

# Predicting visual search performance by quantifying stimuli similarities

**Tamar Avraham**

Computer Science Department, Technion I.I.T., Haifa, Israel



**Yaffa Yeshurun**

Psychology Department, University of Haifa, Haifa, Israel



**Michael Lindenbaum**

Computer Science Department, Technion I.I.T., Haifa, Israel



The effect of distractor homogeneity and target–distractor similarity on visual search was previously explored under two models designed for computer vision. We extend these models here to account for internal noise and to evaluate their ability to predict human performance. In four experiments, observers searched for a horizontal target among distractors of different orientation (orientation search; [Experiments 1 and 2](#)) or a gray target among distractors of different color (color search; [Experiments 3 and 4](#)). Distractor homogeneity and target–distractor similarity were systematically manipulated. We then tested our models' ability to predict the search performance of human observers. Our models' predictions were closer to human performance than those of other prominent quantitative models.

**Keywords:** visual attention modeling, visual search, heterogeneous distractors, target–distractor similarity

**Citation:** Avraham, T., Yeshurun, Y., & Lindenbaum, M. (2008). Predicting visual search performance by quantifying stimuli similarities. *Journal of Vision*, 8(4):9, 1–22, <http://journalofvision.org/8/4/9/>, doi:10.1167/8.4.9.

## Introduction

Visual search is the task of detecting a target among non-relevant distractor stimuli. It has been suggested that some attention mechanism is involved when visual search tasks are performed (e.g., Neisser, 1967). This mechanism controls the search strategy so that some tasks are performed faster and more accurately than others (e.g., Triesman & Gelade, 1980; Wolfe, 1994). Other studies have highlighted the role of sensory factors in visual search, demonstrating that search efficiency is largely determined by low-level factors such as target eccentricity and element density (e.g., Carrasco, Evert, Chang, & Katz, 1995; Carrasco & Frieder, 1997; Carrasco, McLean, Katz, & Frieder, 1998; Geisler & Chou, 1995; Palmer, 1994; Vergheze & Nakayama, 1994).

Several computational models have been suggested to account for search performance (e.g., Bergen & Julesz, 1983; Itti, Koch, & Niebur, 1998; Koch & Ullman, 1985; Tsotsos et al., 1995; Wolfe, 1994). The temporal-serial model, for instance, suggests that when the exposure time is limited, the observer can process only  $k$  out of the  $n$  stimuli present in the display. If the target was one of the  $k$  selected stimuli in a target-present display, a correct decision is made. Otherwise a guess would yield 50% success (Bergen & Julesz, 1983). This model is consistent with serial search models in which the items are searched in random order (e.g., Triesman & Gelade, 1980). Another example is the family of models that are based on signal-detection theory (SDT) (e.g., Eckstein, 1998; Eckstein,

Thomas, Palmer, & Shimozaki, 2000; Green & Swets, 1966; Palmer, Ames, & Lindsey, 1993; Santhi & Reeves, 2004). The SDT models assume that the stimuli are observed with stochastic noise. According to this view, a false detection may occur when one of the distractors in a noisy observation is mistakenly perceived as a target (i.e., as belonging to the target distribution), and a miss may occur when the target is mistakenly perceived as a distractor. Hence, the chances of such detection errors increase with an increase in the number of search items and with an increase in target–distractor similarity.

Recently, a few models inspired by Duncan and Humphreys (1989) have been suggested. Duncan and Humphreys' similarity theory suggests that attention is not drawn to locations but rather to image objects, and that search efficiency depends on similarities between objects in the scene and possible targets (target–distractor similarity) and between objects within the scene (distractor heterogeneity). Specifically, search efficiency deteriorates as target–distractor similarity and distractor heterogeneity increase. Rosenholtz (1999), for instance, has developed a simple measure for a target's saliency that reflects search efficiency (denoted here the *saliency* measure) and implemented it within the *best-normal* model (Rosenholtz, 2001a). Given a one-dimensional feature space relevant to the search task (e.g., orientation) and the points in that space describing the various search items, the standard deviation associated with the distractor set is determined. Then, the saliency measure is the number of standard deviations between the target point and the mean of the points representing the distractors.<sup>1</sup> The saliency measure

