AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: Human Perception and Performance

ISSN: 0096-1523

2021, Vol. 47, No. 9, 1274–1297
https://doi.org/10.1037/xhp0000941

Distractor–Distractor Interactions in Visual Search for Oriented Targets Explain the Increased Difficulty Observed in Nonlinearly Separable Conditions

Zoe (Jing) Xu, Alejandro Lleras, Yujie Shao, and Simona Buetti
Department of Psychology, University of Illinois at Urbana Champaign

The linear separability effect refers to a benefit in search performance observed in a feature-search task, where target and distractor features vary along a continuous feature dimension: Search performance is best when there is a boundary in feature space that separates the distractor features from the target feature. However, the role that distractor heterogeneity plays in this effect is not well understood. Here, we reexamined this effect in the context of a new predictive procedure from Lleras et al. (2019) that quantifies the impact of distractor heterogeneity on search performance. Experiments 1A and 1B measured people's performance in homogeneous search conditions where they searched for the target among one type of distractor. The parameters observed in Experiments 1A and B were then used to predict search times in Experiments 2 and 3, where the target was presented in heterogeneous displays containing two types of distractors. The results show that total variance accounted for was 95% to 98%, without including any factor indexing the linear separability rule. The results demonstrate that heterogeneous search in orientation space is a function of target-distractor similarity and interitem interactions. The study highlights the robustness of the predictive procedure and demonstrates the generalizability of the method to estimate interitem interactions to new stimulus types.

Public Significance Statement
This study demonstrates that there is no linear separability effect in orientation feature space, contrary to suggestions from previous studies. Visual search performance in conditions where the target feature is nested among distractor features in orientation space (nonseparable condition) does not qualitatively differ from search performance in homogeneous conditions. The slowdown in the linearly nonseparable search conditions is attributable to a reduction in interitem interactions that facilitate performance in homogeneous conditions.

Keywords: heterogeneity effects, interitem interactions, linear separability effect, prediction, visual search

Visual search performance suffers both in terms of speed and accuracy when the target is accompanied by distractors that are different from one another (heterogeneous search) compared with when the target is embedded within a field of identical distractors (homogeneous search). Perhaps because of the longer reaction times and increased perceived difficulty, homogeneous and heterogeneous search display also feel very different at the phenomenological level (compare top with bottom rows in Figure 1). However, recent studies have proposed that the mechanisms underlying homogeneous search are not fundamentally different from the ones underlying heterogeneous search because search performance in heterogeneous search can be almost perfectly predicted by parameters observed in homogeneous search conditions (Lleras et al., 2019; Wang et al., 2017). These authors proposed that homogeneous and heterogeneous searches differ mainly in the strength of interitem (or distractor–distractor) interactions, which operate to facilitate distractor rejection.

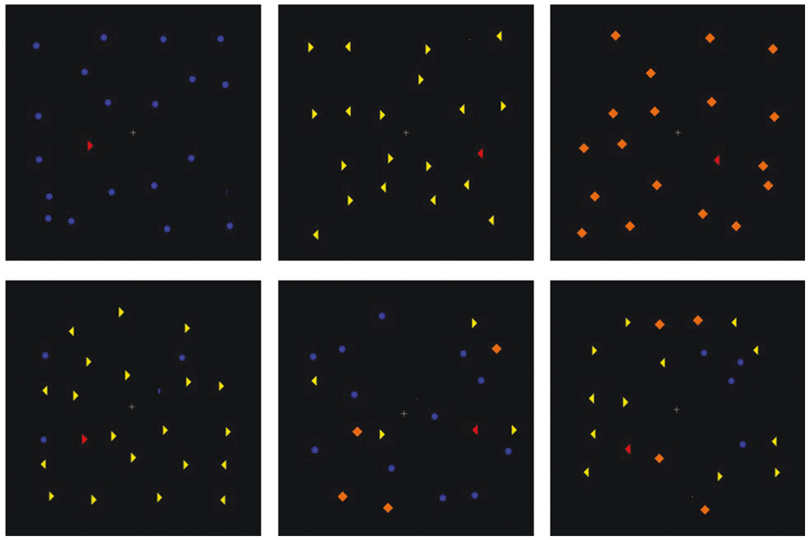
Zoe (Jing) Xu https://orcid.org/0000-0002-2649-418X
Alejandro Lleras https://orcid.org/0000-0003-0391-1355
Simona Buetti https://orcid.org/0000-0002-1718-450X

Zoe (Jing) Xu, Simona Buetti, and Alejandro Lleras contributed to writing and interpretation of results. Simona Buetti and Alejandro Lleras contributed to the design and methodology development. Yujie Shao contributed to software. Yujie Shao and Zoe (Jing) Xu contributed to the investigation. Zoe (Jing) Xu contributed to the formal analysis. The data and code are available on OSF at https://osf.io/wra7h/.

This project was supported by a National Science Foundation Grant to Simona Buetti (Award BCS 1921735). The authors have no competing interests.

Correspondence concerning this article should be addressed to Zoe (Jing) Xu, Department of Psychology, University of Illinois at Urbana Champaign, 603 East Daniel Street, Champaign, IL 61820, United States. Email: jingxu9@illinois.edu

Figure 1
Samples of Homogeneous and Heterogeneous Search Displays in Lleras et al. (2019)



Note. The top row shows samples of homogeneous search displays. The bottom row shows samples of heterogeneous search displays. The target is the red triangle pointing to the left or right. See the online article for the color version of this figure.

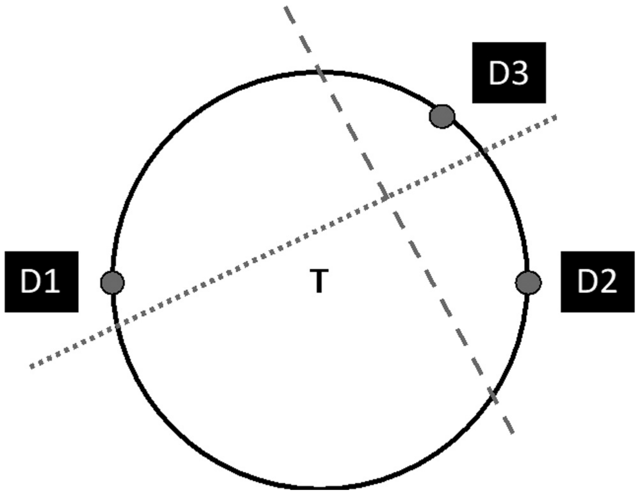
Although homogeneous searches are clearly easier than heterogeneous searches, some heterogeneous conditions are particularly harder than others. To explain the increased difficulty in these harder heterogeneous conditions, novel stimulus-related factors, such as the rule of "linear separability" between target and distractor features, have been proposed (see Bauer et al., 1996a, 1996b, 1998, 1999; Figure 2 for illustration, D'Zmura, 1991). Here we tested the validity of this rule using a model comparison approach that bypasses null hypothesis significance testing and focuses instead on comparing the predictive accuracy of competing models. The models use parameters observed in homogeneous search conditions to make specific reaction time (RT) predictions in various heterogeneous search conditions. The goal was to test whether there is indeed a qualitative difference between easier and harder heterogeneous search conditions or whether both searches rely on the same underlying mechanisms governing homogeneous search.

The Linear Separability Effect

The linear separability effect refers to a benefit in search performance observed in heterogeneous feature-search tasks, where the target and different distractors vary along a continuous feature dimension. According to the linear separability effect, search performance is best when there is a *linear boundary in the feature space that separates the distractor features from the feature that defines the target*. Search is qualitatively more difficult when there exists no such linear boundary separating the target from the distractors (see Figure 2, Arguin & Saumier, 2000; Bauer et al., 1996a, 1996b, 1998, 1999; Blais et al., 2009; D'Zmura, 1991; Hodsoll & Humphreys, 2001; Saumier & Arguin, 2003). Initially, D'Zmura (1991) found the linear separability effect in a CIE color space defined by a green-red axis and a yellow-blue axis. When people searched for a target among two types of distractors that could be linearly separated from the target in this color space, the search slope was flat, reflecting an efficient, parallel search process; in contrast, when the distractors could not be linearly separated from the target in this color space, the search slope was much steeper, which was interpreted as reflecting a serial search process. Later Bauer et al. (1996b) replicated this effect using the CIELUV color space. The linear separability effect was also found when stimuli were defined in orientation space (Blais et al., 2009; Rosenholtz, 2001; Wolfe et al., 1992) and in shape space (e.g., varying curvature & aspect ratios, Arguin & Saumier, 2000; Saumier & Arguin, 2003; varying size, Hodsoll & Humphreys, 2001).

Duncan and Humphreys (1989) demonstrated that search difficulty is determined by two factors: (a) target-distractor similarity and (b) distractor–distractor similarity, also referred to below as distractor heterogeneity. Search becomes more difficult as target and distractors become more *similar* and as distractors become more *dissimilar* to one another. Search difficulty is typically indexed by the steepness of the RT by set size function, aka, the search slope. Although some of the early work on linear separability controlled for target-distractor similarity (Arguin & Saumier, 2000; Bauer et al., 1996b), distractor heterogeneity still remained a potential factor to explain the effect. That is, when equating target-distractor similarity for the two distractor types, the linearly nonseparable condition always has larger distractor heterogeneity than the linearly separable condition. To illustrate this, imagine a circle in a feature space, centered at the target feature, with the two distractors features moving on the periphery along a fixed radius circle, so that the target-distractor similarity is kept the same. Linear separability is only violated (i.e., the target cannot be separated from the distractors) when the two distractors are on opposite

Figure 2
Illustration of the Linear Separability Rule



Note. The circle illustrates the position of features with identical target-distractor similarity in feature space with respect to a target feature located at the center of the circle. According to the linear separability rule, searching for the target T among two different types of distractors is easy if a decision line can be drawn in feature space to separate the target feature from the distractor features. In this example, searching for T among D_1 and D_3 will be easy because the two distractor features are on one side of the dotted line and the target feature is on the other. The same is true when searching for T among D_2 and D_3 distractors (dashed line separation). On the other hand, searching for T among D_1 and D_2 will be hard because no line can be drawn in this space to separate target from distractor features. In the latter example, note that the distance between the two distractor features is maximal, indicating that the dissimilarity between D_1 and D_2 is the largest.

positions on the circle, which is when the distance between these items is also the largest (the distance being an indication of how dissimilar they are from one another, see Figure 2). As demonstrated by Duncan and Humphreys (1989), when the distractor heterogeneity increases, the search also becomes more difficult. Therefore, distractor heterogeneity offers an alternative explanation to why linear nonseparable displays produce steeper search slopes than linear separable displays.

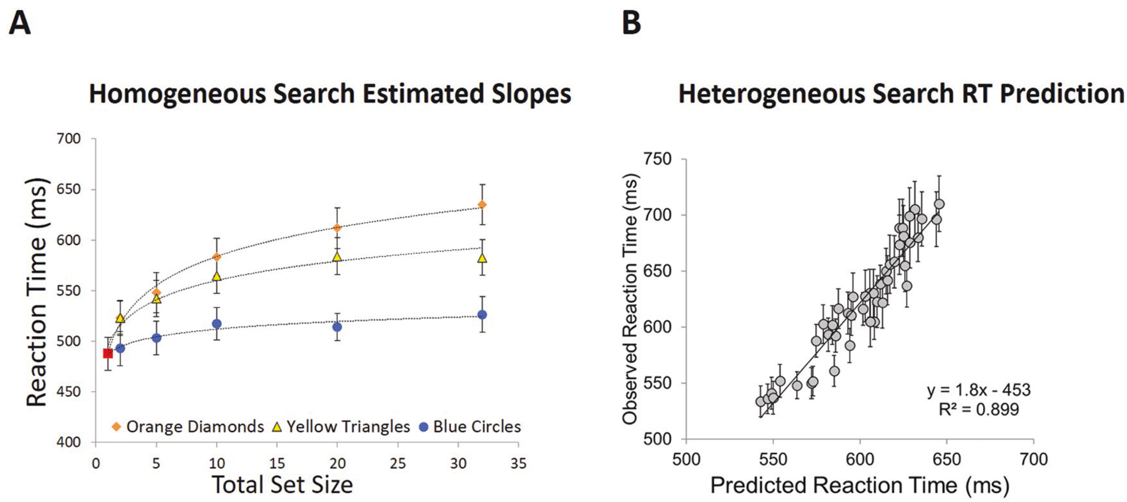
Some studies have tried to evaluate the contribution of distractor heterogeneity to the linear separability effect. For instance, Bauer and colleagues (1996a) conducted a series of experiments in which they attempted to control for distractor heterogeneity by manipulating the ratio of the two types of distractors presented on the display. The logic followed results from Poisson and Wilkinson's (1992), who had found that search times in a conjunction search task were affected by the ratio of the two types of distractors in the display. Bauer et al. (1996a) attributed these distractor-ratio effects on RTs to changes in distractor heterogeneity and thus used manipulations of distractor ratios as a manipulation of distractor heterogeneity. While controlling for distractor heterogeneity across conditions via distractor ratios, the results indicated the same linear separability effect (i.e., the linear nonseparable condition elicited steeper search slopes than the linearly-separable condition). However, although the authors attempted to address the distractor heterogeneity confound through a distractor ratio manipulation, the correspondence between distractor heterogeneity and distractor ratio was not directly tested, casting some doubts on the conclusion that the linear separability effect does not arise from distractor heterogeneity.

More recently, Vighneshvel and Arun (2013) used a model comparison approach to reevaluate whether target-distractor and distractor–distractor similarity are sufficient to explain the linear separability effect, without including a linear-separability rule. The authors used four sets of stimuli: crescent-like shapes varying in curvature and thickness, triangular shapes varying in pointiness and curvature, real-world objects, and English alphabet letters. The approach used was the following. First, for each stimulus set, reaction times in four homogeneous search conditions were evaluated: (a) search for the target among distractors of type I, (b) search for the target among distractors of type II, and (c–d) search for distractors of type I among distractors of type II, and vice-versa. Set size was kept constant across all conditions (i.e., 32). The authors then used both search RTs and the reciprocal of search RTs (1/RT) obtained in these different homogeneous search conditions to predict heterogeneous search time performance, where participants searched for the target among both types of distractors simultaneously. Search RT was used as an index of similarity between stimuli. The use of 1/RT was warranted based on previous work by Arun (2012), showing that 1/RT has useful properties as a quantitative measure and is a relatively good index of the distance between target and distractors in feature space, that is, of target-distractor dissimilarity. Six predictive models were compared. In Model 1, search performance in the heterogeneous conditions was predicted by the similarity index (RT) from the most similar distractor to the target. Model 2 used the similarity of the two distractors to the target. Model 3 was like Model 2 but also included the measure of distractor–distractor similarity. Models 4–6 followed the same rationale but used the 1/RT metric to predict 1/RT performance in the heterogeneous conditions. The results showed that the models based on the reciprocal of RT outperformed the models based on simple search RT. Furthermore, across all four sets of stimuli, the winning model included all three indexes of dissimilarity (Model 6). The results demonstrated that distractor heterogeneity plays a critical role in explaining performance in heterogeneous search conditions because most of the variability in the heterogeneous search performance was predicted by target-distractor dissimilarity and distractor–distractor heterogeneity (r = .91; R^2 = 83%). Thus, the authors concluded that there was no need to incorporate an additional linear separability factor.

