of them may be a good approximation of the exact analytical solution of stage one processing time cost in heterogeneous lure search. Each equation describes time cost of stage-one processing, *T*, as a function of the *D* coefficients and numbers of each type of lure items (*N*), i.e. $T = f(\{D_i\}, \{N_i\})$. The *D* coefficients here are meant to be a proxy of the accumulation threshold values, and thus are assumed to be independent of context (homogeneous or heterogeneous). Therefore, given the *D* coefficients estimated based on homogeneous search, the equations provide different predictions of heterogeneous search time, based on different underlying hypotheses of how search unfold in a heterogeneous scene. For each equation, its form in the case of 3 types of lures will be presented below (rather than the general form with arbitrary number of lure types). D coefficients will have the ordering of $D_3 > D_2 > D_1 > 0$, i.e. we denote lure no.3 to have the highest similarity to the target and lure no.1 with lowest similarity. Note that these equations do not include time costs associated with other processing stages, such as encoding, response selection, and execution, which we assume to be constant in efficient search tasks for a given target.

Equation 1:

$$T = D_1\, ln(N_1 + N_2 + N_3 + 1) + (D_2 - D_1)\, ln(N_2 + N_3 + 1) \\ + (D_3 - D_2)\, ln(N_3 + 1)$$

This equation was a simple extension of the concepts in Buetti et al. (2016). We 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 decision thresholds will be rejected sooner than lures with higher decision thresholds. This reduces the number of "active" accumulators over time. Imagine a display with blue, red and orange lures, and assume that blue is the least similar lure to the target, followed by red, then orange. In this model, blue lures would be rejected first, then red lures, and finally orange lures (on average). As an example of this model, consider the case of the equation above where there are three different types of lures in the scene (i.e., three different decision thresholds, with $D_3 > D_2 > D_1 > 0$), with N_i being the number of lures of type *i* (*i* = 1, 2, 3). The first term represents time cost for all 3 types of lure items to arrive at the evidence threshold for lure type 1 (D_1). Here, lures of type 1 are rejected. Then, evidence for lures of types 2 and 3 continues to accumulate. However, some evidence about these lures has already been accumulated (dictated by D1). Thus, the second term represents the additional time cost to arrive at the decision threshold for lures of type 2 (D_2) for lure types 2 and 3 (hence the term $D_2 - D_1$), and so on.

Equation 2:

$$T = D_1\, ln(N_1 + 1) + D_2\, ln(N_2 + 1) + D_3\, ln(N_3 + 1)$$

Equation 2 above assumes that each group of lures is considered and rejected sequentially. That is, different types of lures are processed in a serial and exhaustive fashion, while within each type of lure, individual items are processed in parallel. This model would mean that first all blue lures are processed and discarded in parallel, then the red ones, and last the orange ones. The big difference between Models 1 and 2 is that in Model 2, accumulation for red ones will start once the blues have been discarded, whereas in Model 1, accumulation for *all* types of lures starts simultaneously and rejected lures "fall off" while the other ones continue to accumulate evidence.

Equation 3:

$$T = max\{D_1, D_2, D_3\}\, ln(N_1 + N_2 + N_3 + 1) \\ = D_3\, ln(N_1 + N_2 + N_3 + 1)$$

Equation 3 represents a model that has a single decision threshold associated with the single D value in the equation. The model predicts that while all items are being processed in parallel and exhaustively, the amount of information required to complete processing is determined by the lure with the highest similarity to the target. This can be understood as there being a single decision threshold for the entire display: items below it will be discarded at the same moment, while items above the threshold will require focal inspection (i.e., are likely targets). This kind of idea has been proposed in the literature in various papers (e.g., Guided Search, Wolfe, 1994; TAM, Zelinsky, 2008). So, in the example above, when all three types of items are present in the display, the decision threshold used would be the one for orange lures, whereas when only blue and red lures are present, the decision threshold for red would be used.

Equation 4:

$$T = \frac{D_1 + D_2 + D_3}{3} ln(N_1 + N_2 + N_3 + 1)$$

Equation 4 serves as an alternative to equation 3. Here the log slope is estimated by the mean of the 3 types of lures (instead of the max), while all items are still processed exhaustively in parallel.

We note here that the above 4 equations include variations in different aspects of processing across lure types. Equation 1 is the strongest extension of our theory since it assumes both parallel processing and independence across lure types. Equation 2 assumes independence but serial processing across lure types, whereas equations 3 and 4 assumes parallel processing but with interaction between different lure types.

Simulation and Analyses

The goal of this simulation is to find out which of the 4 equations above best accounts for simulated time costs of stage-one processing in heterogeneous scenes. The critical parameters are the threshold values (representing different lure-target similarity levels), the drift rate *k* (rate of information accumulation that is sampled from the same Gaussian distribution regardless of scene context) and noise range σ.

We used two sets of parameters representing two different sets of stimuli to simulate stage-one completion times under the same model architecture. Choosing two different sets of parameters estimates allows us to be confident that our simulations and our equations are not overly dependent on any specific parameter, and that in fact, they generalize well across the parameter space. In both runs of simulations, we simulated displays containing at most three types of lure items to ensure a sufficient degree of heterogeneity without requiring too many different display conditions. Simulation no. 1 had a target item whose threshold was 20, and three types of lure items with thresholds of 15, 17, and 19. The drift rate *k* was fixed at 4 and noise range σ was also a constant 2. Simulation no. 2 had a target threshold of 62, three lure thresholds at 48, 53, and 58, with drift rate of 9 and noise range of 4. In each simulation run, threshold values, drift rate, and noise range parameters were held constant.

Given a specific set of parameters, the simulation procedure and algorithm can be described as follows:

(1) For each type of lure item, we simulated homogeneous search time as a function of set size. At set size N, there were 1 target item and N-1 lure items. The target item's processing time was found by randomly sampling from an Inverse Gaussian distribution defined by the target threshold A_t, the drift rate *k* and the noise range σ. Each of the N-1 lure items was similarly simulated by sampling from another Inverse Gaussian distribution with a lure threshold value A_l and the same *k* and σ. The overall processing time was simply the maximum of all individual items' processing times (i.e. the exhaustive processing rule). Because of the randomness in the sampling procedure, we took the mean processing time cost of 2000 repetitions at each set size as the final output.

(2) For each type of lure, we computed a regression of $RT = \hat{D}ln(N) + \hat{a}$ based on the simulated results from step (1). The estimated coefficients $\hat{D}$ and $\hat{a}$ were used for predicting heterogeneous search time cost.

(3) We simulated heterogeneous search time with different combinations of the 3 types of lures. For each simulated display condition, each type of lure could appear 1, 3, 7, or 15 times with one or two other lure types, which yielded a total of 111 unique combinations or conditions (see Appendix A for a complete list of these conditions). Processing time costs were then simulated in the same way as in step (1), with 2000 repetitions per condition.

(4) For each of the 4 approximating equations, we computed the predicted completion times for each display condition simulated in (3) using the estimated $\hat{D}$ and $\hat{a}$ coefficients from (2). We then compared predicted completion times (T) against simulated T values by computing a linear regression for each equation to estimate the equation that best fits the simulated data. We used several diagnostics for goodness of fit including the R square, log likelihood, and the slope and intercept coefficients of the regression models.

Results

We plotted simulated completion times for heterogeneous scenes against predicted completion times as scatterplots, for each equation and for both simulation runs in **Figure 2**. The y = x line is also plotted for reference. **Table 1** summarizes regression model characteristics for each equation in both simulation runs. These characteristics describe how well predicted processing times match simulated processing times. In both simulations, Equation 1 had the highest R-squares, the slope coefficient closest to 1, and the estimated intercept closest to 0. Predictions of Equation 1 also fell closest to the y = x line in **Figure 2** for both simulations.

From these results, we can conclude Equation 1 is our best performing equation for predicting heterogeneous lure search based on performance metrics from homogeneous displays. In the next section, we will consider empirical data based on human participants in both lure-homogeneous (Experiment 1) and lure-heterogeneous (Experiment 2) search tasks.

Experiment 1

Experiment 1 serves two purposes. First, it allowed us to estimate three different lure-target similarity coefficients in homogeneous displays to be used to predict performance in Experiment 2 (heterogeneous displays). In addition, it allowed us to extend the findings from Buetti et al. (2016) to real-world stimuli. Our previous results were based on two groups of simple stimuli with relatively few distinguishing features (group 1: find red triangle or blue half circle among orange diamonds or yellow triangles or

	Simulation 1				Simulation 2			
	R^2	Log likelihood	Slope (Standard Error)	Intercept (Standard Error)	R^2	Log likelihood	Slope (Standard Error)	Intercept (Standard Error)
Eq 1	0.9615	114.031	1.005 (0.019)	–0.109 (0.127)	0.9637	124.179	1.008 (0.018)	–0.152 (0.156)
Eq 2	0.8045	23.816	0.463 (0.021)	3.075 (0.163)	0.8095	32.107	0.482 (0.022)	3.871 (0.203)
Eq 3	0.7934	20.758	0.877 (0.043)	0.570 (0.291)	0.7709	21.878	0.861 (0.045)	0.905 (0.383)
Eq 4	0.8049	23.953	1.112 (0.052)	–0.652 (0.338)	0.7852	25.452	1.137 (0.057)	–1.058 (0.466)

Table 1: Model characteristics of linear regressions of simulated processing times as a function of predicted processing times.