suggests that a search task is more difficult as the distance between the mean distractor value and the target value decreases (i.e., target–distractor similarity increases) and as the variance of the distractors increases (i.e., distractor heterogeneity increases). As in SDT models, it is assumed that the initial internal response of the visual system to the visual display is noisy. While the saliency measure is a qualitative abstract mathematical phrase for search task difficulty, the best-normal model is quantitative and was designed to predict accuracy in 2-interval forced-choice (2IFC) experiments. The best-normal model is a variation of SDT models. While SDT models assume the observer keeps a record of the exact distribution of the distractors, the best-normal model suggests that during visual search the observer uses a simpler approximated representation of distractor distribution. The true distribution is represented only by its mean and variance, that is, by the normal distribution that best fits the distractors' true distribution. Note that whereas both the best-normal and the classical SDT models predict that search performance should get harder as target–distractor similarity increases, only the best-normal model can account for the increase in search difficulty that comes with an increase in distractor heterogeneity.

Another variation of SDT models is the RCref model (relative coding-with-reference; Rosenholtz, 2001a), which is a modification of the relative coding model (Palmer, Verghese, & Pavel, 2000). Like the best-normal model, the RCref model suggests that the observer does not use the exact distribution of the search items. The recorded distribution does not correspond to the feature values themselves but to the relative values. Specifically, the recorded distribution corresponds to the combination of differences between the various items in the display and the differences between the display items and a reference target. Thus, to decide whether an observed display item is a target or a distractor, it is sometimes compared to another display item and sometimes to a reference target.<sup>2</sup>

Avraham and Lindenbaum (2005, 2006) were also inspired by Duncan and Humphreys (1989). They suggested the *cover* difficulty measure and the FLNN algorithm for visual search in the context of automated computerized systems for object recognition and detection. These models offer a novel approach to account for the effects of distractor heterogeneity and target–distractor similarity on the difficulty of visual search tasks. The goal of our current study was to evaluate the relevance of these models for human search performance and to test whether they improve the ability to predict human performance in comparison to other prominent visual search models.

## The cover difficulty measure

The *cover* is a measure that allows us to qualitatively predict the relative difficulty of different search tasks. As

it was originally developed for computer vision, it was assumed that there is no difference between the displayed items and the observed input (i.e., there is no internal noise). Consider a visual search task where the stimuli (a single target and several distractors) differ by a single feature (e.g., color or orientation). In this case, the display items may be represented as points in a one-dimensional feature space (i.e., on a line).<sup>3</sup> The *cover* is calculated as follows: First, the smallest difference between the target's feature value and a distractor's feature value is measured and denoted  $d_T$ . Then, the cover measure is the number of segments of length  $d_T$  required to cover all the points representing the distractors in the feature space. For example, let us calculate the cover for an orientation search task in which the target is a short horizontal line ( $0^\circ$ ) and the distractors are several lines, each oriented at  $15^\circ$ ,  $25^\circ$ , or  $35^\circ$ . Here  $d_T$  is  $15^\circ$ , and 2 segments suffice to cover the distractors' orientations. (i.e., the  $15^\circ$  and  $25^\circ$  points can be covered by a common segment of length  $15^\circ$ , and another segment is required to cover the  $35^\circ$  point). Therefore, the cover measure is equal to 2. Note that for visual search tasks with homogeneous distractors (with noiseless input) the cover is always 1. Intuitively, we can say that the distractors are divided into groups of elements with similar features and the resulting number of groups reflects the difficulty of the search. The variability within such a group is determined by the target–distractor similarity (the length of  $d_T$ ). Thus, the cover grows as the distractors' heterogeneity increases and as they become more similar to the target.