In spite of the overall success of Vighneshvel and Arun's (2013) study, there are some shortcomings that need to be considered. First, the methodology is based on measures that are collected at a fixed set size. As a result, the dissimilarity metric (1/RT) is impractical because the dissimilarity between the same two stimuli will be different for different levels of set size, which consequently harms the generalizability of the model. Further, because set size is not manipulated, the measure 1/RT also compounds search-related processing time with nonsearch related processing time, such as response selection, response execution, and so forth. One problem with this confound is that stimuli that have better stimulus–response compatibility mappings will have smaller RTs, and thus inflated dissimilarity metric ratings, that are

Figure 3

Search Efficiencies Observed in Homogeneous Conditions From Buetti et al. (2016) and Search Time Predictions in Heterogeneous Conditions from Lleras et al. (2019)



Note. (A) Search efficiencies (i.e., logarithmic search slopes) observed in homogeneous search when searching for a red triangle among either blue circles, yellow triangles, or orange diamond. (B) Observed reaction times in heterogeneous search tasks as a function of predicted reaction times. By using the parameters observed in homogeneous search applied to Equation 1, 90% of the variance was accounted for in 45 different heterogeneous search conditions. Search Efficiencies Observed in Homogeneous Conditions is adapted from "Towards a better understanding of parallel visual processing in human vision: Evidence for exhaustive analysis of visual information," by S. Buetti, D. A. Cronin, A. M. Madison, Z. Wang, & A. Lleras, 2016, *Journal of Experimental Psychology: General, 145*(6), p. 680. Copyright [2016] by APA. Adapted with permission. Search Time Predictions in Heterogeneous Conditions adapted from "Predicting search performance in heterogeneous scenes: Quantifying the impact of homogeneity effects in efficient search," by A. Lleras, Z. Wang, A. Madison, S. Buetti, R. Zwaan, & J. Henderson, 2019, *Collabra: Psychology, 5*(1), p. 5. Copyright [2019] by Collabra: Psychology. Adapted with permission. See the online article for the color version of this figure.

unrelated to the actual dissimilarity relation to the distractors they are measured against (Madison et al., 2020). Finally, Arun (2012) showed that the 1/RT metric suffers from a range limitation. That is, the dissimilarity metric is linearly related to 1/RT only over a range of dissimilarity values; at larger dissimilarity values, the measure saturates.

In spite of the findings questioning the existence of a linear-separability effect (Vighneshvel & Arun, 2013), many researchers in the visual search literature continue to believe in this effect, likely biased by the earlier papers (Arguin & Saumier, 2000; Bauer et al., 1996a, 1996b, 1998, 1999; D'Zmura, 1991; Saumier & Arguin, 2003) and perhaps wanting stronger evidence against it.

A Predictive Approach to Understand Search

To test for the validity of the linear separability rule, the present study used the predictive approach first developed by Wang et al. (2017) to study heterogeneity effects in visual search. This approach allows one to evaluate the extent to which factors indexing target-distractor similarity in homogeneous search conditions can predict performance in distractor-heterogeneous search conditions.

The approach follows three steps. First, performance is evaluated in efficient search conditions with only one distractor type present at a time (i.e., the homogeneous search condition), as illustrated for instance in Figure 1 (top row). The goal is to estimate search efficiency—indexed by the logarithmic slope D—for each specific target-distractor similarity level (Figure 3A).

Second, these search efficiencies from Step 1 are then used to predict search performance in heterogeneous search conditions based on Equation 1, when two or three distractor types are simultaneously present in the display (Figure 1 bottom row). As shown in Lleras et al. (2020), Equation 1 represents the closest mathematical solution when one is trying to compute how long it will take to process in parallel N_T distractors (i.e., all locations are processed simultaneously), in a stochastic fashion, where each distractor is processed independently and at its given rate, reflecting its own level of target-distractor similarity. Distractors of type I (N_1) are associated with search efficiency D_1, distractors of type II (N_2) are associated with search efficiency D_2, and so forth (see Lleras et al., 2020; for full details). These search efficiencies are estimated during Step 1, using homogeneous displays. Lleras et al. (2020) proposed that the processing at each location involves accumulating evidence to reject peripheral distractors as nontargets. The evidence being accumulated at each location is the contrast between the item and the target template, and contrast accumulates stochastically. Several factors influence the contrast accumulation rate, such as target-distractor similarity (the larger the similarity, the smaller the contrast, and the slower an item will be rejected) and stimuly (slower accumulation rates as eccentricity increases).

$$RT_{Predicted} = a + \sum_{j=1}^{L}(D_j - D_{j-1}) \times \ln\left(N_T - \left(\sum_{i=1}^{j-1} N_i\right) \times 1_{[2,\infty)}(j) + 1\right) \tag{1}$$

In Equation 1,[1] L represents how many types of distractors there are in the heterogeneous search display being predicted; N_T represents the total number of distractors in the display; N_i represents the number of *type i* distractors; and D_j represents the observed logarithmic slope associated with *type j* distractor from homogeneous displays. As the *type j* distractor becomes more similar to the target, the value of D_j becomes larger. D_j values are organized from smallest (D_1) to largest (D_L), with $D_0 = 0$. Note that the logarithmic slope D is an index of target-distractor similarity and is inversely proportional to the evidence accumulation rate in a drift diffusion process aimed at rejecting peripheral distractors as nontargets (see Lleras et al., 2020). Distractors that are least similar to the target template (smaller Ds, shallower search functions) have larger evidence accumulation rates and are rejected faster than distractors that are similar to the target (larger Ds, steeper search functions), for which accumulation rates are smaller. In Figure 3A, D_1 would correspond to the search slopes associated with blue circle distractors, D_2 the slope associated with yellow triangle distractors and D_3 the slope for orange diamond distractors. The constant a represents the response time when there is only the target and no distractor in the scene. Finally, the index function 1 $[2, \infty]$ (j) means that the sum over i applies only for terms when $j > 1$. That is, when there are multiple distractors in the display, this sum term keeps track of the number of distractors in the display that have been rejected over previous iterations of j. When $j = 1$, this sum term is zero, which indicates that initially all distractors contribute to RT.

In the final step, predicted RTs are then plotted against the Observed RTs that were obtained in separate experiments with the corresponding heterogeneous conditions. As shown in Wang et al. (2017; see also Lleras et al., 2019), the resulting graph is a linear function (see Figure 3B), of the type:

$$RT_{Observed} = C + \beta \times RT_{Predicted} \tag{2}$$

C and β are parameters that are optimized to maximize the fit of the model predictions to the observed RTs, as with any linear regression. β is therefore a free parameter that is only observed in this final step. As $RT_{Predicted}$ in Equation 2 is replaced by the formula in Equation 1, it becomes clear that β amounts to a multiplicative factor that modulates the time cost incurred to reject the distractors:

$$RT_{Observed} = (C + \beta * a) + \beta * \sum_{j=1}^{L}(D_j - D_{j-1}) \times \ln\left(N_T - \left(\sum_{i=1}^{j-1} N_i\right) \times 1_{[2,\infty)}(j) + 1\right) \tag{3}$$

Note that the term in the first parenthesis is a constant that is independent of the number of distractors. Regarding the interpretation of β, Lleras et al. (2019) proposed that β is a quantitative estimate for the strength of distractor–distractor interactions. This conclusion was reached by comparing two heterogeneous search conditions: one where distractors were randomly intermixed around the display and a second one where distractors were segregated by type, such that all distractors of one type would be on the same side of the display. As can be intuited, performance in the intermixed condition was much slower than in the segregated condition. However, Equation 2 correctly accounted for the overwhelming majority of the variance across both conditions (average R^2 in the intermixed conditions: 93%; average R^2 in the segregated conditions: 95%), indicating that target-distractor similarity factors governed both easy segregated search conditions and hard intermixed search conditions. The difference across the two arrangement conditions was entirely accounted for by changes in the β value. In the segregated condition, β was close to 1 (.9) for both simple geometric shapes and real-world objects; whereas in the intermixed condition, the value of β changed as a function of stimulus complexity (1.8 for simple geometric shape stimuli, 1.3 for more complex real-world objects). In sum, when the value of β is close to 1, similar strengths interactions are being observed in homogeneous and heterogeneous conditions. When the value of β is larger than 1, it indicates a slow-down in the processing rate of the distractors in the heterogeneous condition compared with the homogeneous condition. The larger the β value is, the stronger the slow-down.

This approach was successful at predicting people's performance in heterogeneous performance using both geometric shapes (Lleras et al., 2019; Figure 3B) and images of real-world objects (Lleras et al., 2019; Wang et al., 2017).

The Present Study

The goal of the present study was to reevaluate the existence of the linear separability effect using the same predictive approach as Wang et al. (2017) and Lleras et al. (2019) to better quantify the contribution of distractor heterogeneity to the slow down typically observed in nonlinearly separable conditions (Experiment 2). In addition, the approach was also applied to linearly separable conditions (Experiment 3) to demonstrate that these conditions follow the same processing rules as the nonlinearly separable conditions. We also tested a second nonlinearly separable condition (Experiment 4) to confirm our findings and validate the mathematical logic behind the model. Note that the approach used in the present study presents several advantages compared with the one used by Vighneshvel and Arun (2013). First, heterogeneous search times predictions rely on the logarithmic search slopes observed in homogeneous distractor conditions (D values in Equation 1). This parameter indexes only the search component of RT, and not some of the nonsearch processes included in RT and 1/RT measures. It is also a measure of target-distractor similarity that is invariant to set size (by definition). As a result, this similarity index will not change when set size changes. Previous studies have also demonstrated that the D value for a specific target-distractor pair remains stable across groups of participants and across studies, and can therefore be used to predict performance in novel groups of participants and in novel search conditions (Buetti et al., 2019; Lleras et al., 2019; Wang et al., 2017). Finally, the logarithmic slope is theoretically grounded as it indexes the average rejection time for a specific distractor type, given a fixed target.

[1] Equation 1 is consistent with two different models of parallel processing: the one initially put forward by Buetti et al. (2016) and a more recent improved revision of the model proposed in Lleras et al. (2020).

Experiment 1A and 1B aimed at evaluating the logarithmic search efficiency (*D* value parameters) when searching for a target among a set of homogeneous distractors. The stimuli from Experiment 1A were then used to construct heterogeneous displays in Experiments 2 and 4, where the target was nonlinearly separable from the distractors. The stimuli from Experiment 1B were used to construct heterogeneous displays in Experiment 3, where the target was linearly separable from the distractors. Different participants completed Experiments 1A, 1B, 2, 3 and 4. Predictions were made using Equation 1 and two other models to provide a comparison (see Equations 5 and 6)

We made the following predictions. If the linear separability effect exists, then Equation 1 should completely fail to predict the performance in the heterogeneous conditions tested in Experiments 2–4. This follows because Equation 1 does not take into consideration the relationship in feature space between the target and the two distractors. Only the similarity relationship between the target and each distractor *j* is taken into account (D_j). If Equation 1 succeeds at predicting search performance in the heterogeneous displays of Experiments 2–4, this would support the idea that neither easy (linearly separable) nor hard (nonlinearly separable) heterogeneous searches are *qualitatively different* than searching for the same target among homogeneous search displays where only one distractor type is present.

Beyond the linear-separability effect, this study also represents a further test of the validity and generalizability of the predictive approach used by Wang et al. (2017) and Lleras et al. (2019). In particular, the nonlinearly separable condition (which is the hardest of the oriented line search conditions) represents the toughest test of the predictive power of Equation 1 thus far. This is because in our previous studies on distractor heterogeneity, there were no expectations from the literature that intermixing items would somehow change search performance in a qualitatively different manner, quite the way this expectation exists in orientation search.

Finally, from a methodological standpoint, this type of predictive approach sidesteps many of the current issues with Null-Hypothesis Significance Testing. Indeed, the main goal is to evaluate how precisely one can predict performance in novel, complex conditions, based on independent estimates made from somewhat easier conditions. The success of a model is then quantified by how much variance it can predict, rather than by whether or not a critical statistic beats the .05 alpha level. Another advantage of this method is that the same data set can be examined multiple times. Indeed, we make our data publicly available and other investigators can try to fit novel models to the same data and perhaps discover better models than the ones tested here, without the need of further data collection, nor the worries of doing multiple NHST tests on the same data.

Method

The methods and experimental protocols were approved by the Institutional Review Board at the University of Illinois, Urbana-Champaign, and are in accordance with the Declaration of Helsinki.

Participants

Participants for Experiments 1A, 1B, 2, and 3 were recruited from the University of Illinois at Urbana Champaign in exchange of course credit. Participants for Experiment 4 were recruited from either the University of Illinois at Urbana Champaign or Prolific, in exchange for course credit or pay ($6 for 50 minutes). Previous experiments in our lab (e.g., Buetti et al., 2016; Lleras et al., 2019; Madison et al., 2018; Wang et al., 2017) have shown that a sample size of 20 participants is sufficient to obtain an accurate estimate of logarithmic search slopes in homogeneous search (e.g., in Buetti et al., 2019; the mean standard error of the slopes was 3.25 ms/log unit). Note that this sample size consideration differs from the standard ones that focus on statistical power. Because our goal is not to test a null hypothesis but rather to make specific RT point predictions, what we care more about is that our parameter estimates are precise.

Because of scheduling constraints, we initially ran 25 participants in Experiment 1A (3 male, 22 female; mean age = 18.9, age range = 18–22), 24 participants in Experiment 1B (2 male, 22 female; mean age = 19.5, age range = 18–22), 26 participants in Experiment 2 (one participant's demographic information missing because of experimenter error, for the rest 25 participants: eight male, 17 female; mean age = 19.8, age range = 18–28), 24 participants in Experiment 3 (11 male, 13 female; mean age = 19, age range = 18–21), and 25 participants in Experiment 4 (eight male, 17 female; mean age = 23.5, age range = 18–31).

For Experiments 1–2, accuracy rate was calculated as the percentage of trials where participants made a correct response divided by the total number of trials. Because Experiments 3 and 4 were conducted online, we anticipated potential technical problems such as unstable Internet connection and environmental distractions. The accuracy rate was calculated as the percentage of the trials where participants made a correct response divide by the number of trials where participants made a response, i.e., we excluded time-out trials as errors because it was impossible to ascertain the reason for the time out. In Experiment 1A, three participants with an accuracy rate lower than 90% were excluded. In Experiment 1B, one participant who did not complete the experiment and two participants with an accuracy rate lower than 90% were excluded. In Experiment 2, because search difficulty was higher, seven participants with an accuracy rate lower than 70% were excluded. In Experiment 3, one participant who self-reported Internet connection problem and two participants with an accuracy rate lower than 90% were excluded. In Experiment 4, one participant who self-reported not understanding the instruction and four participants with an accuracy rate lower than 90% were excluded.

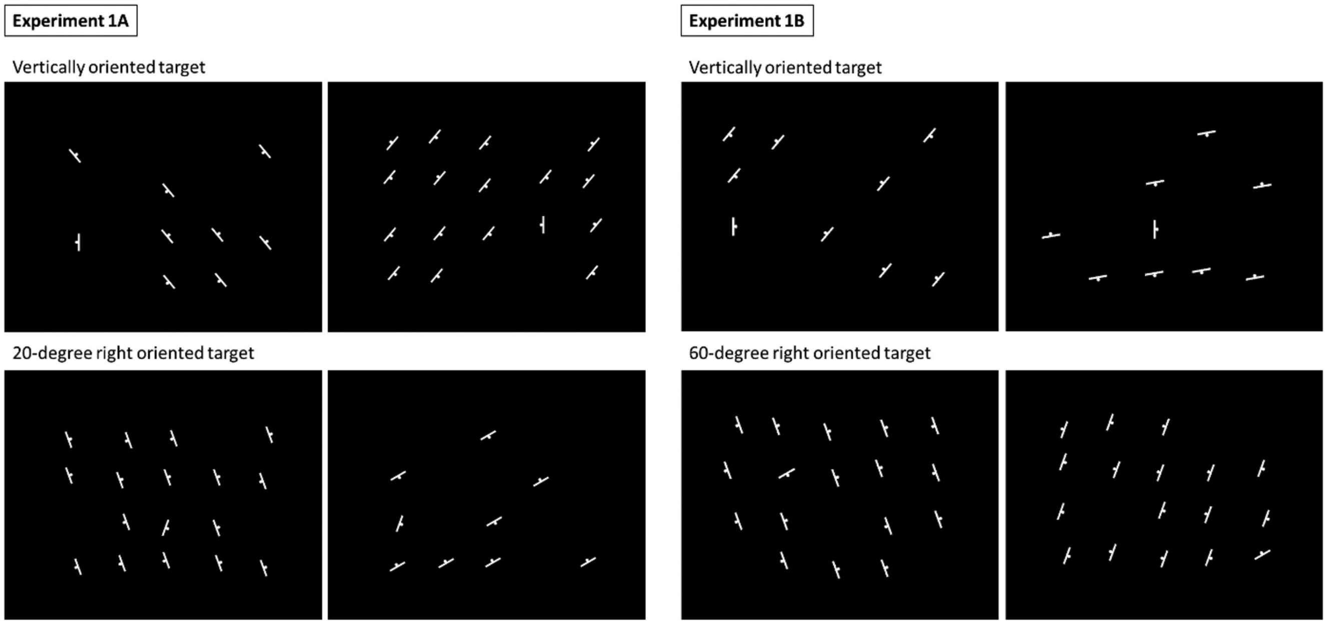
Thus, the data analysis included 22 participants in Experiment 1A (group accuracy = .97, *SD* = .022), 21 participants in Experiment 1B (group accuracy = .98, *SD* = .016), 19 participants in Experiment 2 (group accuracy = .86, *SD* = .081), 21 participants in Experiment 3 (group accuracy = .98, *SD* = .019), and 20 participants in Experiment 4 (group accuracy = .98, *SD* = .022).