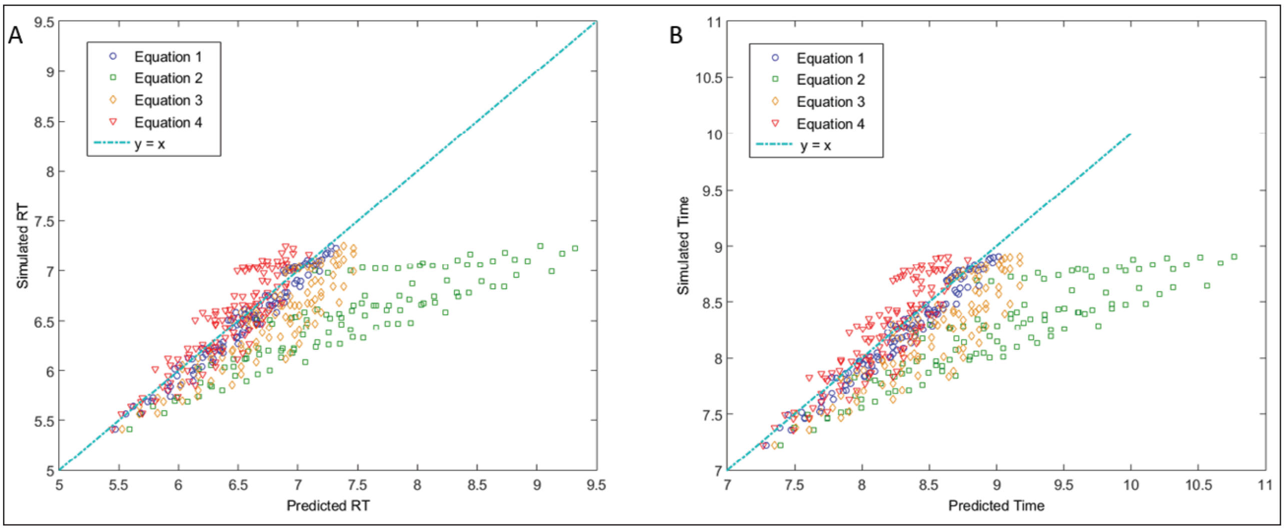

Figure 2: Predicted processing time according to the 4 approximating equations for heterogeneous processing times plotted as a function of simulated processing times. Panels A and B present results from two different simulation runs using different sets of parameters (see text for more details).

blue circles; group 2: find red T among red Ls and thin orange crosses or thick red crosses or orange crosses or orange squares; see **Figure 4** in Buetti et al., 2016 for an illustration of these stimuli).

Methods

Participants. Twenty-six participants were recruited through the Course Credit Subject Pool in the Psychology Department at the University of Illinois at Urbana-Champaign. Participants signed up for the experiment through the Department's subject pool website. Prior to participating in any experiments, participants must fill out a Screening Questionnaire that can be used by experimenters to filter out participants that do not meet recruitment criteria. In our case, we used this questionnaire to make sure only participants without self-reported color-vision deficiencies could sign up for our experiment. Upon arrival to the lab, they were also screened for normal color vision using the Ishihara color test (10 plate edition, with the standard number tests). No participants were excluded due to abnormal color vision or low visual acuity. All participants gave written informed consent before participating in the experiment. We excluded 3 participants whose overall accuracy was below 90%. For the 23 participants included in analysis, their age ranged from 18 to 24 years, 14 are female, 21 were right-handed. This experiment has been approved by the Institutional Review Board of the University of Illinois at Urbana-Champaign.

Apparatus and Stimuli. Stimuli are presented on a 20-inch CRT monitor at 85 Hz refresh rate and 1024*768 resolution. Participants sat in a dimly lit room at a viewing distance of 75 cm. The experiment was programmed with Psychtoolbox 3.0 (Kleiner et al., 2007) in the MATLAB environment, and run on 64 bit Windows 7 PCs.

Search objects were chosen from a collection of images studied by Alexander and Zelinsky (2011), which were originally sampled from Cockrill (2001) and the Hemera Photo-Objects collection. Alexander and Zelinsky (2011) obtained visual similarity ratings on these images using computational models and human subjects' subjective ratings. Using their results, we selected groups of images that were consistently rated as having high or medium similarity to the teddy bear category to be used as distractor items. Specifically, we chose a red humanoid 'carrot man', a white reindeer toy, both of which were consistently rated as highly similar to the teddy bear category, and a gray model car rated as having medium similarity. We also chose a specific teddy bear as the target item.

These images of objects were presented with sizes of approximately 1.3 degrees visual angle horizontal and 1.7 degrees visual angle vertical. All images had a small red dot overlaid on the left or right side, with a diameter of 0.2 degrees of visual angle. In each search display, there was always only one target and at most one type of lure item. The items were randomly allocated onto the screen based on an invisible 6-by-6 square grid that spanned 20 degrees of visual angle horizontally and vertically. Each item's actual location was then randomly jittered within 1 degree horizontally and vertically. On average the minimal distances between two items (i.e. the distance between two adjacent grid points) was 3.5 degrees. The grid was populated with equal (or approximately equal) numbers of items in each of the four quadrants of the screen. A white fixation cross was also presented at the center of the screen, spanning 0.6 degrees vertically and horizontally. All displays had a gray background with a color vector of [121, 121, 121] in RGB color space. **Figure 3** presents examples of search displays for Experiments 1 and 2.

Procedure. At the beginning of the experimental session, instructions were both shown on the screen and delivered verbally to participants. They were told to look for the target teddy bear (whose image was shown on the screen) and respond to whether the red dot appeared on the left or right side on the bear. They were asked to press the left arrow key with their left index finger when the red dot was on the left, and right arrow key with the right index finger when the dot was on the right. Speed and accuracy of response were equally prioritized.

Trials started with a brief presentation of the central fixation cross, with a duration randomly selected from 350 to 550 ms. Then, the search scene was displayed for a maximum duration of 2.5 seconds. The display turned blank as soon as the participant pressed a response key. On error trials, a warning tone (1000Hz sine wave lasting 250 ms) was played. The inter-trial interval was selected randomly between 1.4 to 1.6 seconds. Each experiment session started with a practice block of 32 trials.

Design. The two main independent variables, lure type, and set size, were fully crossed within-subjects. There could be 1, 4, 9, 19, or 31 lures of the same identity on the display along with one target item (so that total set sizes were 2, 5, 10, 20, or 32); additionally, there was a target-only condition where only the target image appeared on the screen. Therefore, there were a total of 3 × 5 + 1 = 16 experimental conditions. The location of the red dot on the target image was pseudo-randomized to ensure that it appeared on the left or right equally often. Locations of red dots on lure images were randomized with 0.5 probability on the left or right. Each condition was repeated 50 times so that there were 800 trials total in one experimental session. All conditions are randomly intermixed. There were short break periods every 20 trials that lasted up to 20 seconds if participants did not resume the experiment sooner.

Results

We compared regression models based on logarithmic and linear RT by set size relationships using R square and log likelihood as measures of goodness of models.[3] In order to test an alternative hypothesis that the results could be better described by a bi-linear model assuming a transition point at set size 2, we also compared the log and linear models using data without the target-only condition. These results are summarized in **Table 2**. When the target-only condition was included, logarithmic models clearly outperformed linear models, indicating that a logarithmic model is more accurate and plausible in describing the data than a simple linear model; without the target-only data point, logarithmic models still consistently had higher R-squares and log likelihoods than corresponding linear models. In **Figure 4** we plotted reaction time against set size, separating the three groups of data by lure type, along with the best-fitting logarithmic curves for each lure type. The estimated logarithmic slope coefficients for each type of lure were: $D_{\text{red carrotman}} = 66.278$, $D_{\text{white reindeer}} = 28.492$, $D_{\text{grey car}} = 26.581$. In sum, the results show that the RT by set size functions found in this experiment are best characterized by a series of logarithmic functions.

It should be noted that the estimated 'linear slopes' were very small and would be categorized as 'efficient' or 'pop-out' search according to traditions in the literature. When the target-only condition was included, the estimated linear slope coefficients were 6.380 ms/item, 2.559 ms/item, and 2.446 ms/item for carrot man, reindeer, and model car lures, respectively. Without the target-only condition, these changed to 4.531 ms/item, 1.727 ms/item, 1.503 ms/item.

A within-subjects ANOVA using lure type and set size as factors on correct RTs was also conducted. Main effects were significant for both lure type, $F(2, 44) = 265.37$, $p < 0.001$, Cohen's $f = 3.47$, and set size, $F(5, 110) = 217.69$, $p < 0.001$, $f = 3.15$. More importantly, the interaction between set size and lure type was significant, $F(10, 220) = 54.13$, $p < 0.001$, $f = 1.57$. These results indicate that the different levels of lure-target similarity lead to different magnitudes of set size effects, i.e. lure-target similarity modulated search processing efficiency. To further understand this difference in search efficiency, we also computed individual subjects' logarithmic slope estimates and use t-tests to compare the mean log slope for different pairs of lures. Consistent with the visual pattern in **Figure 4**, we found that the mean log slope for the red carrot man lure was significantly larger than both the mean log slopes for the white reindeer lure, $t(22) = 15.85$, $p < 0.001$, Cohen's $d = 3.31$, and for the grey model car, $t(22) = 15.54$, $p < 0.001$, $d = 3.24$, while there was no significant difference between the log slopes for reindeer and model car, $t(22) = 1.42$, $p = 0.17$.