The original cover measure gives good indication of machine visual search difficulty, as was demonstrated in Avraham and Lindenbaum (2005, 2006). However, to account for human search difficulty, several drawbacks of the cover measure need to be addressed. First, it is deterministic and discrete, while human responses are not. Second, it does not quantify the increase in search difficulty for homogenous displays in which target–distractor similarity increases. Finally, it does not quantify the increase in search difficulty when there is an increase in the number of distractors (set-size effect). Fortunately, these issues are simply addressed by assuming, as in the SDT models, that a noisy representation of each stimulus is observed rather than the exact feature value. This noise is assumed to be normally distributed with mean 0 and observer-dependent standard deviation  $\sigma$ . In this study, we calculated the cover measure on such noisy input (for details, see Appendix A). The effect of this internal noise on the cover measure depends on the noise level: As the noise level grows, more distractor groups are generated. Stimuli that belonged to one group under the original cover measure can now belong to separate groups, as their noisy representations may be different. Therefore, the cover measure, which is associated with the number of groups, grows as the internal noise level grows. As such, the cover measure reflects the increase in search difficulty that comes with the increase in internal noise. This

relationship between search difficulty and internal noise level can account for the set-size effect: for a given internal noise level, an increase in set size results in a decrease in  $d_T$  and an increase in distractor variability, which leads in turn to a larger cover measure. Similarly, if we compare two homogeneous displays with similar set size but with different target–distractor feature distance, we will get a larger cover for the case in which the target–distractor distance is smaller. Hence, unlike the original calculations of the cover measure (Avraham & Lindenbaum, 2005, 2006), in which the cover depended only on distractor heterogeneity and the distance between the target and the closest distractor, the cover measure calculated on a noisy input depends also on the level of internal noise. As such, it may be different for observers with different levels of internal noise.

## The FLNN model

The cover measure suggests a way to compare different search conditions. The relation between the cover measure and the observer accuracy is not explicit, however. The cover measure and human performance may be linked by modeling the search using a simple mechanism, denoted FLNN (farthest-labeled nearest neighbor), suggested for computerized visual-search tasks (Avraham & Lindenbaum, 2005, 2006). The FLNN algorithm starts by choosing one of the display items randomly. If the currently selected item is not the target, another item is selected. This would be the item that is farthest (feature-wise) from all previously selected items. This procedure is repeated until the target is found. To illustrate the procedure, let us take again the visual search task in which the target is a horizontal line of  $0^\circ$  and the distractors are line segments with orientation of  $15^\circ$ ,  $25^\circ$ , and  $35^\circ$ . FLNN's first step is random, so let us consider all the possible scenarios: If the first selected stimulus has an orientation of  $35^\circ$ , the second selected is the most non-similar, which is the target. The same is true if the first selection is a  $25^\circ$  stimulus. However, if the first selection is a  $15^\circ$  stimulus, the second selection will be a  $35^\circ$  line segment, and only the third choice will be the target. The last case is when the target is selected first. In the worst case, two distractors are selected before the target is found. Note that the cover value for this case was also 2. In Avraham and Lindenbaum (2005, 2006), it was analytically proven that FLNN's worst performance never exceeds the corresponding cover, while its mean performance is usually better, as illustrated in the above example.

As in the aforementioned temporal-serial model, when the search display is presented for a limited time, the FLNN model considers only  $k$  out of the  $n$  items. However, while in the temporal-serial model the  $k$  items are selected randomly, they depend in the FLNN model

on the feature values of the stimuli. In particular, the FLNN selects the  $k$  items that are most dissimilar in terms of their feature values, and the typical outcome is that these  $k$  items include representatives of the various groups of items that are formed due to feature similarity. Thus, the search according to the FLNN model is more akin to a search through similarity-based groups than it is to a search through single elements. Moreover, as with the cover measure, the groups relevant for the FLNN search are not necessarily homogeneous, and the degree of within-group heterogeneity depends on the distance, in feature space, between the target and the distractors. Finally, here too we assume that the observations are noisy. Hence, to characterize accuracy in 2IFC experiments, the FLNN model requires two parameters for each individual observer:  $\sigma$  (the level of internal noise) and  $k$  (for details, see Appendix A).