Stimuli and Design

Experiments 1A, 1B, and 2

All stimuli were presented on a 20-in. CRT monitor at an 85Hz refresh rate and 1024x768 resolution. The experiments were programmed using Psychopy3 and run on 64-bit Windows 7 PCs. The target and distractors were all white rectangular bars with a white dot on either the left or right of the long side, 46 pixels × 13

Figure 4
Examples of Search Displays in Experiment 1A and 1B



Note. Left: Example of homogeneous search displays in Experiment 1A when the target was a vertical line (top panel) among −40-degree distractors (left) and 40-degree distractors (right), and when the target was a line oriented 20 degrees to the right (bottom panel) among −20-degree distractors (left) and 60-degree distractors (right). Right: Example of homogeneous search displays in Experiment 1B when the target was a vertical line (top panel) among 40-degree distractors (left) and 80-degree distractors (right), and when the target was a line oriented 60 degrees to the right (bottom panel) among −20-degree distractors (left) and 20-degree distractors (right).

pixels, and about 1.9 × .6 degrees of visual angle (height × width). Stimuli were randomly placed on the display based on an invisible 5 × 4 cells rectangular grid occupying the entire 20-in. display (40 degrees of visual angle). The background was black.

In Experiment 1A and 1B, we evaluated search performance in separate blocks for two targets, the order being counterbalanced across participants (sample displays are shown in Figure 4).

The displays in these experiments were always homogeneous, meaning that the target was accompanied by one type of distractors (except for the target-only condition). Target type was blocked, and the order was counterbalanced across participants. The goal for running two target conditions in each experiment was to maximize the chances of finding a stimulus set where the two distractors produced different slope (D) values. Indeed, if the two slope values are too similar to one another, predictions made by Equation 2 would be difficult to distinguish from a model where all distractors are treated in identical fashion (see the Appendix for a mathematical discussion of this issue).

In Experiment 1A, the first target was a vertical line (0-degree) that was surrounded by 40-degree left or right oriented lines (referred to as −40 and 40-degree distractors). The second target was a 20-degree right oriented line that was surrounded by 20-degree left oriented lines or 60-degree right oriented lines (referred to as −20 or 60-degree distractors). These conditions were inspired by previous work by Wolfe et al. (1992) that kept the target-distractor distance in orientation space the same (here 40 degrees).

In Experiment 1B, the first target was a vertical line (0-degree) that was surrounded by 40-degree or 80-degree right oriented lines (referred to as 40 or 80-degree distractors). The second target was a 60-degree right oriented line that was surrounded by 20-degree left oriented lines or 20-degree right oriented lines (referred to as −20 or 20-degree distractors).

In Experiments 1A and 1B, for each distractor type, there were five possible set sizes: 0, 2, 4, 8, and 16. In total, there were 20 conditions that were repeated 40 times for a total of 800 trials.

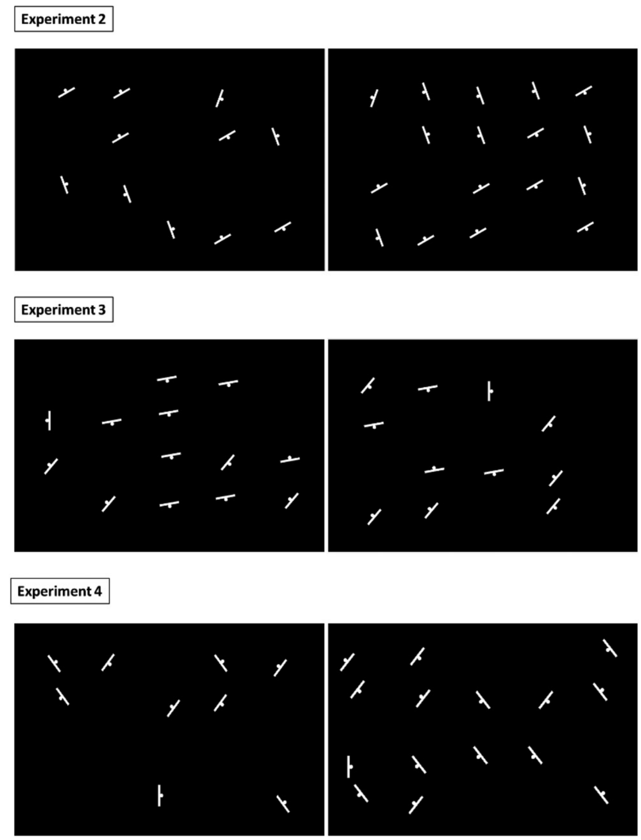
In Experiment 2 (nonlinearly separable heterogeneous search), we used the 20-degree right oriented target stimulus, accompanied by −20-degree and 60-degree distractors. There was a target-only condition. The number of distractors of each type varied independently among four possible values (2, 4, 6, 8), in a fully crossed design. In other words, the total set size varied from 1 (target only condition) to 17 (1 target + 8 distractors of each kind) and the same total set size could be achieved in different combinations of the two distractors (e.g., eight total distractors could result from four of each kind, or two of one kind and six of the other). In total there were 16 possible combinations of the two distractor types, resulting in 17 conditions that were repeated 30 times for a total of 510 trials. The displays in Experiment 2 were always heterogeneous (except for the target-only condition), meaning that there were always two types of distractors in any display.

Experiments 3 and 4

Because of COVID19, data collection in the laboratory was terminated, and Experiments 3 (linearly separable heterogeneous search) and 4 (nonlinearly separable heterogeneous search) were programmed in JavaScript and conducted on Pavlovia, with participants using their own computers remotely. Stimuli were as in Experiment 1A and 1B. Because Experiments 3 and 4 were run online, we had no control over the visual angle of the stimuli on participants' computers. To compensate for this, at the beginning of the experiment we asked participants to rescale the image of a credit card to match the real size of a credit card, so that we could at least ensure that the stimuli across different computer platforms and displays always had the same physical size (1.38 × 0.39cm). Stimuli were randomly placed on the display based on an invisible rectangular 5 by 4 grid occupying an area of 25 × 14cm on the center of participants' screen. This size was chosen to allow participants with screen as small as 12.5 in. to see the full display.

In Experiment 3, we used the vertical (0-degree) target stimulus, accompanied by 40-degree and 80-degree distractors. In Experiment 4, we used the vertical (0-degree) target stimulus, accompanied by −40-degree and 40-degree distractors. The study design in Experiments 3 and 4 was identical to the one used in Experiment 2. Sample displays of Experiments 2–4 are shown in Figure 5.

Figure 5
Examples of Search Displays in Experiments 2–4



Note. Top: Examples of heterogeneous search displays in Experiment 2, where the target (a 20-degree line) was nonlinearly separable from the distractors. Two types of distractors (−20 and 60-degree) were simultaneously presented on the display. Middle: examples of heterogeneous search displays in Experiment 3, where the target (a vertical line) was nonlinearly separable from the distractors. Two types of distractors (40 and 80-degree) were simultaneously presented on the display. Bottom: Examples of heterogeneous search displays in Experiment 4, where the target (a vertical line) was nonlinearly separable from the distractors. Two types of distractors (−40 and 40-degree) were simultaneously presented on the display. In all these experiments, there were always two types of distractors present in the display (aside from trials in the target-only condition) and the number of distractors of each type (2, 4, 6, 8) was fully crossed.

Procedure

At the beginning of each trial, participants were shown a white fixation cross at the center of the screen, followed by the search display. Participants were asked to search for the target and report the left or right location of the dot by pressing the left or right arrow keys on the keyboard. Each display lasted for 4 (Experiment 1A and 1B) or 5 (Experiments 2–4) seconds, or until the participants pressed the response key, whichever occurred earlier. In Experiments 1A, 1B and 2, if the response was wrong, participants heard a beep sound for .2 second. In Experiments 3 and 4, visual feedback of "Correct!" or "Wrong!" was given after each trial, lasting for .5 second. In all experiments, the trial ended with a black background shown for a random interval of .8–1.2 seconds.

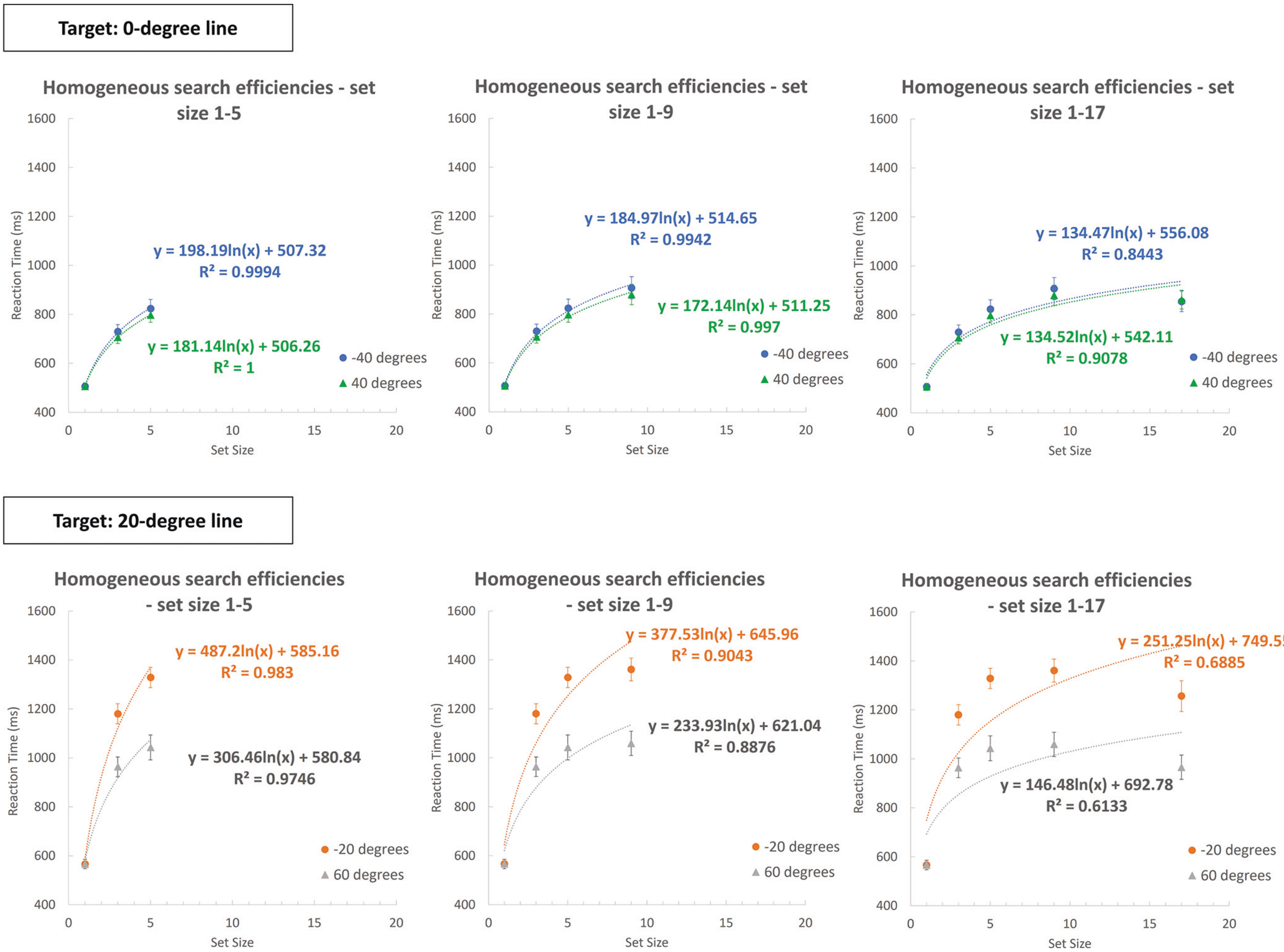
Results

Experiment 1: Search Efficiency in Homogeneous Search

Only correct trials were included in the analysis. Figures 6 and 7 show the logarithmic slopes for each target-distractor combination used in Experiments 1A and 1B, respectively. In both experiments, the target only condition was the anchor point for the logarithmic slopes. In the right-most panels of Figures 6 and 7, one can notice that at larger set sizes, RTs begin to decrease, likely as a result of some sort of texture facilitation. It is unclear whether a logarithmic function should be fit through these data. Therefore, we extracted the logarithmic slopes D using three ranges of set sizes: set size 1 to 5 (i.e., 1, 3, 5; note that RTs monotonically increase in this range), set size 1 to 9 (i.e., 1, 3, 5, 9), and set size 1 to 17 (i.e., 1, 3, 5, 9, 17). The logarithmic fits are best when only the smaller set sizes are considered and as more set sizes are added, there is a systematic decrease in the fit of the logarithmic function for both distractor types. The logarithmic search slopes for the three ranges of set sizes observed in Experiments 1A and 1B are also reported in Table 1.

In Experiment 1A, the set of distractors associated with the vertical target produced very similar search slopes. For instance, when considering set sizes 1–17, the slopes were the same (134.5 ms/log unit of set size) for the −40-degere and 40-degree distractors (Figure 6 top). On the other hand, the set of distractors associated with the 20-degree right oriented line target produced sufficiently different slope values to carry out the predictions in Experiment 2 (Figure 6 bottom). For instance, when considering set sizes 1–17, the slopes for the −20-degee and 60-degree distractors were 251 and 146 ms/log unit of set size, respectively. Therefore, in Experiment 2 we used the stimuli associated with the 20-degree target to be able to use the model comparison approach. We also ran an analogous heterogeneous Experiment (Experiment 4) based on the vertical target with −40-degree and 40-degree distractors to validate our conclusions in yet another nonseparable condition configuration, with the caveat that we expected the

Figure 6
Logarithmic Search Efficiency in Experiment 1A



Note. Top: Reaction times observed in the homogeneous search conditions of Experiment 1A as a function of distractor set sizes and distractor types when the target was a vertical line (0-degree) and the distractors were −40- and 40-degree lines. Bottom: Reaction times observed in the homogeneous search conditions of Experiment 1A as a function of distractor set sizes and distractor types when the target was a 20-degree right oriented line, and the distractors were −20- and 60-degree lines. The left panels show the results when set sizes 1–5 are considered to compute the logarithmic slope, and the middle and right panels, the results for set sizes 1–9 and 1–17, respectively. Dotted lines show the best logarithmic fit. Error bars indicate one standard error of the mean. See the online article for the color version of this figure.

results to be less discriminating between different models because of the similarity in numerical value between the slopes for the two types of distractors (see the Appendix for more on this point).