Discussion

Overall our results provided evidence that a logarithmic function better captured the relationship between reaction time and set size in efficient searches with real-world stimuli than linear models. Importantly, this conclusion was not contingent upon whether the target-only condition was included in the analysis. Additionally, the steepness of the logarithmic curves depended on the similarity between target and lures: the higher similarity, the steeper or more inefficient search functions. This pattern of results extends our previous findings to real-world

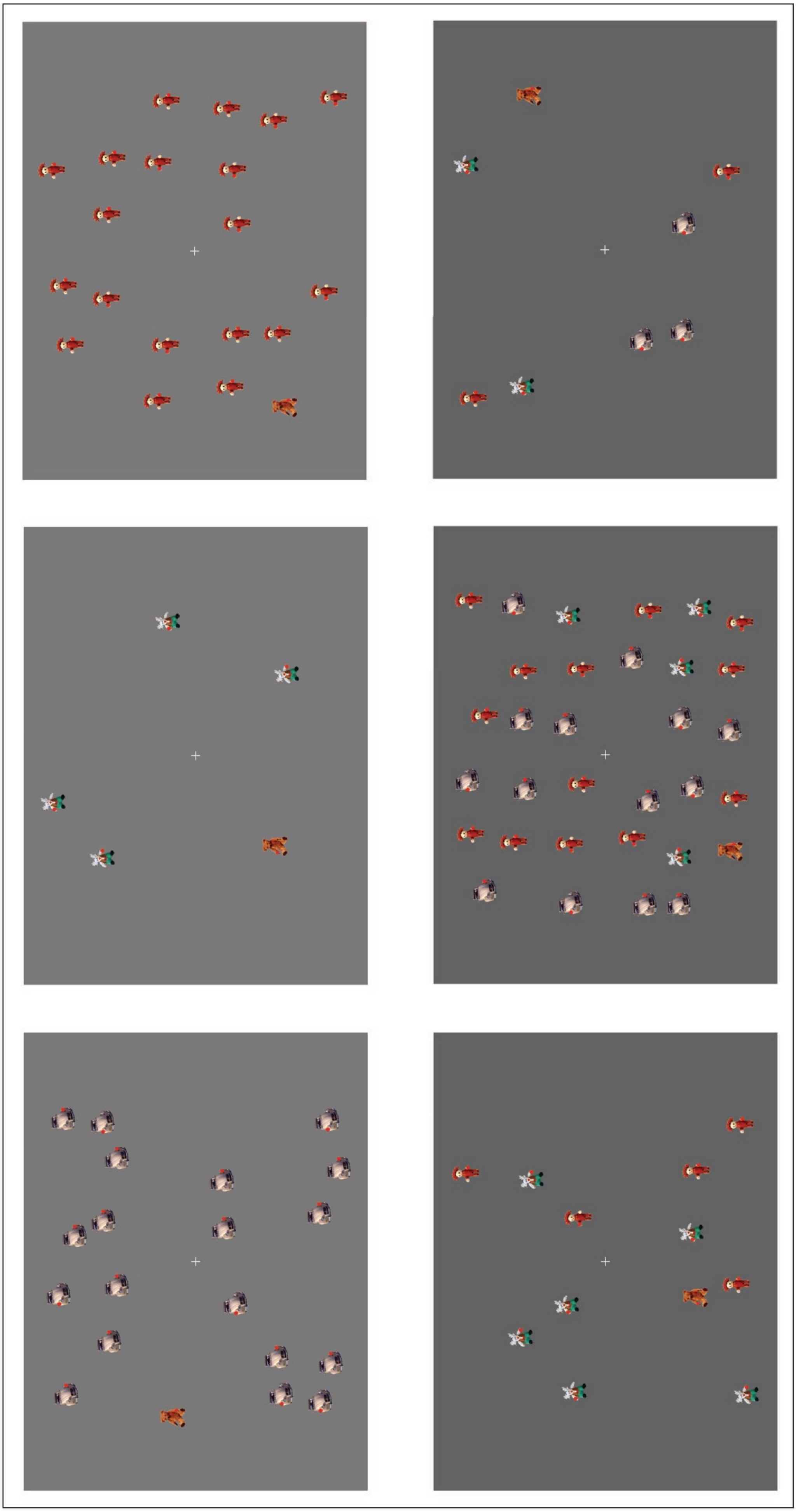

Figure 3: Example search displays for Experiment 1 (top row) and Experiment 2 (bottom row).

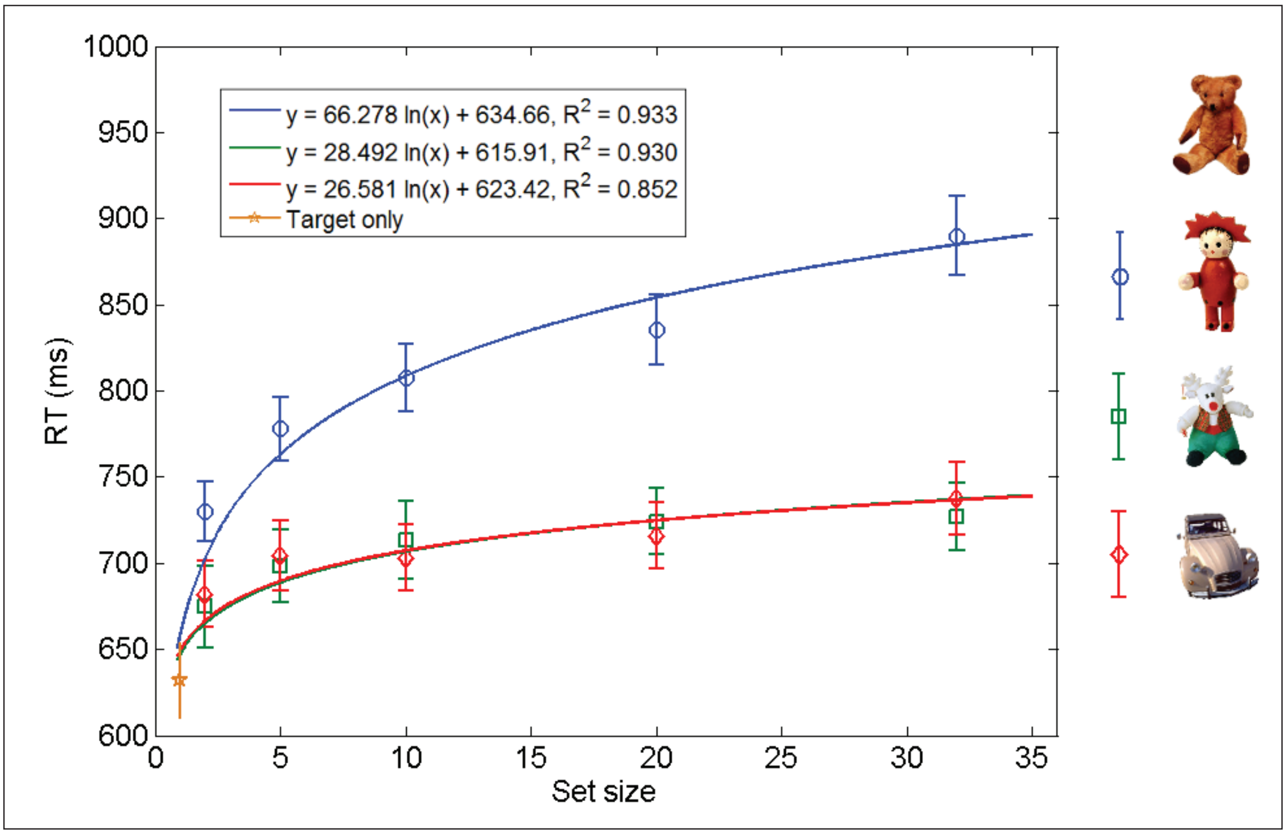

Figure 4: Reaction times for Experiment 1 plotted as a function of set size and lure type. Curves indicate best-fitting logarithmic functions. The legend shows the analytical form of each of these functions as well as corresponding R-squares as a measure of fit. Error bars indicate one standard error of the mean. Images of search stimuli and the corresponding data symbols are presented on the right.

	With target-only condition				Without target-only condition			
	Logarithmic		Linear		Logarithmic		Linear	
	R square	Log Likelihood	R square	Log Likelihood	R square	Log Likelihood	R square	Log Likelihood
Red carrot man	0.933	–27.178	0.713	–31.558	0.965	–18.718	0.924	–20.650
White reindeer	0.930	–22.251	0.619	–27.351	0.951	–15.365	0.711	–19.827
Grey model car	0.852	–24.354	0.595	–27.380	0.938	–14.636	0.919	–15.300

Table 2: Logarithmic vs. linear regression results of RT by set size functions in Experiment 1.

stimuli and corroborates the notion that visual similarity modulates early parallel visual processing, regardless of whether search objects differ from each other along a couple or multiple feature dimensions (Buetti et al., 2016). We can also conclude that all the distractor objects used in Experiment 1 are sufficiently different from the target teddy bear that they can be efficiently processed in the first, parallel stage of visual processing. We can, therefore, use these stimuli to study how this processing stage handles heterogeneous search scenes.

We should note that there is some difference between the similarity relationship reflected in our visual search results and Alexander and Zelinsky (2011)'s ratings. Our data suggested that the white reindeer and grey model car were equally dissimilar to the target teddy (i.e., their search slopes were almost identical). In contrast, in Alexander and Zelinsky's data, the reindeer was rated as being of high similarity to teddy bears, whereas the grey car was rated as having a medium similarity (we used these ratings when we first selected the stimuli for this experiment). Several factors can be identified to account for this apparent inconsistency. Ratings in Alexander and Zelinsky's study were obtained using a ranking method. Five images were presented on screen and participants had to rank-ordered them according to their visual similarity to a teddy bear (no ties allowed). Note that the influence of non-visual factors cannot be ruled out in this ranking procedure: even if a reindeer is in fact equally visually dissimilar to a teddy bear as a car, at the moment of ranking which of the two (reindeer or car) is more similar to the bear, the *conceptual* similarity of the reindeer to the teddy bear (both four-legged animals) might lead observers to give the reindeer a higher similarity rank than the car. More importantly, the requirements of the two tasks are different, so that the nature of 'similarity' computed may be different. In the ranking task, participants try to make a multidimensional decision with as much time as needed to compare stimuli. In this case, participants might decide to weight down differences along single dimensions when there is strong agreement along multiple dimensions. For instance, if a red teddy bear is to be judged in similarity against a green teddy bear, the match along shape, texture, size, and even semantic features might make participants judge them as being highly similar. In contrast, in a search task, the visual system tunes towards feature contrasts between target and distractor stimuli that can quickly locate the target in the scene. In the teddy bear example, the color contrast between the target (green teddy bear) and the distractors (red teddy bears) is likely to override the similarity along the other dimensions, such that participants will find the green teddy bear very fast and efficiently *in spite* of its overall level of similarity to the red teddies. In this sense, perhaps it's more appropriate to consider the modulation of log slopes as an effect of dissimilarity rather than similarity. In our experiment, the log slopes observed for reindeer and car lures might indicate that something along those lines is occurring. It is possible that the color contrast was similar for both target-lure pairings (the target is brown, the lures are white and grey) and that this particular feature-contrast was most responsible for locating the target in the scene.

Experiment 2

In Experiment 2, we used the same stimuli as in Experiment 1 to construct heterogeneous search displays. We then compared observed RTs with predicted RTs from the four equations described in the Predictive Simulation section. We had two goals. The first goal was to determine which of the four equations best predicted human performance. The second goal was to evaluate what kind of systematic deviations exist between our theory-based RT predictions and human data.

Because of the limited number of conditions we could afford within one experimental session (about 50 minutes), we designed five different subsets of conditions of heterogeneous displays, characterized by different types of lure combinations (see **Table 3**). We analyzed RT data from each subset of conditions separately as well as all conditions combined. This allowed us to evaluate whether different ways of mixing the lures produced different patterns of results. This experimental design and planned analyses were pre-registered on the Open Science Framework (https://osf.io/2wa8f/), including the description of the four predicting equations.

Methods

Participants. Using effect size of the two-way interaction effect in Experiment 1, we estimated that in order to achieve power of 0.8, we needed 19 subjects (effect size f = 1.5685, numerator df = 10, denominator df = 7, actual achieved power = 0.815, computed with G-Power,

	# red carrot man	**# white reindeer**	**# grey model car**	**Description**
Subset 1	0	1	2	Comparable numbers of white reindeer and gray cars
	0	3	4	
	0	7	8	
	0	15	16	
Subset 2	1	0	2	Roughly equal numbers of red carrot man and white reindeer
	3	0	4	
	7	0	8	
	15	0	16	
Subset 3	1	6	0	Fixed 6 reindeer, varying number of carrot man
	5	6	0	
	9	6	0	
	21	6	0	
Subset 4	1	1	1	Roughly equal numbers of all 3 types of lures
	2	2	3	
	5	5	5	
	10	11	10	
Subset 5	1	4	2	Fixed 4 reindeer, comparable numbers of carrot men and cars
	3	4	4	
	7	4	8	
	13	4	14	
Target-Only condition	0	0	0	Baseline

Table 3: Description of all the conditions tested in Experiment 2, organized by subset. In Subsets 1–3 only two lure types were presented in the display with the target, whereas in Subsets 4–5 always contained all 3 types of lures in addition to the target.

Faul, Erdfelder, Lang & Buchner, 2007). In anticipation of the need to replace some subjects, we collected data on 26 subjects recruited from the Course Credit Subject Pool at the University of Illinois at Urbana-Champaign. All participants gave written informed consent before participating in the experiment. The same procedure to screen for participants with normal color vision and normal (or corrected-to-normal) visual acuity was used as Experiment 1. No participants were excluded due to abnormal color vision or low visual acuity. No participants in this experiment participated in Experiment 1. Two participants were excluded from analysis because their overall accuracy was lower than 90%. For the 24 subjects included in the analysis, their age ranged from 18 to 22 years and had a mean of 19 years, 12 were female, and 22 were right-handed. This experiment has been approved by the Institutional Review Board of the University of Illinois at Urbana-Champaign.

Stimuli and Apparatus. In contrast to Experiment 1 where only one type of lure is present in each search display, displays in Experiment 2 contain 2 or 3 types of lures. These lure items were randomly intermixed across all possible spatial configurations under the constraint that each quadrant of the screen contained the same number of items. All other aspects of the stimuli and apparatus were the same for Experiments 1 and 2. See **Figure 3** for examples of search displays.

Instruction and Procedure. The experiment procedure and instructions were the same for both Experiments 1 and 2, with the exception that the practice session at the beginning of Experiment 2 had 27 trials.

Design. There were 21 total conditions in this experiment, where each condition is specified by the number of carrot men, reindeer and model cars in the display. They are organized into 5 different subsets, each exhibiting a specific kind of variation in the number of lure items, which are detailed in **Table 3**. Each condition was repeated 38 times, for a total of 798 trials. Location of the red dot on the target image was pseudo-randomized with equal probability of left or right, while for lure images they were randomized with 0.5 probability.

Notice that in both Experiments 1 and 2 we had included a target-only condition where the only item in the display is the target. We consider reaction time in this condition to be an important baseline to compare performance across both groups. Mean RT in this condition represents all the RT components that do not depend on set size, e.g. time for visual information to arrive at the cortex, response selection processes, motor response time, etc. In the case of efficient search for a target among lures, the only component depending on set size should be the stage-one processing time, which can be computed by subtracting target-only RT from RT in each of the conditions with mixtures of lures. Notice that this operation is consistent with the property of the logarithmic function, i.e. $ln\,(1) = 0$. Since the set size of target-only condition is 1, stage-one processing time is 0 under our current formulation. Thus, subtracting out target-only condition leaves us with a direct measure of stage-one processing time.

Data analysis. According to the hypotheses laid out in the introduction, the key analysis for this experiment is a linear regression of predicted RT data to observed RT data. We used the log slope coefficients for each type of lure estimated in Experiment 1. The same fixed parameters were used for all four equations and all experimental conditions (i.e., the RT in the target-only condition and the three log slope estimates from Experiment 1). To predict RTs in a specific condition, we used the numbers of each type of lures (as indicated in **Table 3**). Predicted RT for each condition is *the sum of the predicted stage-one time cost* (derived separately from Equations 1–4) *plus the mean RT of the target-only condition* (which contains the time costs for other cognitive stages such as encoding, response selection, and execution).

The analysis consisted in linear regressions for observed RTs as a function of predicted RTs for each subset of conditions. Because there were four equations to be compared, there were be four different set of predicted RT values. Each set of predicted RT values are based on all 20 non-target-only conditions of the experiment. Thus, four regressions were performed using observed RTs as the dependent variable and each set of predicted RTs as the independent variable. To compare the performance or 'goodness of fit' across the four equations, we computed the R-square and the Akaike Information Criterion (AIC; Akaike, 1974) values for each model. These computations were carried out using the fitlm() function, the R-squared and the ModelCriterion methods of the Linear Model class in Matlab.

In a further analysis focused on the best-performing equation, we analyzed whether the human data had any systematic deviations from the model. We interpret this deviation as the effect of heterogeneity/homogeneity on stage-one processing. As described before, we were interested in identifying either additive or multiplicative deviations. To estimate them, we performed six regressions on stage-one time costs obtained from the best-performing equation, one for each subset of conditions (5 total), plus one overall regression combining all conditions. In this manner, the estimated intercept coefficients become a useful indicator of any systematic lure-to-lure interaction effects in homogeneous search. If this interaction does not cause an additive difference, then the estimated intercept should be equal to zero, assuming that our best performing equation provides a truthful prediction. Hence, any substantial difference between the estimated intercept and zero becomes a measure of the magnitude of this additive effect. By the same logic, deviation of the estimated slope coefficient from 1 represents a multiplicative effect of inter-lure interaction in homogeneous search displays.

Results

Descriptive statistics. Mean RTs are plotted for each condition, grouped by Subset, in **Figure 5**. Visual inspection shows that Subset 1 was uniquely separated from the other subsets. Subset 1 was the only one that did not contain images of the carrot man stimulus. This pattern was confirmed when we performed logarithmic regressions to each Subset of data (see **Table 4**). The log slope of Subset 1 (D = 26.881) was very close to the homogeneous slope

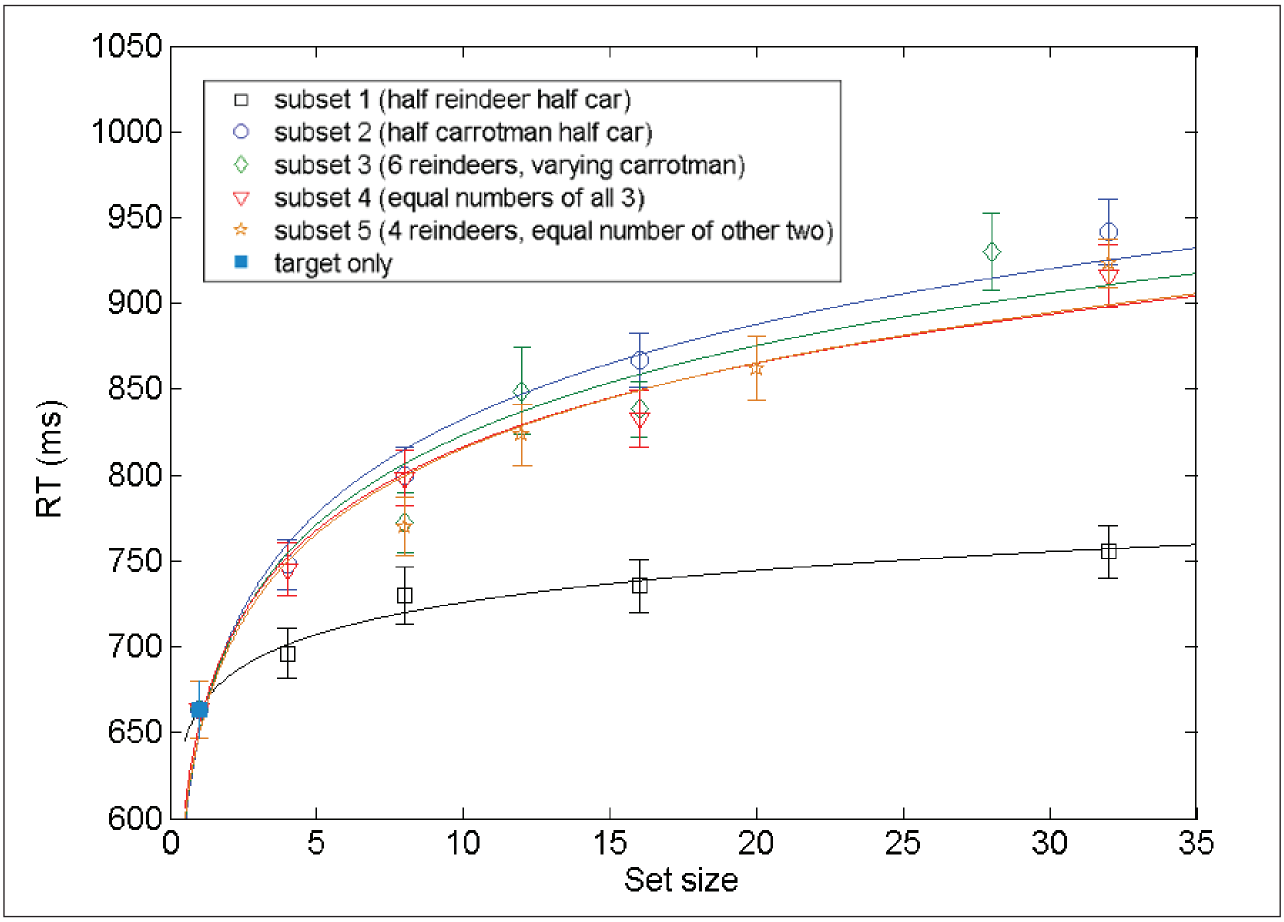

Figure 5: Reaction times in Experiment 2 as a function of set size, grouped by the different subsets of conditions. Error bars indicate one standard error of the mean. Curves are best-fitting logarithmic functions, see **Table 4** for regression model coefficients and R-squares.