In Experiment 1B, the set of distractors associated with the vertical target produced substantially different search slopes. For instance, when considering set sizes 1–17, the slopes for the 40-degree and 80-degree distractors were 123 and 49 ms/log unit of set size, respectively (Figure 7 top); The difference between slopes was smaller for the set of distractors associated with the 60-degree target (83 and 99 ms/log unit of set size, Figure 7 bottom). In Experiment 3 we used the stimuli that produced the larger slope difference (vertical target with 40- and 80-degree distractors).

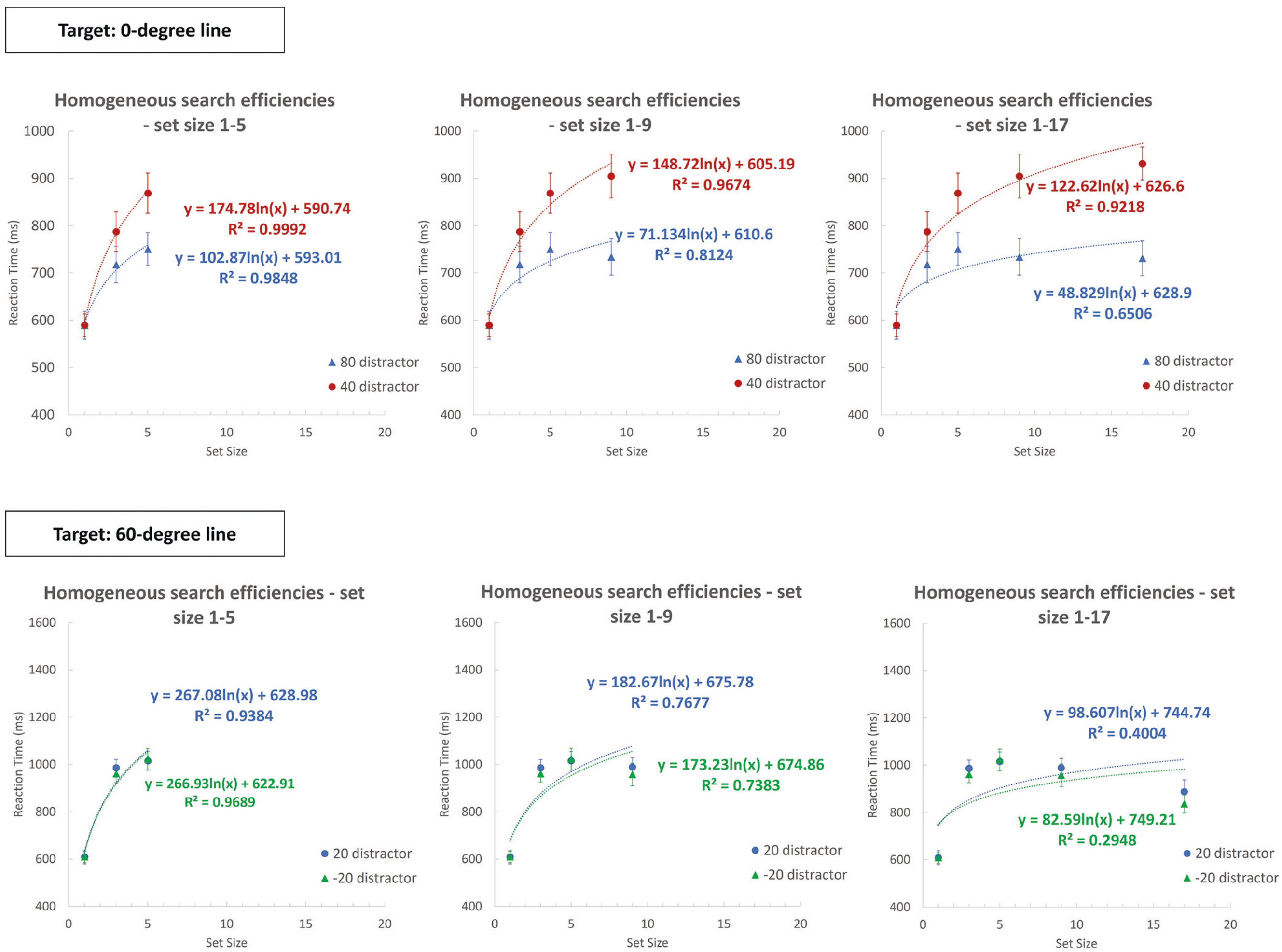
We note that the results from Experiment 1A confirm Wolfe et al.'s (1992) findings that target-distractor distance in orientation space does not solely determine search efficiency in orientation search. In spite of the fact that both target conditions in Experiment 1A had identical target-distractor separations (40 degrees), search for the 20-degree oriented target was half as efficient as search for the vertical target. Similarly, although both target conditions in Experiment 1B had one type of distractor that differed from the target by 80 degrees, the search for the 0-degree oriented target among 80-degree distractors was twice more efficient than the search for a 60-degree target among −20-degree distractors.

Experiment 2: Search Performance Observed in Nonlinearly Separable Heterogeneous Displays

Only correct trials were included in the analysis. The observed response times for all distractor combinations are shown in Table 2.

In previous studies using heterogeneous displays, it is often the case that the same number of distractors of each type are used in a

Figure 7
Logarithmic Search Efficiency in Experiment 1B



Note. Top: Reaction times observed in the homogeneous search conditions of Experiment 1B as a function of distractor set sizes and distractor types when the target was a vertical line (0-degree), and the distractors were 40- and 80-degree lines. Bottom: Reaction times observed in the homogeneous search conditions of Experiment 1B as a function of distractor set sizes and distractor types when the target was a 60-degree line, and the distractors were −20 and 20-degree lines. The left panels show the results when set sizes 1–5 are considered to compute the logarithmic slope, and the middle and right panels, the results for set sizes 1–9 and 1–17, respectively. Dotted lines show the best logarithmic fit. Error bars indicate one standard error of the mean. See the online article for the color version of this figure.

given set size condition (e.g., Wolfe et al., 1992). Figure 8 plots the data from those conditions and allows for a more direct comparison to the results from the homogeneous search condition. A quick glance at the data confirms that RTs in the heterogeneous search conditions (black dots in Figure 8) are indeed much slower than RTs in the homogeneous conditions in Experiment 1A (orange triangles and blue squares in Figure 8). As can be seen on Figure 8, in Experiment 1A, RTs in the homogeneous conditions ranged from 963 to 1,361 ms (excluding the target-only condition). In contrast, in Experiment 2, RTs were almost twice as long, ranging from 1,578 to 2,570 ms (see Table 2). The logarithmic fit for the heterogeneous search function was 636 ms/log unit of set size, which was 2.5 times steeper than the slope observed when searching for the 20-degree target among −20-degree distractors and 4.4 times steeper than the slope observed for the 20-degree target among 60-degree distractors. This confirms previous findings from the literature and would suggest a linear separability effect is at play because of the much more elevated RTs.

Predictive Approach: Predicting RTs in the Nonlinearly Separable Conditions in Experiment 2 by Using the Parameters Observed in Experiment 1A

We used three models to predict what response times ought to be in the nonlinearly separable heterogeneous search conditions in Experiment 2 based on the parameters from the homogeneous conditions in Experiment 1A.

Model 1: Distractor Rejection Cost Model

When solved for two distractor types, Equation 1 becomes Equation 4:

Table 1
Logarithmic Search Slopes Observed in Experiment 1A and 1B

Target orientation	Distractor orientation	Logarithmic search slope (D) Set size 1−5	Set size 1−9	Set size 1−17
Experiment 1A				
0	−40	198.2	185.0	134.5
	40	181.1	172.1	134.5
20	−20	487.3	377.5	251.2
	60	306.5	233.9	146.5
Experiment 1B				
0	40	174.8	148.7	122.6
	80	102.9	71.3	48.8
60	−20	266.9	173.2	82.6
	20	267.1	182.7	98.6

Note. Logarithmic slopes (D, in ms/log unit of set size) and conditions tested in Experiment 1A and 1B. In Experiment 1A, the vertically oriented target was presented among −40 and 40-degree distractors, and the 20-degree target was presented among −20 and 60-degree distractors. In Experiment 1B, the vertically oriented target was presented among 40 and 80-degree distractors, and the 60-degree target was presented among −20 and 20-degree distractors.

$$RT_{Predicted} = a + D_1 \times \ln(N_T + 1) + (D_2 - D_1) \times \ln(N_2 + 1) \tag{4}$$

Equation 4 is the equation previously used by Lleras et al. (2019) and Wang et al. (2017). Here is a brief explanation to get an intuition for what the different terms in the equation mean. This equation assumes that all distractors on the display start to be processed simultaneously. Type I distractors are less similar to the target and thus finish getting processed and rejected first with a temporal constant of D_1 ms per distractor. The total number of type I distractors is N_1. But importantly, all distractors are being processed during this time, even those of type II, which is why the first term in the equation includes all items (N_T, with $N_T = N_1 + N_2$). Type II distractors are more similar to the target; thus, they take longer to reject. As a result, they continue to be processed after type I distractors are rejected. Their accumulators continue to gather evidence until they reach a threshold and get rejected. Alone, they would be rejected at a time cost of $D_2 \times \ln(N_2 + 1)$. Given that during the time taken to reject type I distractors, type II distractors were also being processed, that corresponding time cost must be subtracted from the overall processing time, leading to the term $(D_2 - D_1) \times \ln(N_2 + 1)$. The constant a corresponds to the RT observed in the target only condition.

Model 2: Single-Threshold Model

As a comparison model for Equation 4, Equation 5 assumes a single rejection decision structure, which predicts that all distractors are processed as the distractor that is the most similar to the target (i.e., D_2 ms/log unit of set size per distractor). This type of rejection criterion approximates several ideas proposed in other visual search models, where only one decision threshold is applied to reject noncandidates (e.g., Guided Search, Wolfe, 1994; TAM, Zelinsky, 2008).

$$RT_{Predicted} = a + D_{max} \times \ln(N_T + 1) \tag{5}$$

Why is this equation representative of single-decision threshold models? Single-decision threshold models like Guided Search and TAM assume that there is an initial parallel analysis of the scene, though these models do not specify how long that process takes. They do, however, propose that following that initial analysis, a single decision criterion is applied across the whole scene/priority map to separate obvious nontargets from stimuli that could potentially be the target (candidate stimuli). This decision criterion is directly inspired by signal-detection models, and it produces misses (i.e., candidates that fall below criterion) and false alarms (i.e., very dissimilar distractors that make it above threshold and thus, might get inspected by focused attention). As demonstrated by Townsend and Ashby (1983) and further developed by Buetti et al. (2016), a model that proposes parallel analysis of the scene with stochastic variability and an exhaustive processing rule produces processing costs that increase logarithmically with total set size. The assumption that processing in the parallel stage is exhaustive is implied by the fact that the priority map in these models is computed over the entire scene (i.e., every element must have a priority score) before a decision criterion is applied to the priority map. In other words, single-decision threshold models assume that activation signals are accumulated during the parallel analysis stage at all locations; once this evidence accumulation process ends, a single decision threshold is applied to the resulting activation map to differentiate target-likely locations (above threshold) from target-unlikely locations (below threshold). Thus, if one introduces a temporal duration to the parallel stage of the single-decision threshold models, this temporal duration will increase as a logarithmic function of the number of elements in the scene, with all items being processed at some constant evidence accumulation rate, likely determined by the sensory processing rate. Here, we used D_{max} as the parameter that characterizes the time constant associated with this evidence accumulation process, because it tends to indicate the duration required to reject the most similar distractors to the target in parallel. That said, as explained below, the actual numerical value chosen for the logarithmic slope of this model does not

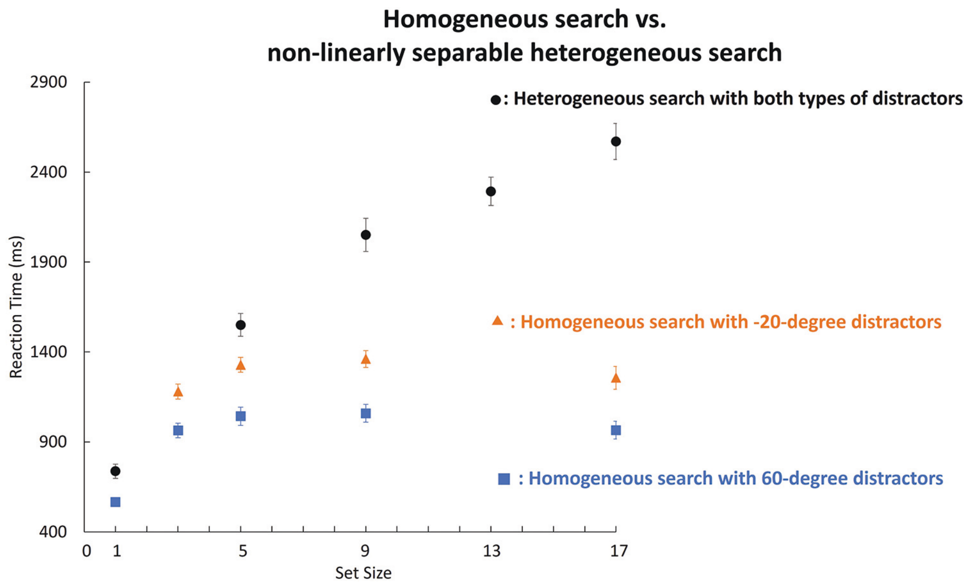
Table 2
Response Times Observed in Experiment 2

N of −20-degree distractors	N of 60-degree distractors	Observed RT (SE) (ms)
0	0	750.88 (39.7)
2	2	1,578.38 (66.4)
2	4	1,784.22 (78.1)
2	6	1,904.67 (89.0)
2	8	1,969.50 (72.2)
4	2	1,847.60 (91.8)
4	4	2,074.13 (89.9)
4	6	2,096.09 (87.8)
4	8	2,239.30 (90.7)
6	2	2,120.77 (81.6)
6	4	2,241.79 (83.5)
6	6	2,331.64 (83.6)
6	8	2,395.76 (80.4)
8	2	2,227.61 (65.9)
8	4	2,402.16 (65.5)
8	6	2,465.13 (91.6)
8	8	2,570.02 (95.4)

Note. RT = reaction time. Mean reaction time (standard error) observed in each condition in Experiment 2. In each condition, there was always a target (20 degrees) and a combination of two types of distractors (−20 and 60 degrees), each with its own set size.

Figure 8

Comparison of Heterogeneous Search and Homogeneous Search Performance



Note. Orange triangles and blue squares: Reaction times observed in the homogeneous conditions of Experiment 1A for the 20-degree target, as a function of distractor set sizes (for set size 1–17) and distractor types, replotted for comparison. Black dots: The figure shows a subset of reaction times observed in Experiment 2 for the 20-degree target. RTs were plotted as a function of total distractor set sizes but only in conditions where there were equal numbers (0, 2, 4, 6, and 8) of −20 and 60-degree distractors in the display (that is when the total set size was 1, 5, 9, 13, and 17), as is typically done in the linear-separability literature. Error bars indicate one standard error of the mean. See the online article for the color version of this figure.

alter how much variance this model will account for in the prediction stage. By comparison, the Distractor Rejection Cost model is mathematically equivalent to arguing that each distractor has its own "rejection decision threshold." This follows because in this model distractors are analyzed at a rate that depends on the overall contrast signal between that item and the target: distractors with different levels of similarity to the target will require different amounts of processing time before they can be ruled out as nontargets, with the least similar distractors requiring much fewer processing time than more similar ones. Having a model that has varying processing rates with a fixed threshold (like TCST) is mathematically equivalent to a model that has a constant processing rate and a variable decision threshold. In sum, Equation 5 represents models where target-unlikely locations are rejected at about the same speed[2] (i.e., where rejection time is independent of an item's similarity to the target), whereas Equation 4 represents models where target-unlikely locations are rejected at varying speeds that are directly determined by their degree of dissimilarity to the target (the target contrast signal).