	Log Slope (D)	Intercept	R square
Subset 1	26.881	663.97	0.9722
Subset 2	79.59	649.38	0.9811
Subset 3	75.183	650.16	0.9307
Subset 4	70.196	654.85	0.9801
Subset 5	71.974	649.76	0.9573

Table 4: Logarithmic regression results of search RT for each subset of conditions.

estimates of white reindeer ($D_{\text{white reindeer}}$ = 28.492) and of the grey car ($D_{\text{grey car}}$ = 26.581) from Experiment 1, in spite of the fact that in Subset 1, there were two types of lures always present in the display. In contrast, the log slopes for Subsets 2 to 5 were all *greater* than the homogeneous slope estimate of the red carrot man ($D_{\text{red carrotman}}$ = 66.278) from Experiment 1. That is, even though in each of the Subsets 2–5, the red carrot man was paired with stimuli that were *lower* in similarity to the target, the processing time was increased compared to displays containing *only* carrot man stimuli. Finally, it is worth noting that the regressions for all Subsets had very large R-squares. This indicates that for all subsets, the underlying processing was consistent with the parallel, exhaustive nature of stage one proposed by our theory. It should be noted, however, that the logarithmic RT-set size pattern for each of the Subsets is dependent on the fact that within each Subset, the proportions of each type of lure were roughly constant (by our design). If such proportion constancy is absent, there's no a priori reason to believe any group of heterogeneous search data will exhibit a logarithmic function.

Model comparison. We computed predicted RTs for each of the four equations and regressed those predicted RTs to observed mean RTs. The linear fitting R-squares, AIC values (corrected for small sample sizes) and root mean square errors (RMSE) are summarized in **Table 5**. Consistent with the results from the Predictive Simulation section, all three measures indicated that Equation 1 was the best-performing equation in terms of precision and likelihood. Specifically, the corrected AIC values indicated that the model based on Equation 1 was 166 times more likely than the second best model (Equation 3).[4] Also, the R-square value for Equation 1 (0.9681) is roughly the same as the R-square obtained in the Predictive Simulation section (0.9615, **Table 1**), when Equation 1 was used to predict simulated heterogeneous search data. This might indicate an upper-bound of predictive accuracy for this equation. The RMSE (which is an indicator of the average amount of prediction error) for Equation 1 was 14.520 ms, which compares favorably to the smallest observed standard error of mean RTs (S.E. = 14.181 ms., in the condition where there were 13 carrot men, 4 reindeer and 14 grey cars). The accuracy of Equation 1's predictions is all the more remarkable given that these predictions were based on parameters estimated from Experiment 1's data coming from a different group of participants. Further, the two groups of subjects saw qualitatively different displays. Participants in Experiment 1 only saw homogeneous displays, whereas participants in Experiment 2 never saw homogeneous displays (i.e., they only saw heterogeneous displays).

In sum, Equation 1 represents an architecture that is equally successful at predicting performance in simulations as well as in human experiments.

Estimating homogeneity effects. To investigate any potential effect of homogeneity facilitation between identical lure items, *observed* RT were first transformed to observed stage-one processing time by subtracting out the target-only RT. Then, we fitted observed stage-one processing times to predicted stage-one processing times based on Equation 1. Regressions were computed for all conditions combined as well as for each subset of conditions. The resulting coefficients are listed in **Table 6** along with standard error of estimates.

To evaluate whether there was an additive effect of homogeneity on stage-one processing times, we computed and reported 95% confidence interval of both coefficients. The regression on all 20 conditions combined had 19 degrees of freedom, whereas the regressions on each subset had 3. All 6 intercepts' confidence intervals included zero, indicating that there was no meaningful additive deviation when we predict heterogeneous search time using efficiency parameters (D values) from homogeneous search data.

Next, to evaluate whether there was a multiplicative effect of homogeneity on stage-one processing times, we can compare 95% confidence intervals of slope coefficients to 1. The results indicated that slope coefficients were significantly larger than 1 when all conditions were combined, as well as in 3 out of 5 Subsets (specifically, subsets 2, 4, 5 had confidence intervals that were larger than 1). In other words, our best-predicting equation *systematically under-predicted* stage-one processing time in heterogeneous search tasks by a *multiplicative* factor. This multiplicative factor is approximately 1.3 for this particular set of search stimuli, and can be viewed as a quantitative estimate of the pure effect of heterogeneity. This pattern of results is visualized in **Figure 6**. Recall that when making predictions using Equation 1, the log slope parameters for each lure type were estimated from homogeneous search data and that Equation 1 assumed processing independence between individual items. The most straightforward explanation of this multiplicative deviation, then, is that in Experiment 1 when adjacent lures were identical, processing of individual lure items sped up by a multiplicative factor of about 1.3, in logarithmic efficiency space. In other words, this would mean that the estimated D parameters from Experiment 1 were, in fact, *under-estimating the true 'standalone' processing efficiency of each individual lure item* because of the presence of lure-to-lure interactions in homogeneous displays. In contrast, in heterogeneous search, adjacent items are less likely to be identical and thus this type of suppression is less likely to take place and improve performance.

Discussion

Equation 1 provided the best predictions of heterogeneous search reaction time using log slope values estimated from homogeneous search data, just it did for simulated data. Thus, the predictive power of our theory was confirmed by empirical data. Therefore, Equation 1 represents a formula that will allow investigators to predict performance in heterogeneous search scenes. In the present study, Equation 1 accounted for 96.81% of the variance for a total of 20 different experimental conditions. This

	Equation 1	Equation 2	Equation 3	Equation 4
R squared	0.9681	0.9178	0.9480	0.9153
AICc	174.533	194.392	184.757	195.003
RMSE (ms)	14.520	23.298	18.522	23.639

Table 5: Linear regression results of predicted RT to observed RT in Experiment 2.

	Intercept			Slope		
	Estimate	Std. Error	95% C.I.	Estimate	Std. Error	95% C.I.
All subsets	−10.961	7.268	[−26.17; 4.25]	1.3328	0.0555	[1.22; 1.45]
Subset 1	0.948	5.986	[−18.11; 19.99]	0.9554	0.0938	[0.66; 1.25]
Subset 2	−4.392	6.585	[−25.35; 16.56]	1.3587	0.0515	[1.19; 1.52]
Subset 3	2.038	16.573	[−50.71; 54.78]	1.2063	0.1179	[0.83; 1.58]
Subset 4	−1.242	10.217	[−33.76; 31.27]	1.2905	0.0857	[1.02; 1.56]
Subset 5	−1.242	6.277	[−21.22; 18.74]	1.2890	0.0474	[1.14; 1.44]

Table 6: Regression coefficients of the regression of Equation 1's predicted stage-one processing times to observed stage-one processing times.

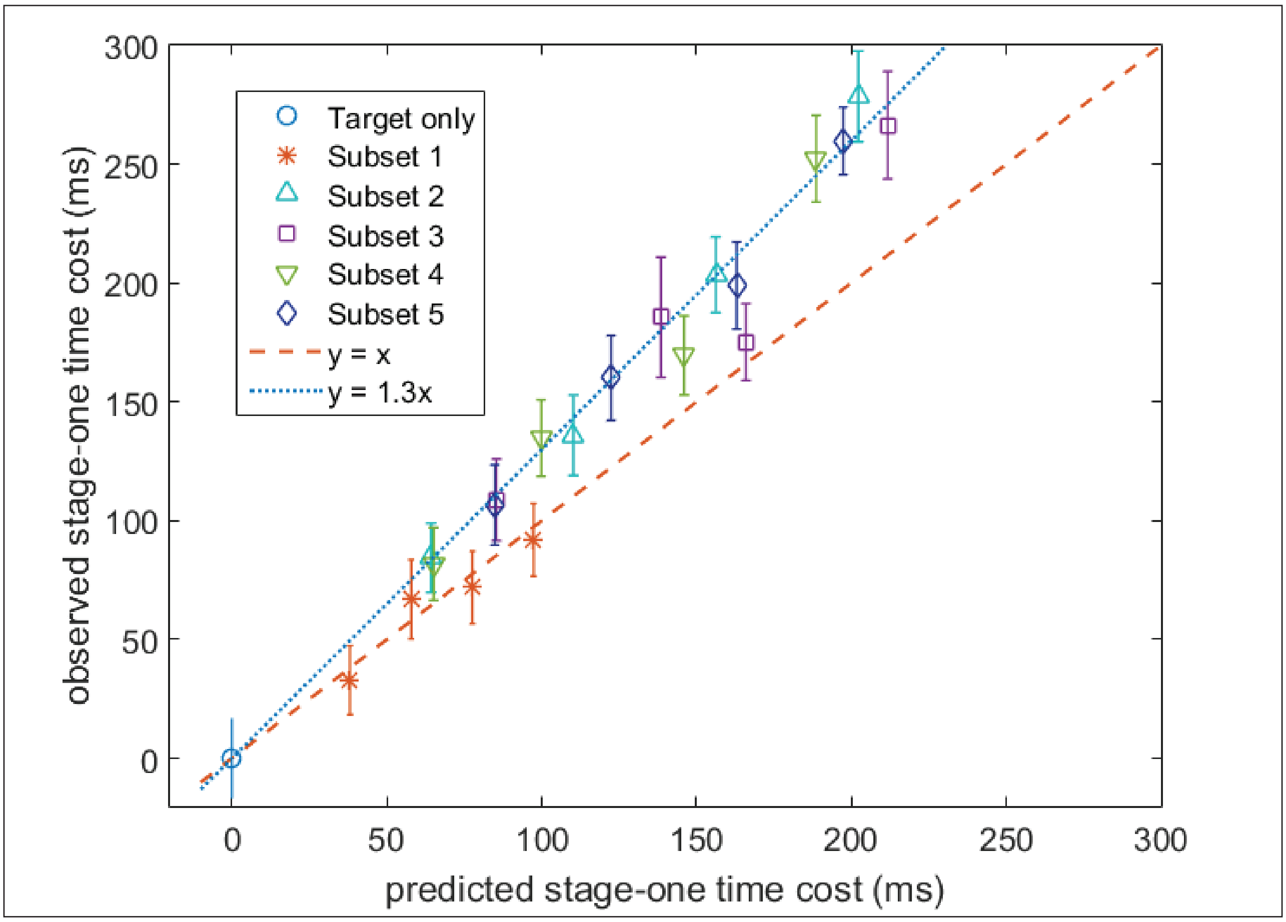

Figure 6: Predictions of Equation 1 for each Subset plotted against observed stage-one processing time. Error bars indicate one standard error of the mean. The lines $y = x$ and $y = 1.3x$ were plotted for referencing purpose. Observed stage-one processing time was computed by subtracting target-only RT from RT of other conditions. The data points for Subset 1 fitted $y = x$ line closely, while points for the other Subsets were close to the $y = 1.3x$ line, except for a couple of points in Subset 3 and 4.

predictive success is all the more compelling given that predictions were based on parameter estimates from different participants.

Further, since our simulation on both homogeneous and heterogeneous search assumed processing independence between individual items, systematic deviations from Equation 1's predictions can be used to estimate quantitatively, for the first time, the extent and effect of homogeneity facilitation in efficient search tasks. The results indicated a systematic multiplicative deviation that suggests that in homogeneous displays, identical items do interact in a facilitative fashion and are not truly independently processed. *This facilitation effect can be characterized by a constant multiplicative factor that does not depend on set size.* Because the general formula to describe the RT by set size functions in efficient visual search takes the form of $RT = a + Dln\ (N)$, where N is set size, we can infer that the facilitation effect resulted in an underestimation of the D coefficients in our Experiment 1. And since D coefficients were found to be directly related to the thresholds of accumulators, we propose that the facilitation was a result of a multiplicative lowering of the thresholds between adjacent, identical items. We discuss this finding further in the General Discussion.

The slope coefficients for all conditions combined, as well as for Subsets 2-5 all indicated a systematic multiplicative under-prediction by a factor somewhere between 1.2 and 1.3, in a fairly consistent pattern. It should be noted, however, that regression analysis on stage-one processing times for Subset 1 showed a slope coefficient that deviate from the other groups. It was much closer to 1 (estimate = 0.9554, standard error = 0.0938). Recall that Subset 1 also had a cluster of RTs that substantially differed from the other Subsets, as shown in **Figure 5**. Finally, it is also important to acknowledge that for Subset 3, although numerically larger than 1, the estimated slope coefficient was not significantly different from 1 (estimate = 1.2063, standard error = 0.1179). That said, we still view the results from Subset 3 as being in line with our interpretation of multiplicative lure-to-lure interaction effects in homogeneous displays. We think the slope coefficient failed to reach significance due to the relatively large standard error in that condition. It may also be that simply because of a lack of power, we could not simultaneously detect that all four slopes for Subsets 2, 3, 4 and 5 were greater than one (see Francis, 2012). That leaves open the question of why the slope for Subset 1 was so different from the slope of all the other subsets: why was the multiplicative effect of homogeneity absent in Subset 1? It is worth recalling that Subset 1 displays were constructed with mixtures of reindeer and model cars. Both of these lures have very low levels of similarity to the target, as indexed by their D coefficients from Experiment 1. In fact, their D values are very close to each other. This can be interpreted as reflecting that, in spite of reindeer and model cars being clearly different stimuli, they are both *equally dissimilar* to the target teddy bear.

More data is needed to understand why Subset 1's results differed from those of the other Subsets. There are at least two possible explanations. The first is that, at extremely low levels of lure-target similarity, homogeneity facilitation effects are absent. If so, the D values observed in Experiment 1 are good predictors of performance in heterogeneous displays, simply because the D values truly represented stand-alone processing efficiency of those items. A second possibility is that, for this pair of stimuli, the same lure-to-lure interaction effects are present in both homogeneous and heterogeneous display. That is, perhaps when two types of lures are *equally dissimilar* to the target, they can mutually facilitate each other as if they were identical lures. If so, D values in Experiment 1 did reflect lure-to-lure suppression effects, but these values produced accurate predictions in Experiment 2 because mixing reindeer and cars (when looking for a teddy bear) allows for reindeer and cars to mutually facilitate each other to the same extent as when they are each presented in isolation.

General Discussion

Recent work in our lab has uncovered that there is systematic variability in stage-one processing times and that much can be learned about the architecture of early visual processing by studying this variability (Buetti et al., 2016). Typically, the literature has assumed that on fixed-target efficient visual search tasks, reaction times (RTs) do not meaningfully vary as a function of set size or other display characteristics. In a series of experiments, we demonstrated that RTs increase logarithmically as a function of the number of items in the display and further, that the steepness of the log function is modulated by the similarity between the target and the distractors. In the present study, we followed-up on this research and tested four different specific computational implementations of stage-one processing that produced specific RT predictions for different visual search conditions in heterogeneous scenes. Both computer simulations and human data indicated that Equation 1 was the best performing equation to predict stage-one completion times. This equation assumed parallel, unlimited capacity and exhaustive processing, with complete inter-item processing independence, as initially proposed in Buetti et al. (2016).

Using data from homogeneous search tasks with real-world objects (Experiment 1), we were able to predict heterogeneous search RTs (Experiment 2), accounting for as much as 96.8% of the variance and with high precision, as indicated by the 14.520 ms RMSE. This prediction was made across participants: that is, parameters were estimated on one set of participants and predictions were confirmed on an entirely new set of participants that had never participated in one of these search experiments and who never saw any homogeneous displays like the ones used to estimate D parameters. The only common condition across experiments was the target-only condition. Finally, we used systematic deviations from the predictions of our model to estimate quantitatively, for the first time in the literature, the effects of homogeneity facilitation on performance in homogeneous displays, similar to the ones traditionally used in the literature to study efficient, a.k.a. pop-out, search (all elements identical but one). We found evidence that in homogeneous displays, there is a facilitatory processing effect whereby evidence thresholds are systematically reduced. This results in an improvement in overall search efficiency (in logarithmic space) in homogeneous scenes and thus, D coefficients estimated in homogeneous scenes end up under-predicting performance, by a multiplicative factor, in heterogeneous scenes where lure-to-lure interactions are absent (or much reduced).

Implications regarding homogeneity effects in search

The idea that the degree of heterogeneity (or homogeneity) in a scene influences visual search processing efficiency is not new. Duncan and Humphreys (1989) referred to it as the nontarget-nontarget similarity effect. Their claim that nontarget-nontarget similarity increases processing efficiency was based on their Experiments 3 and 4. However, in both experiments, search slopes for homogeneous displays were collapsed across different distractors and compared to heterogeneous search slopes, which were also collapsed across easy and difficult conditions in Experiment 4. Perhaps most important, there was not a direct manipulation of the degree of heterogeneity in these experiments: there was always a nearly equal number of both types of distractors in all heterogeneous scenes. Thus, the evidence in Duncan and Humphreys is in fact quite limited to an observed difference between homogeneous scenes and a specific type of heterogeneous scene (a 50-50 mix of items). Duncan and Humphreys proposed that the search slope should increase continuously as the degree of nontarget-nontarget similarity decreases (**Figure 3**, Duncan & Humphreys, 1989), but there was no direct evidence for this continuum.

In contrast, here we conducted a more systematic evaluation of heterogeneity and differences in processing between heterogeneous and homogeneous scenes. We analyzed homogeneous search separately for three different types of lures and designed displays with varying degrees of heterogeneity using those stimuli. Whereas Duncan and Humphreys suggested increasing linear search slopes with increasing degree of heterogeneity, we found that different mixtures (e.g. mixing two or three types of lure items) can be accounted for by a single constant factor (around 1.3). This result suggests that processing in heterogeneous scenes is somewhat insensitive to variations in the types of heterogeneity in those scenes. This finding somewhat contradicts the Duncan and Humphreys' spreading suppression account of homogeneity, because according to their account, the efficiency with which items are processed ought to be affected by distractor context (i.e. the composition of distractor set), whereas our results suggests it is not. Granted, Duncan and Humphreys had theorized this modulation of search efficiency as occurring in stage two, whereas, here, we only focused on changes in efficiency during stage-one processing. More data are needed to continue evaluation of these conclusions.

At the theoretical level, according to Duncan and Humphreys, items are given different attentional weights or different amounts of resources from a limited pool, depending on their similarity to the target template. Items (or 'structural units') compete for access to visual short-term memory by their weight. Further, the more items perceptually group with each other, the stronger the weights of those items will covary. This *spreading suppression* mechanism thus entails an overall bias (i.e., weight) to rejected grouped items that is a result of lower-level perceptual grouping mechanisms. Importantly, spreading suppression is *not* a form of lateral inhibition. But rather, it is a description of how the weight given to an item will "spread" to other items as a *function of the strength of grouping* between those items. Homogeneous scenes therefore produce faster RTs than heterogeneous scenes because the grouping strength amongst elements in homogeneous scenes is much stronger than in heterogeneous scenes. The spreading suppression account was thought to be a further advance from the traditional 'perceptual grouping' accounts (Bundesen & Pedersen, 1983; Farmer & Taylor, 1980) because it described *how grouping strength affected attentional priorities*. There is reason, however, to doubt this spreading suppression account, at least in this simple form, because our results (here and in Buetti et al., 2016) demonstrate that items are *not* rejected as groups in homogeneous scenes as proposed by Duncan and Humphreys. Rather, the fact that parallel search exhibits an exhaustive processing rule (and logarithmic efficiency) implies that every element in a scene matters (with each additional element contributing a non-zero cost to RT), in spite of whatever grouping effects might be observed amongst lures.