A second important consideration regarding the importance of using Equation 5 as a comparison model is that it represents *all* RT functions that are a simple logarithmic function of total set size. That is to say, if RTs in the heterogeneous conditions are simply a logarithmic function of total set size, Equation 5 would account for the entirety of the variability in the predicted results (i.e., R-squared would be close to 1), irrespective of the value of the slope parameter in Equation 5. That follows because the slope value D_{max} in Equation 5 works as a linear transformation of log of total set size (i.e., $D_{max} \times \ln(N_T + 1)$). As a result, the actual value of D_{max} would not impact the total variance accounted for by Equation 5; it would only impact the slope of the function relating observed RTs to predicted RTs.

In sum, the Single-threshold model is a tough test for the Distractor Rejection Cost model because its ability to capture variance does not represent a specific model but rather an entire family of models of the form RT = $a + b \times \ln(N_T + 1)$. If Equation 5 fails to win in the model comparison, it would represent strong evidence that (a) RTs in heterogeneous search conditions are not a simple function of log of total set size, as RTs in homogeneous conditions are; (b) that distractors are rejected at rates that are determined by their own level of similarity to the target, rather than by a single, overall decision threshold applied over the entire priority map; and (c) that the parallel analysis of the scene in a visual search model, is best understood as a multithreshold signal detection problem, as opposed to a single decision threshold model, as it is usually studied (Eckstein et al., 2000; Palmer et al., 1993; Verghese, 2001).

[2] All else being equal. Other factors, such as eccentricity, might impact processing speeds at different locations but the key is that these other factors are independent of target-distractor similarity.

Model 3: Swap Model

The third comparison model is a swapped similarity relationship model (Equation 6) based on Equation 4. Remember that the logarithmic slope values (Dj) represent the time to process and reject a given type of distractor j given a specific target template in mind. Distractors of type I (N_1) are associated with search efficiency D_1, distractors of type II (N_2) are associated with search efficiency D_2, and so forth (see Lleras et al., 2020). Model 3 pairs the number of distractors of type II (N_2) to the search efficiency D_1, and vice-versa. This model serves as a sanity check for Equation 4 because it demonstrates the importance of matching the specific number of a given type of distractor present in the display to its corresponding temporal constant (Dj). When simplified, the equation becomes:

$$RT_{Predicted} = a + D_1 \times \ln(N_T + 1) + (D_2 - D_1) \times \ln(N_1 + 1) \tag{6}$$

Note that the first term in the equation is identical to the first term in Equation 4. That is because all items are initially processed with the smallest temporal constant (D_1). However, the last term corresponds to the number of items that were *not* successfully rejected during that initial time period, thus, if N_1 is associated with the term ($D_2 - D_1$), it means that N_2 items were rejected during the first time period associated with D_1. In sum, the Swap model uses the same parameters to predict performance as the Distractor Rejection Cost model, except that the parameters are paired in a slightly different manner.

Model Comparison

The predicted RTs from these three models were then compared with the observed RTs from Experiment 2. Figure 9 shows the prediction accuracy, and Table 3 shows the R^2 and Akaike information criterion (AIC) for the three models across the three ranges of set sizes. For the three ranges of set sizes, the R^2 associated to the Distractor Rejection Cost model were always substantially higher than the R^2 of both Single-threshold model and Swap model. Furthermore, we computed the AIC relative likelihood for the three models, separately for the three ranges of set sizes, using $\exp((\mathrm{AIC}_{min}\text{-}\mathrm{AIC}_i)/2)$ to evaluate the strength of evidence in favor of the winning model. For set sizes 1–5, the AIC model comparison results showed that the Distractor Rejection Cost model (AIC = 163.3) better accounted for the variability in the observed data and was 1.67×10^6 and 4.87×10^{10} times more likely than Single-threshold model (AIC = 191.95) and Swap (AIC = 212.51) model respectively. For set sizes 1–9, the AIC model comparison results showed that the Distractor Rejection Cost model (AIC = 164.24) better accounted for the variability in the observed data and was 1.03×10^6 and 3.61×10^{10} times more likely than Single-threshold model (AIC = 191.95) and Swap (AIC = 212.86) model, respectively. Finally, for set sizes 1–17, the AIC model comparison results showed that the Distractor Rejection Cost model (AIC = 168.41) better accounted for the variability in the observed data and was 1.29×10^5 and 8.6×10^9 times more likely than Single-threshold model (AIC = 191.95) and Swap (AIC = 214.16) model, respectively. Overall, the AIC model comparisons on the three ranges of set sizes showed that predictions from the Distractor Rejection Cost model were a much superior fit to the data than the ones from the Single-threshold model and Swap model.

Experiment 3: Search Performance Observed in Linearly Separable Heterogeneous Displays

Only correct trials were included in the analysis. The observed response times for all distractor combinations are shown in Table 4.

Compared with Experiment 2, in Experiment 3, RTs in the heterogeneous condition do not appear to be slower than RTs in the homogeneous conditions (see Figure 10). This is in accordance with prior literature suggesting that in linearly separable conditions, search unfolds in much the same way in both homogeneous and heterogeneous conditions, unlike in the nonlinearly separable conditions (see Figure 8).

Predictive Approach: Predicting RTs in the Linearly Separable Conditions in Experiment 3 by Using the Parameters Observed in Experiment 1B

We used the same three models to predict what response times ought to be in the linearly separable heterogeneous search conditions in Experiment 3, based on parameters from the homogeneous conditions in Experiment 1B.

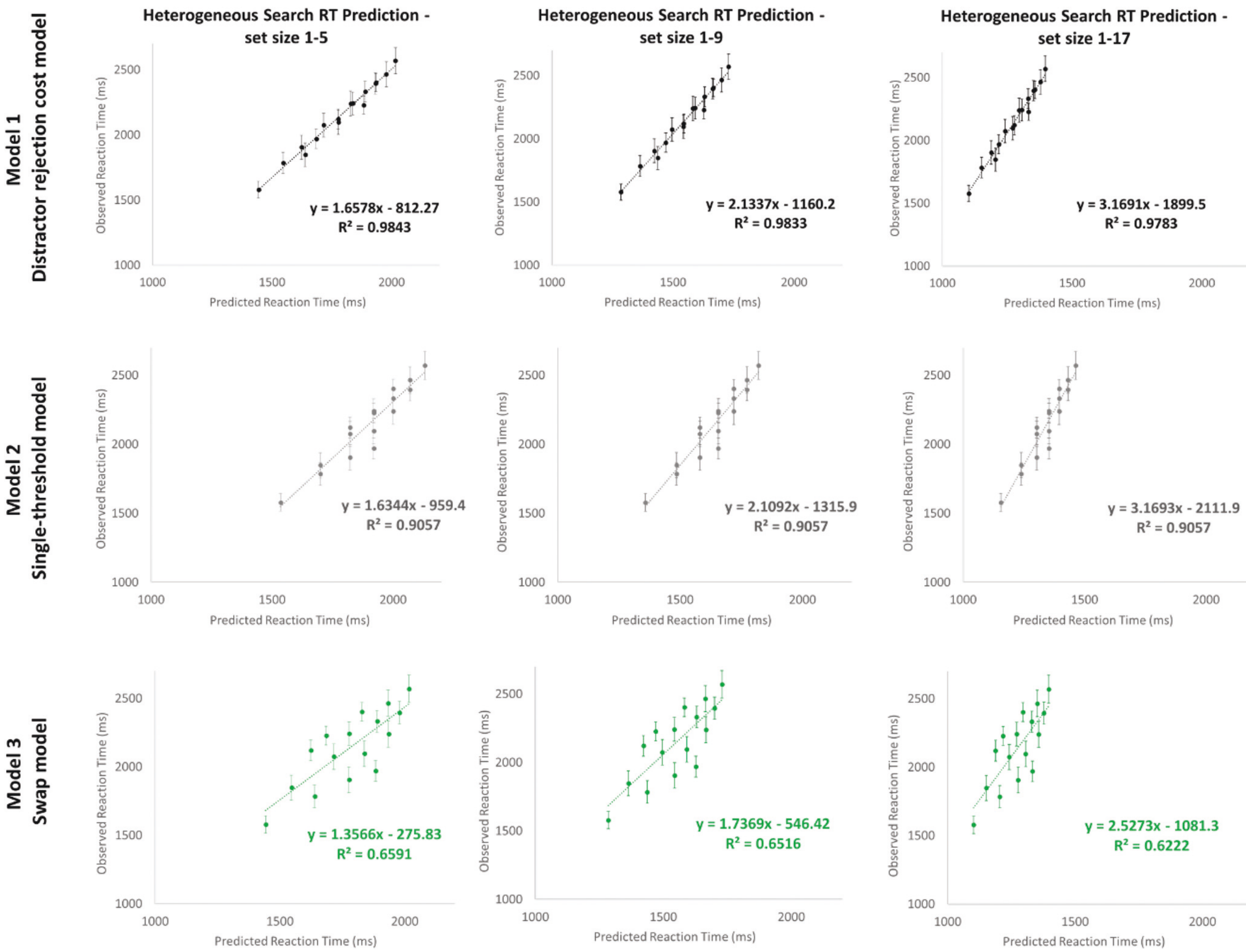
The predicted RTs from these three models were then compared with the observed RTs from Experiment 3. Figure 11 shows the prediction accuracy, and Table 5 shows the R^2 and AIC for the three models across the three ranges of set sizes. For the three ranges of set sizes, the R^2 associated with the Distractor Rejection Cost model were always substantially higher than the R^2 of both Single-threshold model and Swap model. Furthermore, for set sizes 1–5, the AIC model comparison results showed that the Distractor Rejection Cost model (AIC= 127.36) better accounted for the variability in the observed data and was 7.48×10^5 and 5.81×10^8 times more likely than Single-threshold model (AIC = 154.41) and Swap (AIC = 167.72) model, respectively. For set sizes 1–9, the AIC model comparison results showed that the Distractor Rejection Cost model (AIC = 121.44) better accounted for the variability in the observed data and was 1.44×10^7 and 3.18×10^{10} times more likely than Single-threshold model (AIC = 154.41) and Swap (AIC = 169.8) model, respectively. Finally, for set sizes 1–17, the AIC model comparison results showed that the Distractor Rejection Cost model (AIC = 121.1) better accounted for the variability in the observed data and was 1.71×10^7 and 6.93×10^{10} times more likely than Single-threshold model (AIC = 154.41) and Swap (AIC = 171.02) model, respectively. Overall, the AIC model comparisons on the three ranges of set sizes showed that predictions from the Distractor Rejection Cost model were a better fit to the data than the ones from the Single-threshold model and Swap model. It is important to note that the Distractor Rejection Cost model was able to predict performance across two very different sets of conditions: the nonlinearly separable conditions of Experiment 2 and the linearly separable conditions of Experiment 3.

Experiment 4: Search Performance Observed in a Second Nonlinearly Separable Heterogeneous Condition

Only correct trials in Experiment 4 were included in the analysis. The observed response times for all distractor combinations are shown in Table 6.

Figure 9

Plot of Observed RTs in Experiment 2 as a Function of Predicted RTs for the Three Models Being Compared



Note. The figures show the observed RTs in Experiment 2 plotted against the predicted RTs. Data are shown for the Distractor Rejection Cost model (top row), Single-threshold model (middle row), and Swap model (bottom row). For each row, the Dj parameters from Experiment 1A (homogeneous conditions) were extracted for the three different ranges of set sizes 1–5 (left column), 1–9 (middle column), and 1–17 (right column), respectively. Error bars on each data point indicate the standard error of the observed reaction time for each specific condition. See the online article for the color version of this figure.

Compared with Experiment 2, in Experiment 4, RTs in the heterogeneous condition are not slower than RTs in the homogeneous conditions (see Figure 12). This is different from the nonlinearly separable condition presented in Experiment 2 (see Figure 8 for comparison), and it also differs from prior literature suggesting that in nonlinearly separable conditions, search ought to unfold in a more difficult way than both homogeneous and separable heterogeneous conditions. Instead, this result serves as evidence that linear separability per se does not necessarily determine the search difficulty. Note that this result replicates the pattern observed in this same search condition in Wolfe et al. (1992), the study that inspired our selection of stimuli. These authors also failed to find a substantial increase in difficulty in the 0 among −40 and 40-degree distractors.

Predictive Approach: Predicting RTs in Experiment 4 by Using the Parameters Observed in Experiment 1A

We used the same three models to predict what response times ought to be in this nonlinearly separable heterogeneous search conditions in Experiment 4, based on parameters from the homogeneous conditions in Experiment 1A. That said, as demonstrated in the Appendix, we did not anticipate that the model comparison approach would be able to meaningfully discriminate between models, given the numerical similarity between D_1 and D_2 values. Yet, as the results revealed, Experiment 4 does contribute meaningfully to the discussion regarding the plausibility of a linear separability rule in visual search.

The predicted heterogeneous search RTs from the same three models (Equations 4–6) were compared with the observed RTs in

Table 3
R^2 and AIC Results in Experiment 2 for the Three Models, Evaluated Across Three Different Ranges of Set Size: 1–5, 1–9, and 1–17

Index	Model	Set size 1−5	Set size 1−9	Set size 1−17
R^2	Distractor Rejection Cost model	0.984	0.983	0.978
	Single-threshold model	0.906	0.906	0.906
	Swap model	0.66	0.652	0.622
AIC	Distractor Rejection Cost model	163.3	164.24	168.41
	Single-threshold model	191.95	191.95	191.95
	Swap model	212.51	212.86	214.16

Note. AIC = Akaike information criterion.

Experiment 4. Figure 13 shows the prediction accuracy, and Table 7 shows the R^2 and AIC for the three models across the three ranges of set sizes.

As was done for Experiments 2 and 3, we computed the AIC relative likelihood for the three ranges of set sizes separately, using $\exp((\mathrm{AIC}_{\min}-\mathrm{AIC}_i)/2)$. For set size 1–5, the AIC model comparison results showed that the Distractor Rejection Cost model (AIC = 142.15) better accounted for the variability in the observed data and was 4.31 and 14.78 times more likely than Single-threshold model (AIC = 145.08) and Swap (AIC = 147.54) model, respectively. For set size 1–9, the AIC model comparison results also showed that the Distractor Rejection Cost model (AIC = 142.59) better accounted for the variability in the observed data and was 3.46 and 8.89 times more likely than Single-threshold model (AIC = 145.08) and Swap (AIC = 146.96) model, respectively. Finally, for set size 1–17, the AIC model comparison results showed that the Single-threshold model (AIC = 145.08) and Swap (AIC = 145.06) model were almost as equally likely (relative likelihood of 1.004 and 1.01, respectively) as the Distractor Rejection Cost model (AIC = 145.09), making the three models' fits indistinguishable from one another.

In conclusion, as mentioned above, when the two slope values from the homogeneous search conditions are too similar to one another, the model comparison approach does not provide strong evidence to pick a winning model among the three. As a comparison, the largest relative likelihood in this analysis came from predictions using set sizes 1–5, in which the Distractor Rejection Cost model was about 15 times more likely than the Swap model, whereas in Experiment 2, the Distractor Rejection Cost model beat the other models by at least a factor of several thousands. This much larger level of differentiation allows for much stronger conclusions to be drawn.

Discussion

The goal of the present study was to test whether heterogeneous searches rely on the same underlying mechanisms governing homogeneous search. We focused on the linear separability rule as a case study because this rule proposes a fundamental qualitative distinction between two subsets of heterogeneous search conditions: the linearly-separable conditions (easier cases of heterogeneous search) where a single decision rule in feature space permits the correct categorization of features as target versus nontarget features, and the nonlinearly separable conditions (harder cases of heterogeneous search), where no such simple decision rule exists in feature space. If true, this would mean that the visual system treats certain heterogeneous search displays in a fundamentally different manner than how it deals with homogeneous and linearly separable heterogeneous conditions.