As a result, our results imply that a mechanism different from Duncan and Humphreys spreading suppression is at play in homogeneous search. We foresee at least two possible mechanisms. First, it is possible that instead of grouping similar search items, decisions are still made for each individual item, but *adjacent identical distractor items facilitate each other by reducing the amount of information needed (i.e. threshold of accumulators) to reach a decision of rejection*. This lowering of the thresholds could be due to the knowledge that only a single target exists in the display, which implies that for any-two adjacent items, the more similar they are to each other, the less likely it is that *either* of them is the target. This can be easily tested, for example, by controlling how often two identical items appear next to each other in a heterogeneous search scene. One possible extreme is when scenes consist of homogeneous regions, each containing a different type of lure, so that within each region, all adjacent items are identical to each other, and facilitation over the search scene should be maximized. The opposite extreme case would be when different types of lures are carefully 'interlaced' with each other so that adjacent lures are always different from one another. Our first hypothesis would predict that when facilitation is maximized, stage-one processing time should be nearly perfectly predicted by homogeneous search coefficients (i.e., the slope estimates reported in **Table 6** should all be close to 1). On the other hand, when such inter-item facilitation is minimized, stage-one processing time should deviate even further from the predictions based on homogeneous search data and Equation 1 (i.e., the slope estimates should be larger than the ones reported in **Table 6**).

Alternatively, the lowering of thresholds for identical items could reflect the presence of an evidence monitoring mechanism. An evidence monitoring mechanism is one that observes (i.e., monitors) how evidence accumulates at all local accumulators and sums up (or averages) evidence over all (or large) regions of the scene, much like global motion detectors sum/average local motion signals to extract a global motion direction. Applied to lure processing, as information accumulates, regions containing identical lure items will produce stronger evidence against target presence, compared to regions containing different lure items. Thus, homogeneous regions can then be discarded sooner as being unlikely to contain the target. Precise location information would not be available for these global accumulators because they represent large regions, but that is not a big problem: representing lure-location information is unnecessary for task completion, what is needed, rather, is a representation of the target location. Rejecting large regions of the display as unlikely to contain the target does help to reduce the uncertainty about the target location. Further, an advantage of such an evidence monitoring mechanism is that it can facilitate the orienting response towards regions that are more likely to contain the target (if one is present). This might happen even *before* evidence accumulation for all items within the region containing the target completes. In other words, imagine a scene where low lure-target similarity items are to the left of fixation and high lure-target similarity items are to the right (where the target is). On average, the left region will finish processing sooner than the right region. Once the left region is rejected, the eyes can start moving to the right of fixation, even before information about the specific target location is represented. This possibility too can be tested in future work.

It is also interesting to consider how our results relate to other accounts of heterogeneity effects in visual search. The Signal Detection Theory model of visual search has been shown to offer a natural explanation of the heterogeneity effect (Palmer, Verghese, & Pavel, 2000), based on the increased external noise as a result of heterogeneity in the search scene. This account assumes that representations of individual items are independent, while the heterogeneity effect arises from statistical influences of a reduced signal-to-noise ratio in a decision stage. This independence assumption is different from our current proposal as well as Duncan & Humphreys (1989)'s. The Attention based on Information Maximization model

(AIM) of visual saliency proposed by Bruce & Tsotsos (2009) also accounted for common heterogeneity effects gracefully, based on the idea that when distractor set is heterogeneous, each item is intrinsically rarer than when the set is homogeneous. Hence heterogeneity effect arises purely from bottom-up saliency computation according to AIM, without the comparison of search items to a target template. While it is not immediately clear how these models can account for our present finding that heterogeneity seems to introduce a relatively constant increase in search efficiency (within our experiment conditions), both models are highly specific and can make testable predictions with appropriate adjustments. Future work is needed to contrast these accounts with ours, although neither of the two models seem compatible with our basic finding of a logarithmic time cost function of stage-one processing.

In sum, current data present a challenge to traditional views of the distractor homogeneity effects and suggest that further study is needed to understand the mechanisms underlying this search facilitation effect. A potential avenue for further testing this facilitative interaction between homogeneous items is through the use of the capacity coefficient (Townsend & Wenger, 2004; for an example of application, see Godwin, Walenchok, Houpt, Hout, & Goldinger, 2015), which could provide more direct evidence for the violation of the independent processing in stage one.

Limitations

It is important to note that the visual search task used in this study was a target discrimination task, where there is always a target present in the display, and the participants had to locate it in order to make a decision about its details (i.e., the relative location of the red dot). This is different from the target detection task also used in the literature, where the presence or absence of a target is to be reported (e.g. Jonides & Gleitman, 1972; Treisman & Gelade, 1980), and linear search slopes are often reported. How much does the nature of the search task matter? Detection and discrimination tasks very likely induce different processing strategies, especially in the case of efficient visual search. In a target discrimination task, the need to extract details from the target compels participants to fixate or at least focus spatial attention on the target (Bravo & Nakayama, 1992), even when all distractors can be efficiently rejected or filtered out. In a target detection task, there is no such demand, and when target and distractors are highly dissimilar, the presence or absence of the target may create strong differences in the global pattern or topology of the search scene. Hence, the whole search scene could be processed as a single 'structural unit' (Duncan & Humphreys, 1989) or 'object file' (Kahneman, Treisman & Gibbs, 1992). As soon as a global topological feature is detected, a response decision can be made, and there is no need to find out the exact location of the target. Going back to the evidence monitoring mechanism proposed above, presence/absence tasks with lures that are very dissimilar from the target may be completed solely based on information at that more global level of analysis where precise locations are not represented. In contrast, in target-discrimination tasks, even if a global or regional evidence monitoring mechanism were to help reject lure regions, an observer would still have to recover precise target location information (at the local accumulators' level) to be able to make a response in a trial. That said, we should point out that in spite of these differences, there are indications in the literature (Treisman & Gelade, 1980; Palmer, Ames & Lindsey, 1993) that even in a target detection tasks, logarithmic functions can be observed if one samples set size appropriately (see also, Lleras, Madison, Cronin, Wang & Buetti, 2015).

Conclusion

In a target discrimination task with a fixed target, efficient visual search is best characterized as arising from a system that processes all items in a parallel, unlimited capacity, and exhaustive fashion. Under this conceptualization, a lawful relationship between heterogeneous and homogeneous search performance was predicted by simulation and confirmed by experiments with a novel methodology. Results indicated that, rather than being completely independent, individual items *facilitate* each other's processing when they appear in the context of other identical items. This facilitation effect can be characterized by a multiplicative factor in logarithmic space that does not change with set size. This result presents a challenge for traditional accounts of distractor homogeneity effects, like spreading suppression. These findings also extend the application of Buetti et al.'s (2016) theory to real-world objects and heterogeneous search tasks and demonstrate the computational specificity of our model of stage-one processing. Therefore, early parallel processing in visual search is non-trivial: it *systematically contributes to reaction time, plays an important role in achieving the search goal, and can be mechanistically understood.* More generally, this paper presents a novel approach for studying visual search: a predictive inference approach. While most studies in visual search draw mechanistic inferences based on descriptive data for a given set of manipulated conditions (i.e., the mean/slope in condition A is smaller than the mean/slope in condition B, therefore . . .), here we suggest that great experimental and theoretical validity is afforded by *making specific predictive inferences* to new experimental conditions. More specifically, making predictions about what processing times *ought to be* in heterogeneous displays allowed us to quantitatively estimate for the first time the effects of homogeneity facilitation, independently of other factors like lure-target similarity.

Additional File

The additional file for this article can be found as follows:

- **Appendix A.** Conditions Used in Simulation. DOI: https:/doi.org/10.1525/collabra.53.s1

Competing Interests

The authors have no competing interests to declare.

Notes

[1] An additional underlying assumption is that measures of lure-target similarity generalize across subjects, at least at the group level. Thus, we can use the estimates from one set of participants and use them to predict performance in a new set of participants.

[2] With 2 types of items, for example, the essential integral to be solved takes the following form:

$$\int\left(1-\Phi\left(\sqrt{\frac{\lambda_1}{t}}\left(\frac{t}{\mu_1}-1\right)\right)-e^{\frac{2\lambda_1}{\mu_1}}\Phi\left(-\sqrt{\frac{\lambda_1}{t}}\left(\frac{t}{\mu_1}+1\right)\right)\right)^{n_1-1}$$

$$\left(1-\Phi\left(\sqrt{\frac{\lambda_2}{t}}\left(\frac{t}{\mu_2}-1\right)\right)-e^{\frac{2\lambda_2}{\mu_2}}\Phi\left(-\sqrt{\frac{\lambda_2}{t}}\left(\frac{t}{\mu_2}+1\right)\right)\right)^{n_2}$$

$$\left(e^{\frac{-\lambda_1}{t}\left(\frac{t}{\mu_1}-1\right)^2}+e^{\frac{2\lambda_1}{\mu_1}}e^{\frac{-\lambda_1}{t}\left(\frac{t}{\mu_1}+1\right)^2}\right)tdt$$

where n_1, n_2 are the numbers of two types of items, and λ_1, μ_1, λ_2, μ_2 are corresponding parameters, $\Phi(x)$ is the CDF of standard normal distribution. We welcome any ideas or suggestions about how to find an analytical solution to this problem.

[3] The measure of log likelihood is a measure of how likely the regression model is given the observed data. The higher log likelihood value, the more likely a specific model is. The relative likelihood ratio between two models can be computed by $exp\,(L_1 - L_2)$ where L_1 and L_2 are log likelihood values.