No Qualitative Difference Between Linearly and Nonlinearly Separable Conditions: Both Searches Can Be Predicted by Homogeneous Search Parameters to the Same Extent

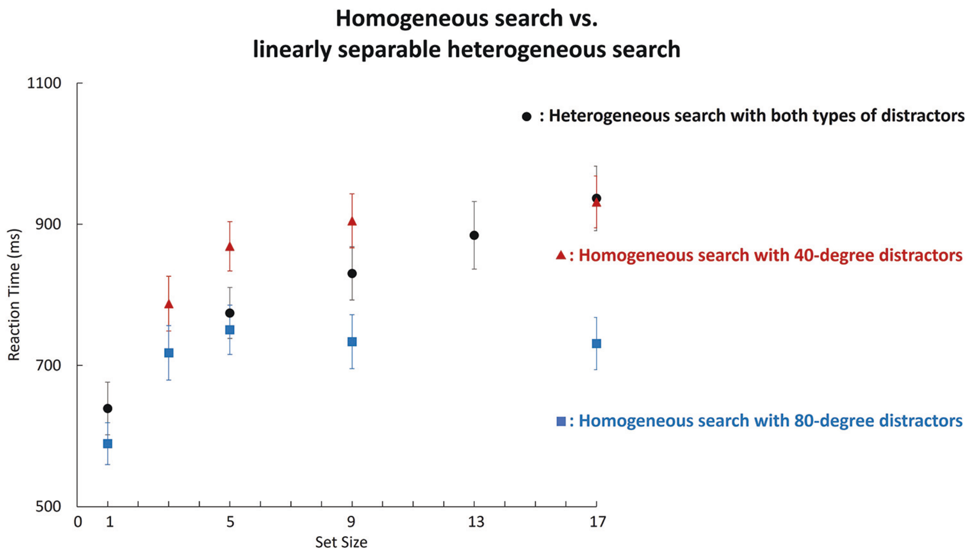
The success of the predictions made by Distractor Rejection Cost model in both linearly separable (Experiment 3) and nonlinearly separable (Experiments 2 and 4) search conditions indicates that the visual system fundamentally treats the two search displays in the same way. Specifically, the same Distractor Rejection Cost model was used to make RT predictions in all of these experiments, and the results suggest there is no qualitative difference between searching for a target among homogeneous distractors and searching for a target that is either linearly separable or nonlinearly separable from the distractors in feature space. Indeed, the Distractor Rejection Cost model accounted for 97% to 98% of the total variance observed in the nonlinearly separable heterogeneous conditions data in Experiment 2, across 16 separate conditions, over a prediction range of nearly 1,000 ms (RT range: 1578 to 2570 ms), and in an overall difficult search task (accuracy range:

Table 4
Response Times Observed in Experiment 3

N of 40-degree distractors	*N* of 80-degree distractors	Observed RT (*SE*) (ms)
0	0	638.93 (37.27)
2	2	774.34 (36.02)
2	4	788.24 (39.08)
2	6	818.36 (46.53)
2	8	820.10 (46.88)
4	2	838.34 (48.14)
4	4	830.34 (37.71)
4	6	864.43 (43.26)
4	8	850.16 (43.17)
6	2	870.65 (42.74)
6	4	883.84 (47.15)
6	6	884.17 (47.81)
6	8	911.51 (54.5)
8	2	902.46 (49.2)
8	4	924.17 (51.18)
8	6	931.30 (51.91)
8	8	936.67 (45.7)

Note. RT = reaction time. Mean reaction time (standard error) observed in each condition in Experiment 3. In each condition, there was always a target (0 degree) and a combination of two types of distractors (40 and 80 degrees), each with its own set size.

Figure 10
Comparison of Heterogeneous Search and Homogeneous Search Performance

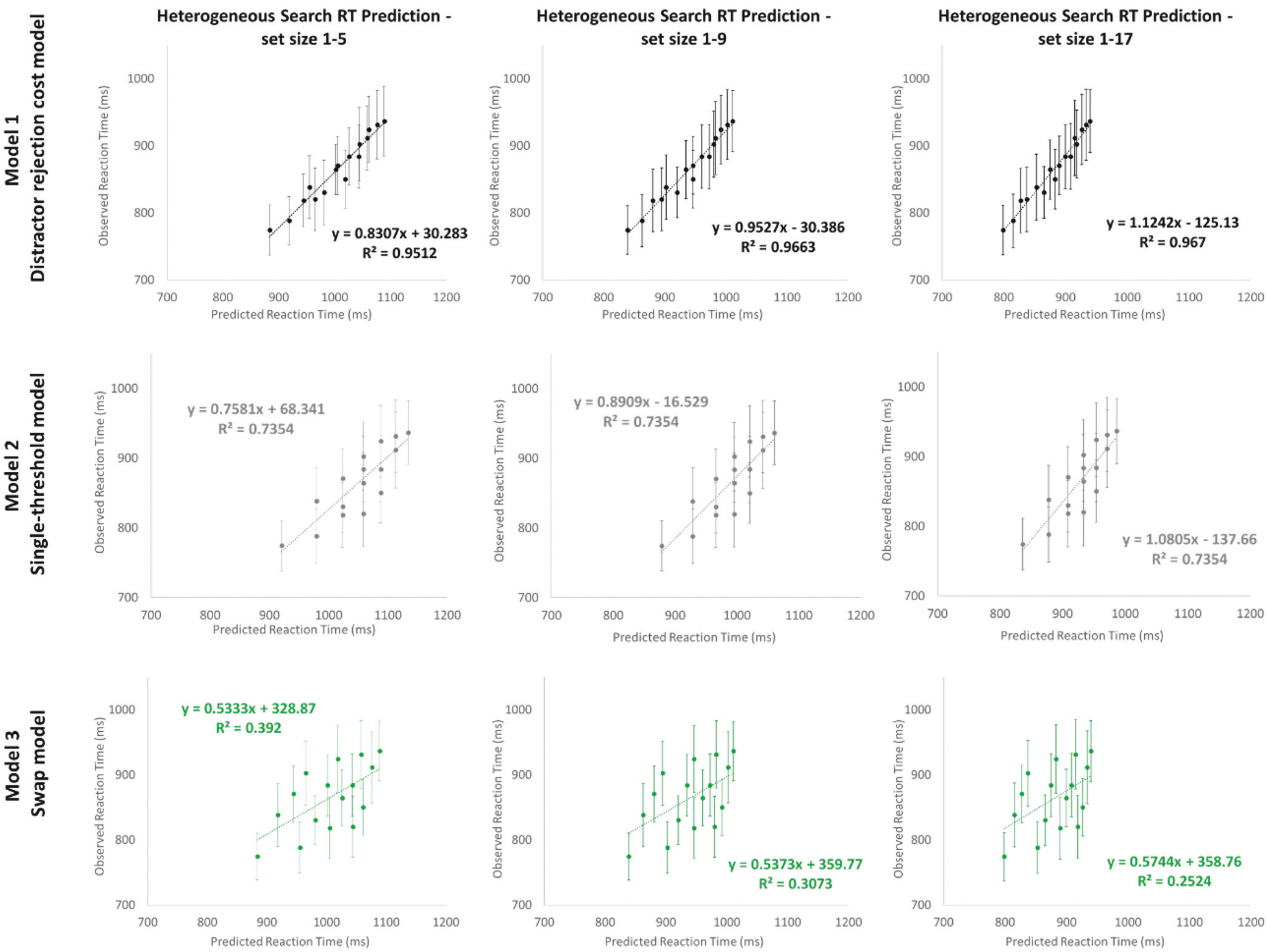


Note. Red triangles and blue squares: Reaction times observed in the homogeneous conditions of Experiment 1B for the vertical target, as a function of distractor set sizes (for set size 1–17) and distractor types, replotted for comparison. Black dots: The figure shows a subset of RTs observed in Experiment 3 for the vertical target. RTs were plotted as a function of total distractor set sizes in conditions where there were equal numbers (0, 2, 4, 6, and 8) of 40 and 80-degree distractors, that is when the total set size was 1, 5, 9, 13, and 17. Error bars indicate one standard error of the mean. See the online article for the color version of this figure.

.71 to .97, mean accuracy: .86). The parameters used to make these predictions were obtained in the much easier homogeneous search task (RT range: 963 to 1,361 ms, accuracy range: .83 to 1, mean accuracy: .95), and were extracted from a separate sample of participants, implying strong robustness and generalizability of the results. Furthermore, the Distractor Rejection Cost model predicted performance in both hard (Experiment 2) and easy (Experiment 4) nonseparable conditions. Indeed, the Distractor Rejection Cost model accounted for 90% to 92% of the total variance observed in in Experiment 4, which was an overall much easier search task than Experiment 2 both in terms of RT (RT range: 815 to 1,014 ms, compared with 1,578–2,570 ms in Experiment 2) and accuracy (accuracy range:.93 to 1, mean accuracy: .98, compared with .86 in Experiment 2). The parameters used to make these predictions were also obtained in a homogeneous search task (RT range: 706 to 907 ms, accuracy range: .96 to 1, mean accuracy: .98). Furthermore, the same Distractor Rejection Cost model also accounted for 95% to 97% of the total variance observed in the linearly separable heterogeneous conditions data (Experiment 3), over a prediction range of 163 ms (RT range: 774 to 937 ms), which was a relatively easier search task (accuracy range: .92 to 1, accuracy mean: .98). The parameters used to make the predictions for Experiment 3 were again obtained in a homogeneous search task (RT range: 718 to 932 ms, accuracy range: .96 to 1, mean accuracy: .99), and were also extracted from a separate sample of participants. One initial conclusion from these results is that the success of the predictive approach demonstrates that the logarithmic slope parameters *D* are a useful index of target-distractor similarity. These slope parameters are robust across different tasks and different groups of subjects (see also Buetti et al., 2019, for further demonstrations; Lleras et al., 2019; Wang et al., 2017).

The conclusion that there is no qualitative difference between separable and nonseparable searches might appear to run counter to some specific aspects of the data. Results from Experiment 2 indicated that distractor heterogeneity in nonlinearly separable conditions almost doubled search time compared with conditions where distractors were homogeneous (Experiment 1A, Figure 8). In contrast, results from Experiment 3 showed that distractor heterogeneity in linearly separable conditions produced comparable performance to conditions where distractors were homogeneous (Experiment 1B, Figure 10). In terms of the coefficient of the predicted by observed RT function, β, the nonseparable condition in Experiment 2 was associated with a β value larger than 1 (1.7–3.2), indicating that the speed at which the distractors were rejected in nonseparable displays was slower than they were in homogeneous displays. In contrast, the linearly separable condition in Experiment 3 was associated with a β value around 1 (.8–1.1), meaning that the distractors were rejected at similar speed in homogeneous and linearly separable displays. One could argue that this is evidence in favor of a linear separability effect. However, results from the nonseparable condition in Experiment 4 similarly showed a β coefficient around 1, suggesting that in some nonseparable conditions, distractors are rejected at the same speed as in homogeneous conditions. That is, search time in the nonseparable condition was in the same range as in the homogeneous search conditions (see Figure 12). As a result, one must conclude that the difficult search observed in Experiment 2 was not due to the nonseparability of the distractors in feature space. Instead, we propose the slowdown was a result of two factors:

Figure 11
Plot of Observed RTs in Experiment 3 as a Function of Predicted RTs for the Three Models Being Compared



Note. The figures show the observed reaction times plotted against the predicted reaction times in Experiment 3. Data are shown for the Distractor Rejection Cost model (top row), Single-threshold model (middle row), and Swap model (bottom row). For each row, the D_j parameters from Experiment 1B (homogeneous conditions) were extracted for the three different ranges of set sizes 1–5 (left column), 1–9 (middle column), and 1–17 (right column), respectively. Error bars on each data point indicate the standard error of the observed reaction time for each specific condition. See the online article for the color version of this figure.

interitem interactions which are driven by the overall similarity between neighboring items (more on this below) and the fact that this similarity measure might depend on specific characteristics of the feature space where those distractor features are selected. For instance, in orientation feature space, it has been demonstrated that categorical status can strongly influence search performance (Wolfe et al., 1992). That is, in addition to the absolute angular difference between two stimuli, the similarity between them also depends on the categorical status of the stimuli (same category vs different category), with categories like "steep," "oblique" and

Table 5
R^2 and AIC Results in Experiments 3 for the Three Models, Evaluated Across Three Different Ranges of Set Size: 1–5, 1–9, and 1–17

Index	Model	Set size 1–5	Set size 1–9	Set size 1–17
R^2	Distractor Rejection Cost model	0.951	0.966	0.967
	Single-threshold model	0.735	0.735	0.735
	Swap model	0.392	0.307	0.252
AIC	Distractor Rejection Cost model	127.36	121.44	121.1
	Single-threshold model	154.41	154.41	154.41
	Swap model	167.72	169.8	171.02

Note. AIC = Akaike information criterion.

Table 6
Response Times Observed in Experiment 4

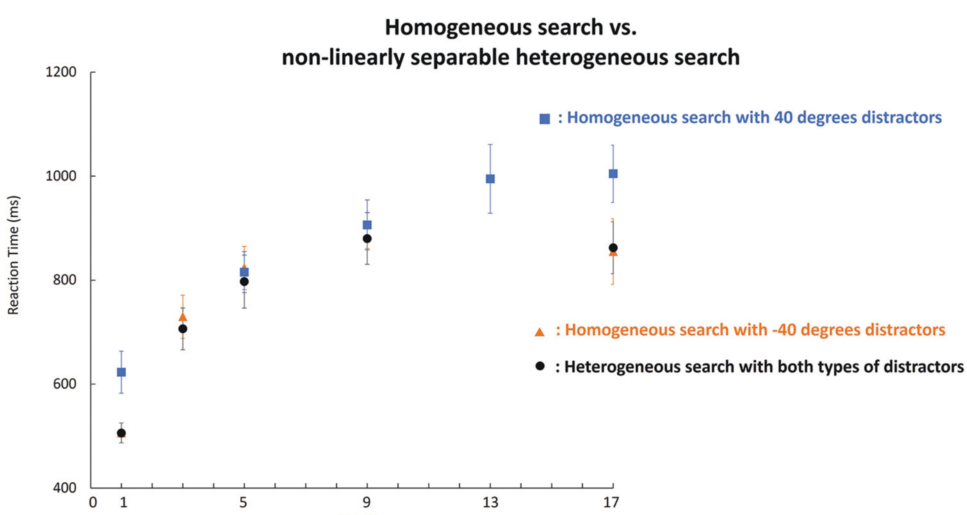
N of −40-degree distractors	*N* of 40-degree distractors	Observed RT (*SE*)(ms)
0	0	622.97 (40.4)
2	2	815.40 (39.4)
2	4	850.56 (48.4)
2	6	889.95 (47.6)
2	8	887.05 (54.1)
4	2	861.18 (51.3)
4	4	906.17 (48.0)
4	6	946.09 (57.3)
4	8	963.05 (59.5)
6	2	911.02 (56.8)
6	4	924.50 (52.4)
6	6	994.87 (66.2)
6	8	1,010.30 (79.5)
8	2	918.18 (56.6)
8	4	979.04 (58.5)
8	6	1,013.83 (51.2)
8	8	1,004.54 (55.4)

Note. RT = reaction time. Mean reaction time (standard error) observed in each condition in Experiment 4. In each condition, there was always a target (0-degree) and a combination of two types of distractors (-40 and 40-degree), each with its own set size.