[4] The relative likelihood ratio between two linear models can be computed using AIC values by the formula $exp\,((AIC_1 - AIC_2)/2)$. Thus the regression model based on Equation 1 was $exp\left(\frac{194.392-174.533}{2}\right)=20527.07$ times more likely than the Equation 2 based model, $exp\left(\frac{184.757-174.533}{2}\right)=166.00$ times more likely than the Equation 3 based model, and $exp\left(\frac{195.003-174.533}{2}\right)=27861.47$ times more likely than the Equation 4 based model.

References

Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, *19*(6), 716–723. DOI: https://doi.org/10.1109/TAC.1974.1100705

Alexander, R. G., & **Zelinsky, G. J.** (2011). Visual similarity effects in categorical search. *Journal of Vision*, *11*(8), 9–9. DOI: https://doi.org/10.1167/11.8.9

Awh, E., Barton, B., & **Vogel, E. K.** (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological science*, *18*(7), 622–628. DOI: https://doi.org/10.1111/j.1467-9280.2007.01949.x

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

Breitmeyer, B. G. (1992). Parallel processing in human vision: History, review, and critique.

Bruce, N. D., & **Tsotsos, J. K.** (2009). Saliency, attention, and visual search: An information theoretic approach. *Journal of vision*, *9*(3), 5–5. DOI: https://doi.org/10.1167/9.3.5

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

Bundesen, C. (1990). A theory of visual attention. *Psychological review*, *97*(4), 523. DOI: https://doi.org/10.1037/0033-295X.97.4.523

Bundesen, C., & **Pedersen, L. F.** (1983). Color segregation and visual search. *Perception & Psychophysics*, *33*(5), 487–493. DOI: https://doi.org/10.3758/BF03202901

Bylinskii, Z., Judd, T., Durand, F., Oliva, A., & **Torralba, A.** (2014). Mit saliency benchmark.

Chhikara, R. (1988). *The Inverse Gaussian Distribution: Theory: Methodology, and Applications* (Vol. 95). CRC Press.

Chong, S. C., & **Treisman, A.** (2005a). Statistical processing: Computing the average size in perceptual groups. *Vision research*, *45*(7), 891–900. DOI: https://doi.org/10.1016/j.visres.2004.10.004

Chong, S. C., & **Treisman, A.** (2005b). Attentional spread in the statistical processing of visual displays. *Perception & Psychophysics*, *67*(1), 1–13. DOI: https://doi.org/10.3758/BF03195009

Cockrill, P. (2001). *The teddy bear encyclopedia*. New York: DK Publishing.

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

Farmer, E. W., & **Taylor, R. M.** (1980). Visual search through color displays: Effects of target-background similarity and background uniformity. *Perception & Psychophysics*, *27*(3), 267–272. DOI: https://doi.org/10.3758/BF03204265

Faul, F., Erdfelder, E., Lang, A. G., & **Buchner, A.** (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, *39*(2), 175–191. DOI: https://doi.org/10.3758/BF03193146

Fei-Fei, L., Iyer, A., Koch, C., & **Perona, P.** (2007). What do we perceive in a glance of a real-world scene? *Journal of vision*, *7*(1), 10–10. DOI: https://doi.org/10.1167/7.1.10

Francis, G. (2012). Publication bias and the failure of replication in experimental psychology. *Psychonomic Bulletin & Review, 19*(6), 975–991. DOI: https://doi.org/10.3758/s13423-012-0322-y

Godwin, H. J., Walenchok, S. C., Houpt, J. W., Hout, M. C., & **Goldinger, S. D.** (2015). Faster than the speed of rejection: Object identification processes during visual search for multiple targets. *Journal of experimental psychology: human perception and performance*, *41*(4), 1007. DOI: https://doi.org/10.1037/xhp0000036

Haberman, J., & **Whitney, D.** (2009). Seeing the mean: ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 718. DOI: https://doi.org/10.1037/a0013899

Itti, L., & **Baldi, P.** (2009). Bayesian surprise attracts human attention. *Vision research*, *49*(10), 1295–1306. DOI: https://doi.org/10.1016/j.visres.2008.09.007

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

Itti, L., & **Koch, C.** (2001). Computational modelling of visual attention. *Nature reviews neuroscience*, *2*(3), 194–203. DOI: https://doi.org/10.1038/35058500

Jonides, J., & **Gleitman, H.** (1972). A conceptual category effect in visual search: O as letter or as digit. *Perception & Psychophysics*, *12*(6), 457–460. DOI: https://doi.org/10.3758/BF03210934

Kahneman, D., **Treisman, A.**, & **Gibbs, B. J.** (1992). The reviewing of object files: Object-specific integration of information. *Cognitive psychology*, *24*(2), 175–219. DOI: https://doi.org/10.1016/0010-0285(92)90007-O

Kleiner, M., **Brainard, D.**, **Pelli, D.**, **Ingling, A.**, **Murray, R.**, & **Broussard, C.** (2007). What's new in Psychtoolbox-3. *Perception*, *36*(14), 1.

Kruthiventi, S. S., **Ayush, K.**, & **Babu, R. V.** (2015). DeepFix: A Fully Convolutional Neural Network for predicting Human Eye Fixations. *arXiv preprint arXiv:1510.02927*.

Li, F. F., **VanRullen, R.**, **Koch, C.**, & **Perona, P.** (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences, 99*(14), 9596–9601. DOI: https://doi.org/10.1073/pnas.092277599

Lleras, A., **Madison, A.**, **Cronin, D.**, **Wang, Z.**, & **Buetti, S.** (2015). Towards a better understanding of the role of parallel attention in visual search. *Journal of vision*, *15*(12), 1255–1255. DOI: https://doi.org/10.1167/15.12.1255

Luck, S. J., & **Vogel, E. K.** (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281. DOI: https://doi.org/10.1038/36846

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

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

Nordfang, M., & **Wolfe, J. M.** (2014). Guided search for triple conjunctions. *Attention, Perception, & Psychophysics*, *76*(6), 1535–1559. DOI: https://doi.org/10.3758/s13414-014-0715-2

Oliva, A. (2005). Gist of the scene. *Neurobiology of attention*, *696*(64), 251–258. DOI: https://doi.org/10.1016/B978-012375731-9/50045-8

Palmer, J., **Ames, C. T.**, & **Lindsey, D. T.** (1993). Measuring the effect of attention on simple visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(1), 108. DOI: https://doi.org/10.1037/0096-1523.19.1.108

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

Parkes, L., **Lund, J.**, **Angelucci, A.**, **Solomon, J. A.**, & **Morgan, M.** (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature neuroscience*, *4*(7), 739–744. DOI: https://doi.org/10.1038/89532

Pashler, H. (1992). Attentional limitations in doing two tasks at the same time. *Current Directions in Psychological Science*, *1*(2), 44–48. DOI: https://doi.org/10.1111/1467-8721.ep11509734

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of experimental psychology: human learning and memory*, *2*(5), 509. DOI: https://doi.org/10.1037/0278-7393.2.5.509

Potter, M. C., & **Levy, E. I.** (1969). Recognition memory for a rapid sequence of pictures. *Journal of experimental psychology*, *81*(1), 10. DOI: https://doi.org/10.1037/h0027470

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), 14–14. DOI: https://doi.org/10.1167/12.4.14

Schyns, P. G., & **Oliva, A.** (1994). From blobs to boundary edges: Evidence for time-and spatial-scale-dependent scene recognition. *Psychological science*, *5*(4), 195–200. DOI: https://doi.org/10.1111/j.1467-9280.1994.tb00500.x

Shapiro, K. L., **Raymond, J. E.**, & **Arnell, K. M.** (1997). The attentional blink. *Trends in cognitive sciences, 1*(8), 291–296. DOI: https://doi.org/10.1016/S1364-6613(97)01094-2

Sigman, M., & **Dehaene, S.** (2008). Brain mechanisms of serial and parallel processing during dual-task performance. The Journal of neuroscience, 28(30), 7585–7598. DOI: https://doi.org/10.1523/JNEUROSCI.0948-08.2008

Sperling, G. (1960). The information available in brief visual presentations. *Psychological monographs: General and applied*, *74*(11), 1. DOI: https://doi.org/10.1037/h0093759

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

Townsend, J. T., & **Wenger, M. J.** (2004). A theory of interactive parallel processing: new capacity measures and predictions for a response time inequality series. *Psychological review*, *111*(4), 1003. DOI: https://doi.org/10.1037/0033-295X.111.4.1003

Treisman, A. M., & **Gelade, G.** (1980). A feature-integration theory of attention. *Cognitive psychology*, *12*(1), 97–136. DOI: https://doi.org/10.1016/0010-0285(80)90005-5

Verghese, P. (2001). Visual search and attention: A signal detection theory approach. *Neuron*, *31*(4), 523–535. DOI: https://doi.org/10.1016/S0896-6273(01)00392-0

Vig, E., Dorr, M., & **Cox, D.** (2014). Large-scale optimization of hierarchical features for saliency prediction in natural images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2798–2805. DOI: https://doi.org/10.1109/cvpr.2014.358

Vogel, E. K., Luck, S. J., & **Shapiro, K. L.** (1998). Electrophysiological evidence for a postperceptual locus of suppression during the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance, 24*(6), 1656. DOI: https://doi.org/10.1037/0096-1523.24.6.1656

Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychonomic bulletin & review*, *1*(2), 202–238. DOI: https://doi.org/10.3758/BF03200774

Wolfe, J. M., Alvarez, G. A., Rosenholtz, R., Kuzmova, Y. I., & **Sherman, A. M.** (2011). Visual search for arbitrary objects in real scenes. *Attention, Perception, & Psychophysics*, *73*(6), 1650–1671. DOI: https://doi.org/10.3758/s13414-011-0153-3

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*(6), 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. DOI: https://doi.org/10.1037/a0013118

Zhang, J., & **Sclaroff, S.** (2013). Saliency detection: A boolean map approach. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 153–160. DOI: https://doi.org/10.1109/iccv.2013.26

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.53.pr

How to cite this article: 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

Submitted: 22 August 2016 **Accepted:** 19 November 2016 **Published:** 10 March 2017

UNIVERSITY of CALIFORNIA PRESS | Collabra

Collabra: Psychology is a peer-reviewed open access journal published by University of California Press.

OPEN ACCESS