"horizontal," much like the perceptual similarity between two colors depends not just on the distance between two colors in color space but also on whether the colors cross color categories (red vs blue) or not (red vs. pink). Finally, it is worth remembering that the numerical modulation provided by β represents a quantitative shift in processing, not a qualitative one. That is, β simply represents the extent to which distractor rejection is slowed down under certain conditions compared with others (as demonstrated by the manipulations of spatial configurations in Lleras et al., 2019): the same underlying architecture and form of visual processing is occurring under all the conditions that were successfully predicted by Equation 4. That is, (a) separable, (b) nonseparable easy, and (c) nonseparable hard conditions all obey the same processing rules. There was no need to postulate new rules or new forms of processing to account for differences in observed variability between these three sets of conditions. These processing rules are the same ones that seem to govern heterogeneous search performance more generally.

One major contribution of this article is to provide a mathematical rule to understand and predict heterogeneous search performance. More specifically here, we tested whether the rule that applies to the majority of heterogeneous search conditions where the target is separable from the distractors by a linear boundary in the feature space also applies to a special case of heterogeneous search that has long been considered a theoretical unknown– the nonlinearly separable search condition. As has been known since at least Duncan and Humphreys (1989), performance in heterogeneous displays is usually worse than in homogeneous displays. That said, the *variability* in search data under heterogeneous distractor conditions can be accounted for in terms of parameters observed under homogeneous distractor conditions. Wang et al. (2017) used Equation 1 to predict performance in displays containing up to three different types of distractors, across twenty different conditions, accounting for 97% of the overall variance. The stimuli used were complex images of real-life objects. Lleras et al. (2019) used the same equation to predict performance across three different groups of participants, with a total of 45 different conditions, intermixing two or three different types of distractors, also with great success (R^2 = .9), using colored geometric shapes as stimuli. Together with the current results, the success of this predictive approach across different types of stimuli suggests that

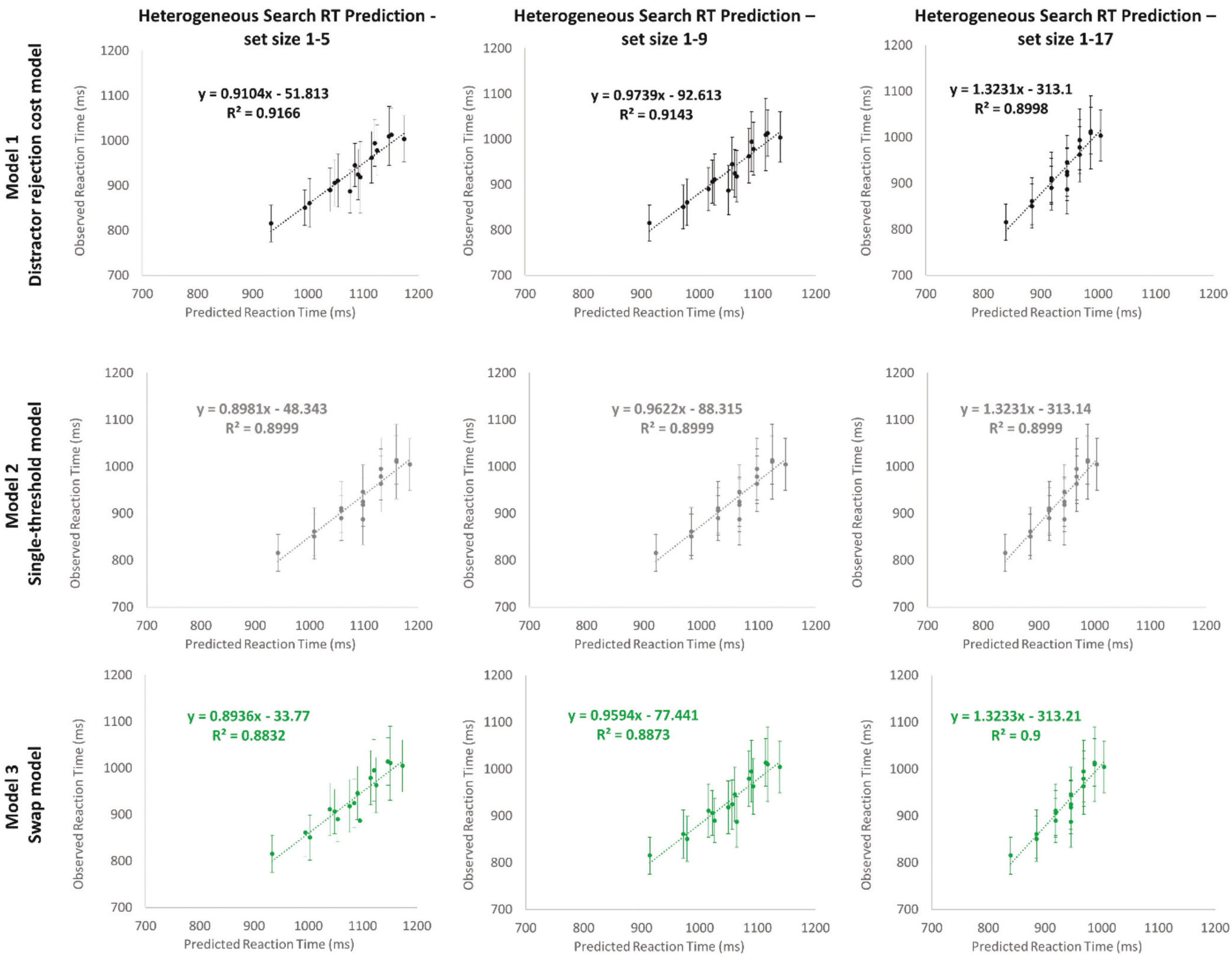
Figure 12
Comparison of Heterogeneous Search and Homogeneous Search Performance



Note. Blue squares and orange triangles: Reaction times observed in the homogeneous conditions of Experiment 1A for the vertical target, as a function of distractor set sizes (for set size 1–17) and distractor types, replotted for comparison. Black dots: The figure shows a subset of RTs observed in Experiment 4 for the vertical target. RTs were plotted as a function of total distractor set sizes in conditions where there were equal numbers (0, 2, 4, 6, and 8) of −40 and 40-degree distractors, that is when the total set size was 1, 5, 9, 13, and 17. Error bars indicate one standard error of the mean. See the online article for the color version of this figure.

Figure 13

Plot of Observed RTs in Experiment 4 as a Function of Predicted RTs for the Three Models Being Compared



Note. The figures show the observed reaction times plotted against the predicted reaction times in Experiment 4. Data are shown for the Distractor Rejection Cost model (top row), Single-threshold model (middle row), and Swap model (bottom row). For each row, the D_j parameters from Experiment 1A (homogeneous conditions) were extracted for the three different ranges of set sizes 1–5 (left column), 1–9 (middle column), and 1–17 (right column), respectively. Error bars on each data point indicate the standard error of the observed reaction time for each specific condition. See the online article for the color version of this figure.

there is a fundamental commonality across all three studies and that there is nothing unique to picking a target that is nonlinearly separable from the distractors: in some cases, heterogeneous performance worsens just as it did in Wang et al. (2017) and Lleras et al. (2019), even though in those studies targets and distractors did not meet the nonlinearly separable criterion; and in other cases, nonlinearly separable heterogeneous search can be as easy as homogeneous search (Experiment 4), further indicating that linearly separability per se does not determine the search difficulty in any meaningful or qualitatively different way.

Our results are in line with findings by Vighneshvel and Arun (2013) showing that there was no need to consider a linear separability effect factor to understand heterogeneous visual search performance. In their experiments, performance was fitted using three parameters reflecting the dissimilarity between the stimuli (i.e., dissimilarity between the target and type I distractors, between the target and type II distractors, and between type I and type II distractors), plus an additional constant. In the Introduction we reviewed the shortcomings of the Vighneshvel and Arun's (2013) modeling approach but, nonetheless, their model showed great success achieving a correlation of .91 (R^2 = .83). Taken together with the success of our approach, we see our current study as a confirmation of Vighneshvel and Arun (2013) conclusion that there is no linear separability effect in visual search (at least for oriented targets).

It should also be noted that Rosenholtz (1999) had also proposed a simple model of saliency to account for the linear-separability effect, where the target salience determines the search efficiency rather than the distance between the target and the boundary line separating target from distractor features. The equation determining the target salience in that model is the distance between the target feature and the mean of the distractor distribution features, and it

Table 7
R^2 and AIC Results in Experiment 4 for the Three Models, Evaluated Across Three Different Ranges of Set Size: 1–5, 1–9, and 1–17

Index	Models	Set size 1−5	Set size 1−9	Set size 1−17
R^2	Distractor Rejection Cost model	0.917	0.914	0.9
	Single-threshold model	0.9	0.9	0.9
	Swap model	0.883	0.887	0.9
AIC	Distractor Rejection Cost model	142.15	142.59	145.08
	Single-threshold model	145.08	145.08	145.08
	Swap model	147.54	146.96	145.06

Note. AIC = Akaike information criterion.

also takes into account the standard deviation of the distractor distribution. To our knowledge, this model has not been directly tested with oriented lines (nor with color). One complication with Rosenholtz's model is the measure of the target-distractor feature distance. As highlighted by Vighneshvel and Arun (2013), parametric distance in feature spaces (e.g., orientation space) from which the search stimuli are created do not necessarily correspond to people's subjective experience of dissimilarity between stimuli in those spaces. To use Rosenholtz's model, one would first need to characterize how distances in the orientation feature space correspond to the perceptual dissimilarity as perceived by the human visual system. This is not easy to accomplish because, as demonstrated by Wolfe et al. (1992), dissimilarity depends both on the distances between target and distractor features as well as the location of those features in feature space. Our current results from Experiment 1 also illustrate this point: a 40-degree difference between target and distractors produced relatively more efficient search when the target feature was a vertical line (134–135 ms/log unit of set size) than when it was a 20-degree oriented line (146–251 ms/log unit of set size). Such feature neighborhood effects cannot be captured by Rosenholtz's model. In contrast, the approach used in the present study assumes that the logarithmic slope parameters D represent a true measure of perceptual distance between two specific stimuli. Specifically, Lleras et al. (2020) proposed that the slope parameter D is inversely proportional to the contrast signal between the target and distractor features (see also Buetti et al., 2019).

The idea that contrast or difference between two stimuli is critical for determining search performance is also consistent with prior evidence from neuroscience showing search speed is proportional to the discriminability between patterns of neuronal activity in visual cortex that respond to target and distractors (e.g., Cohen et al., 2017; Lee et al., 2002; Sripati & Olson, 2010). In fact, the proposal that attention in visual search is driven by target-distractor contrast (or dissimilarity) stands in contrast with many previous models of visual attention and search that emphasis feature-based attention as the mechanism guiding attention (e.g., Bundesen, 1990; Palmer et al., 1993; Verghese, 2001; Wolfe, 1994; Wolfe & Gray, 2007). That is to say, according to those theories, attention is tuned to target-specific features (say, attending to red elements in the scene when one is looking for a red target), and this tuning boosts or prioritizes locations containing those target specific features. Some more recent feature-based theories also propose that attention is *optimally* tuned to a specific feature that most discriminates between the target and distractor features present in the display (e.g., Navalpakkam & Itti, 2007; Scolari & Serences, 2010). For instance, when looking for an orange target among red distractors, attention would be tuned to yellow items because boosting the processing of yellow in the scene would boost the processing of the target (orange) and not the distractors (red). Feature-specific attention theories have been recently challenged by dissimilarity-based theories (e.g., Arun, 2012; Becker, 2008; Becker et al., 2013; 2014; Cohen et al., 2017; Lleras et al., 2020; Vighneshvel & Arun, 2013), where what matters most is not so much the specific value of a target feature but rather, the computation of a difference signal between the expected target feature and the distractor features.

The Role of Interitem Interactions in Heterogeneous and Homogeneous Displays

One key factor that determines search performance is the extent to which nearby items interact with one another (Lleras et al., 2019). The present work suggests that heterogeneous searches rely on the same underlying mechanisms governing homogeneous search. Specifically, the same processes that improve performance in homogeneous search conditions are responsible for slowing down performance in heterogeneous search conditions. As proposed by Lleras et al. (2019), we believe these processes are interitem interactions, which facilitate distractor rejection when nearby items are similar to one another and/or slow down distractor rejection when nearby items are dissimilar to one another.

In homogeneous search conditions, nearby items are always identical to one another and interitem interactions maximally facilitate the rejection of nearby distractors. Search efficiency under these conditions mostly indexes target-distractor similarity. Specifically, search efficiency decreases (i.e., the logarithmic slope becomes steeper) as the similarity between distractors and the target increases. In contrast, in heterogeneous search conditions, nearby items tend to be different from one another (in spatially intermixed conditions), slowing down the rejection of distractors. Wang et al. (2017) and Lleras et al. (2019) initially proposed that the strength of interitem interactions is indexed in the data by the β factor, which is obtained when plotting observed heterogeneous RTs as a function of predicted RTs. Because the predicted RTs are based on data from homogeneous conditions, the magnitude of the β factor indicates the extent to which the interitem facilitatory effects observed in homogeneous displays are still observed in heterogeneous conditions. When the value of β is close to 1, similar strengths interactions are being observed in homogeneous and heterogeneous conditions. When the value of β is larger than 1, it indicates a slowdown in the processing rate of the distractors in the heterogeneous condition compared with the homogeneous condition. The larger the β value is, the stronger the slowdown.

Initial evidence in favor of this proposal comes from Lleras et al. (2019). In that study, the authors used Equation 1 to predict search performance in two heterogeneous display conditions: the intermixed distractors condition (i.e., when distractors of every kind could appear anywhere in the display) and in the spatially segregated condition (i.e., when distractors of the same type always appeared together in the same region of the display). When search performance from homogeneous displays was used to predict performance in intermixed displays, the β was much larger than 1 (1.8 for colored geometric shapes; 1.3 for real world objects, from Wang et al., 2017). When search performance from homogeneous displays was used to predict performance in spatially segregated displays, the β was close to 1 (∼.9 for both stimulus types). Thus, distractor heterogeneity per se did not slow down performance. The slow down occurred when different stimuli were spatially intermixed, weakening local interitem interactions.

The results from the predictive approach in the present study provides additional support to Lleras et al.'s (2019) interpretation of the β factor. First, the results from homogeneous displays showed that as set size increased, interitem interactions facilitated performance so much that RTs began to decrease at larger set sizes, a sort of textural facilitation effect. This textural facilitation was consistently observed across all distractor orientations at larger set sizes. According to the logic described above, a large β factor should be observed when there is a big difference in the processing rate of the distractors between homogeneous and heterogeneous conditions. Supporting this idea, the β values were larger when all set sizes (1–17) were used to compute D parameters in the homogeneous conditions, compared with when only small set sizes (1–5) were used to compute them. For instance, in Experiment 2, a β of 3.2 and a β of 1.7 were observed when Ds were computed over set sizes 1–17 and set sizes 1–5, respectively. We interpreted the increase in β as an indication that interitem facilitation effects were stronger at the larger set sizes in the homogeneous conditions, and these facilitatory effects were much reduced in the heterogeneous condition when distractors were intermixed.

Second, a comparison of the β values between the linearly separable (Experiment 3) and nonlinearly separable (Experiment 2) conditions suggests that, in heterogeneous displays, the interitem facilitatory effects varied as a function of distractor–distractor similarity. Although we do not have a direct measure of the distance in perceptual similarity space between the two distractors used in Experiments 2 and 3, there is reason to believe that the distractors used in Experiment 2 (−20 and +60 degree lines, with a target at +20) were more different from one another than the ones used in Experiment 3 (40 and 80 degree distractors, with a target at 0). For one, the angular difference between the distractors was larger in Experiment 2 (80 degree) than in Experiment 3 (40 degree). And further, from a categorical perspective (see Wolfe et al., 1992), the distractors in Experiment 2 likely belong to different categories (-20 degree lines are tilted "left" and close to a "steep" category, whereas the 60 degree lines are tilted "right" and closer to a "shallow" category). In contrast, in Experiment 3, both distractors are tilted in the same general orientation (right), and neither is particularly "steep," suggesting Experiment 3 distractors are more similar to one another than Experiment 4 distractors. The same is true when comparing the distractors in Experiments 2 and 4, which were both chosen in nonseparable configuration. Although the angular difference between the two distractors were 80 degrees in both Experiments 2 and 4 (−40 and 40 degree distractors), in Experiment 4, both distractors are tilted to the same extent (though in different directions), and both likely belong to "shallower" or "slanted" category, while the target is the only"steep" stimulus in the display.

As noted by Duncan and Humphreys (1989), distractor heterogeneity is higher when distractor–distractor similarity is low, and distractor heterogeneity is low when distractor–distractor similarity is high. As a result, one would predict larger β values would obtain in Experiment 2 compared with Experiments 3 and 4 because the distractor heterogeneity in Experiment 2 is higher than in Experiments 3 and 4. This expectation was borne out in the data: the β values in Experiment 2 were larger (from 1.66 to 3.17, Figure 9) than in Experiment 3 (from .83 to 1.12, Figure 11) and Experiment 4 (.91 to 1.32, Figure 13). Thus, this seems to indicate that in heterogeneous searches, the more distractors are similar to one another, the stronger the interitem facilitatory effects will be, and the more performance will become comparable to the one observed in homogeneous conditions. This can be seen in Figure 10, where performance in the heterogeneous condition of Experiment 3 (black dots) is very similar to performance in corresponding homogeneous conditions from Experiment 1B (red triangles and blue squares), and it is also evident in Figure 12, where performance in the heterogeneous condition of Experiment 4 (black dots) is very similar to performance in corresponding homogeneous conditions from Experiment 1A (orange triangles and blue squares). In other words, it is the distractor–distractor similarity that is at heart of the slowdown in search time, not linear separability.

We should note that different types of stimuli (real-world objects, geometric shape, and oriented lines) and perhaps different feature values within a feature dimension (different orientation of stimuli used in the current study) might afford different magnitudes of β. Therefore, though we expect β to increase when distractor heterogeneity increases, further research is needed to better understand the mathematical relationship between the two, and of the factors (if any) that may moderate that relationship.

Limitations and Generalizability

We should caution that β is not uniquely affected by interitem interactions. Because it is a multiplicative factor in Equation 3, any factor that will systematically impact all D parameters in the same way will be factored out of the sum and multiplicatively modulate the value of β. The present results illustrate this when one looks at the predictions made by D parameters computed over the small set sizes compared with the predictions based on D computed over all set sizes: The β factor increased in systematic fashion as the underlying D parameters decreased. We propose the decrease in the D parameters was caused by textural facilitation effects arising when a certain number of identical lines were presented on the same display. Because the decrease was systematic for both distractor types, β increased. β's increase reflected the extent to which the magnitude of textural facilitation effect that was at play in homogeneous conditions was reduced in corresponding heterogeneous displays. So, one should be mindful that other factors could also come into play and impact the value and interpretation of β, such as changes in the eccentricity of stimuli between Step 1 in the study (estimating Dj in homogeneous conditions) and Step 3 (measuring heterogeneous performance).[3] So, when interpreting β, it is important to consider all systematic factors that may come into play.

[3] This might rarely be a problem when experiments are conducted in a laboratory, although it might come into play when collecting data online where controlling for stimulus and display properties is more difficult.

In terms of generalizability, Equation 1 seems to be able to predict performance across a varied set of stimuli: from single features (oriented lines, present study), to colored geometric shapes (of "medium" complexity, Lleras et al., 2019), to complex stimuli (images of real-world stimuli, Wang et al., 2017). We feel confident that it demonstrates how rising complexity in displays (here, complexity in terms of intermixing different types of distractors, as in heterogeneous conditions) translates into slowdowns in search.

Regarding the generalizability of the predictive approach used here, recently, we used this same predictive approach to understand how color and shape dimensions combine to guide attention in efficient search (Buetti et al., 2019). The results indicated that distinctiveness along color and shape combine linearly to determine the overall distinctiveness of a stimulus that differs from distractors along these two features. In that case, the slope of the predicted versus observed function (the equivalent factor to the β studied here) was close to one, suggesting that color and shape distinctiveness did not coactivate: the presence of distinctiveness signals across the two feature dimensions did not facilitate performance above and beyond what would be expected when only distinctiveness signals are present in one of these dimensions. According to Garner's (1974) work, there are reasons to believe that other feature dimensions, specifically the ones referred to as integral dimensions, might coactivate. The effect of this coactivation should be reflected in the slope factor of the predicted *vs* observed function. We found evidence for coactivation when evaluating how texture and shape (two integral features) information combine to guide attention (Xu et al., 2021). This measure-and-predict approach can therefore shed light into fundamental mechanisms in vision and attention and allows us to estimate quantitatively for the first time the strength of some of those mechanisms (e.g., distractor–distractor interactions, interdimension coactivation).

We should acknowledge that our results even in the homogeneous search condition show RTs that are longer than previously reported RTs in similar oriented search conditions (Wolfe et al., 1992). Our displays were generally larger than the ones used in Wolfe et al. (1992), thus it is quite likely that the relatively faster RTs in their study comes from the fact that interitem interactions and textural facilitation effects were stronger in those displays because display density was greater in those displays. It would not be surprising if orientation search were particularly sensitive to display density. In addition, the present study used a relatively difficult left-right discrimination task that might have also contributed to longer RTs.

Finally, we want to briefly note that our approach focuses on predicting correct RTs and sidesteps the issue of predicting accuracy in heterogeneous displays. We have not started any efforts aimed at understanding differences in error rates across homogeneous and heterogeneous conditions. It is often found that accuracy in the latter is lower than in the former condition. This was evident in Experiment 2, which had an average accuracy of 86%, whereas all the other conditions had accuracies near ceiling (in the 97% to 98% range). Understanding the interaction between evidence accumulation rates and error rates is certainly a subject worthy of future investigation. As a reminder to the reader, data reported in this paper is publicly available, and thus provides an opportunity for other investigators interested in pursuing this line of inquiry.

Conclusion

Wang et al. (2017) introduced a new measure-then-predict methodology whereby stimulus parameters are first estimated to make specific point predictions for what performance ought to be on a novel set of experimental conditions. The current findings represent another success of this new methodology, here applied to the search for oriented lines, in a nonlinearly separable feature arrangement. The results indicated the same equation that describes the slowdown that occurs when real world objects are intermixed predicted 95% to 98% of the variance observed when oriented lines were used as distractors in both linearly separable and nonlinearly separable configurations. These findings demonstrate the generalizability of Equation 1 to predict performance in heterogeneous displays, irrespective of stimulus type, based on parameters observed in homogeneous displays. With regard to the linear-separability effect, the present findings provide strong evidence against the existence of a linear separability rule in feature search. These conclusions align well with previous studies in the literature that have also called into question the existence of this rule (Rosenholtz, 1999; Vighneshvel & Arun, 2013).

References

Arguin, M., & Saumier, D. (2000). Conjunction and linear non-separability effects in visual shape encoding. *Vision Research*, *40*(22), 3099–3115. https://doi.org/10.1016/S0042-6989(00)00155-3

Arun, S. P. (2012). Turning visual search time on its head. *Vision Research*, *74*, 86–92. https://doi.org/10.1016/j.visres.2012.04.005

Bauer, B., Jolicoeur, P., & Cowan, W. B. (1996a). Distractor heterogeneity versus linear separability in colour visual search. *Perception*, *25*(11), 1281–1293. https://doi.org/10.1068/p251281

Bauer, B., Jolicoeur, P., & Cowan, W. B. (1996b). Visual search for colour targets that are or are not linearly separable from distractors. *Vision Research*, *36*(10), 1439–1466. https://doi.org/10.1016/0042-6989(95)00207-3

Bauer, B., Jolicoeur, P., & Cowan, W. B. (1998). The linear separability effect in color visual search: Ruling out the additive color hypothesis. *Perception & Psychophysics*, *60*(6), 1083–1093. https://doi.org/10.3758/BF03211941

Bauer, B., Jolicoeur, P., & Cowan, W. B. (1999). Convex hull test of the linear separability hypothesis in visual search. *Vision Research*, *39*(16), 2681–2695. https://doi.org/10.1016/S0042-6989(98)00302-2

Becker, S. I. (2008). Can intertrial effects of features and dimensions be explained by a single theory? *Journal of Experimental Psychology: Human Perception and Performance*, *34*(6), 1417–1440. https://doi.org/10.1037/a0011386

Becker, S. I., Folk, C. L., & Remington, R. W. (2013). Attentional capture does not depend on feature similarity, but on target-nontarget relations. *Psychological Science*, *24*(5), 634–647. https://doi.org/10.1177/0956797612458528

Becker, S. I., Harris, A. M., Venini, D., & Retell, J. D. (2014). Visual search for color and shape: When is the gaze guided by feature relationships, when by feature values? *Journal of Experimental Psychology: Human Perception and Performance*, *40*(1), 264–291. https://doi.org/10.1037/a0033489

Blais, C., Arguin, M., & Marleau, I. (2009). Orientation invariance in visual shape perception. *Journal of Vision*, *9*(2), 14. https://doi.org/10.1167/9.2.14

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

Buetti, S., Xu, J., & Lleras, A. (2019). Predicting how color and shape combine in the human visual system to direct attention. *Scientific Reports*, *9*(1), 20258. https://doi.org/10.1038/s41598-019-56238-9

Bundesen, C. (1990). A theory of visual attention. *Psychological Review*, *97*(4), 523–547. https://doi.org/10.1037/0033-295X.97.4.523

Cohen, M. A., Alvarez, G. A., Nakayama, K., & Konkle, T. (2017). Visual search for object categories is predicted by the representational architecture of high-level visual cortex. *Journal of Neurophysiology*, *117*(1), 388–402. https://doi.org/10.1152/jn.00569.2016

D'Zmura, M. (1991). Color in visual search. *Vision Research*, *31*(6), 951–966. https://doi.org/10.1016/0042-6989(91)90203-H

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

Eckstein, M. P., Thomas, J. P., Palmer, J., & Shimozaki, S. S. (2000). A signal detection model predicts the effects of set size on visual search accuracy for feature, conjunction, triple conjunction, and disjunction displays. *Perception & Psychophysics*, *62*(3), 425–451. https://doi.org/10.3758/BF03212096

Garner, W. R. (1974). *The processing of information and structure*. Erlbaum.

Hodsoll, J., & Humphreys, G. W. (2001). Driving attention with the top down: The relative contribution of target templates to the linear separability effect in the size dimension. *Perception & Psychophysics*, *63*(5), 918–926. https://doi.org/10.3758/BF03194447

Lee, T. S., Yang, C. F., Romero, R. D., & Mumford, D. (2002). Neural activity in early visual cortex reflects behavioral experience and higher-order perceptual saliency. *Nature Neuroscience*, *5*(6), 589–597. https://doi.org/10.1038/nn0602-860

Lleras, A., Wang, Z., Madison, A., & Buetti, S. (2019). Predicting search performance in heterogeneous scenes: Quantifying the impact of homogeneity effects in efficient search. *Collabra Psychology*, *5*(1), 2. https://doi.org/10.1525/collabra.151

Lleras, A., Wang, Z., Ng, G. J. P., Ballew, K., Xu, J., & Buetti, S. (2020). A target contrast signal theory of parallel processing in goal-directed search. *Attention, Perception & Psychophysics*, *82*(2), 394–425.

Madison, A., Buetti, S., & Lleras, A. (2020). Target discrimination vs. detection in efficient search: What's the difference? *Journal of Vision*, *20*(11), 1380. https://doi.org/10.1167/jov.20.11.1380

Madison, A., Lleras, A., & Buetti, S. (2018). The role of crowding in parallel search: Peripheral pooling is not responsible for logarithmic efficiency in parallel search. *Attention, Perception & Psychophysics*, *80*(2), 352–373. https://doi.org/10.3758/s13414-017-1441-3

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

Palmer, J., Ames, C. T., & Lindsey, D. T. (1993). Measuring the effect of attention on simple visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(1), 108–130. https://doi.org/10.1037/0096-1523.19.1.108

Poisson, M. E., & Wilkinson, F. (1992). Distractor ratio and grouping processes in visual conjunction search. *Perception*, *21*(1), 21–38. https://doi.org/10.1068/p210021

Rosenholtz, R. (1999). A simple saliency model predicts a number of motion popout phenomena. *Vision Research*, *39*(19), 3157–3163. https://doi.org/10.1016/S0042-6989(99)00077-2

Rosenholtz, R. (2001). Visual search for orientation among heterogeneous distractors: Experimental results and implications for signal-detection theory models of search. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(4), 985–999. https://doi.org/10.1037/0096-1523.27.4.985

Saumier, D., & Arguin, M. (2003). Distinct mechanisms account for the linear non-separability and conjunction effects in visual shape encoding. *The Quarterly Journal of Experimental Psychology Section A*, *56*(8), 1373–1388. https://doi.org/10.1080/02724980343000134

Scolari, M., & Serences, J. T. (2010). Basing perceptual decisions on the most informative sensory neurons. *Journal of Neurophysiology*, *104*(4), 2266–2273. https://doi.org/10.1152/jn.00273.2010

Sripati, A. P., & Olson, C. R. (2010). Global image dissimilarity in macaque inferotemporal cortex predicts human visual search efficiency. *The Journal of Neuroscience*, *30*(4), 1258–1269. https://doi.org/10.1523/JNEUROSCI.1908-09.2010

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

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

Vighneshvel, T., & Arun, S. P. (2013). Does linear separability really matter? Complex visual search is explained by simple search. *Journal of Vision*, *13*(11), 10. https://doi.org/10.1167/13.11.10

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

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

Wolfe, J. M. (2007). Guided search 4.0. In W. Gray (Ed.), *Integrated models of cognitive systems* (pp. 99–119). Oxford University Press.

Wolfe, J. M., Friedman-Hill, S. R., Stewart, M. I., & O'Connell, K. M. (1992). The role of categorization in visual search for orientation. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(1), 34–49. https://doi.org/10.1037/0096-1523.18.1.34

Xu, Z. J., Lleras, A., & Buetti, S. (2021). Predicting how surface texture and shape combine in the human visual system to direct attention. *Scientific Reports*, *11*(1), 6170. https://doi.org/10.1038/s41598-021-85605-8

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

Appendix

Mathematical Mechanisms Behind the Models

In the predictive approach used by Wang et al. (2017) and Lleras et al. (2019), the slopes observed in homogeneous search conditions are used to predict RTs in heterogeneous search conditions. In the article, we mentioned the problem that arises when slopes in the homogeneous conditions are too similar to one another. Specifically, when the slopes of the two distractors are too similar to one another, the predictions made by the different models (Distractor Rejection Cost Mode, Single-Threshold Model, and Swap Model) become too similar as well.

Mathematically, when looking at Equation 4 (Distraction Rejection Cost Model), when D_1 and D_2 are numerically similar, the term $(D_2 - D_1)$ will approach 0, and therefore $(D_2 - D_1) \times \ln(N_2 + 1)$ will also approach 0, leaving the following:

$$RT_{Predicted} = a + D_1 \times \ln(N_T + 1) \quad (7)$$

The same reasoning can be applied to Equation 6 (Swap Model). Overall, when D_1 and D_2 are similar to one another, Equations 4 and 6 will become indistinguishable from one another and they will also become indistinguishable from Equation 5, because $D_1 = D_2 = D_{\max}$.

$$RT_{Predicted} = a + D_{max} \times \ln(N_T + 1) \quad (8)$$

To illustrate this limitation, we ran Experiment 4 using the stimuli from Experiment 1A: the 0-degree target accompanied by −40-degree and 40-degree distractors. Results were reported in the main text.

Received September 5, 2020
Revision received May 5, 2021
Accepted May 31, 2021 ■