To test the relevance of the cover measure and the FLNN model to human search performance, we conducted four experiments in which target–distractor similarity and distractor heterogeneity were systematically manipulated. We then evaluated the ability of the cover measure and the FLNN model to predict the data collected in these experiments and compared their predictive ability to that of other prominent quantitative models of visual search, including the saliency measure (Rosenholtz, 1999), the standard SDT model (Palmer et al., 1993), the best-normal model (Rosenholtz, 2001a), the RCref model (Rosenholtz, 2001a), and the temporal-serial model (Bergen & Julesz, 1983; Eckstein, 1998). In addition, we compared the predictive abilities of these various models for two additional experiments reported in Rosenholtz (2001a). The overall results show that the cover measure and the FLNN model predict the participants' performance better than the models to which they were compared. Some possible further improvements of the models are suggested and discussed.

## Experiments

This study included four visual search experiments designed to test whether the cover measure and the FLNN model have advantages over previous models of visual search. All four experiments employed the 2IFC paradigm and included either an orientation search task or a color search task. In the orientation experiments (Experiments 1 and 2), the target is always a horizontal line segment and the distractors are oblique line segments (Figures 1a–1b). In the color search experiments (Experiments 3 and 4), the target is always a gray disk, while the distractors are disks in different shades of red and green (Figures 1c–1d). Figure 2 depicts a schematic description of the feature values employed in each condition of each experiment and

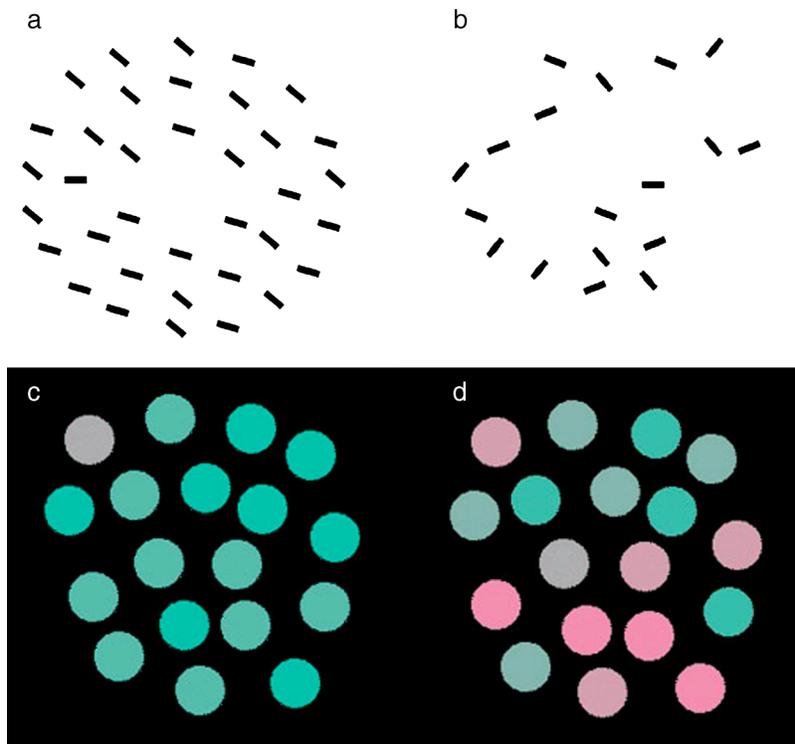


Figure 1. Examples of target-present displays from each experiment: (a) Experiment 1, (b) Experiment 2, (c) Experiment 3, and (d) Experiment 4.

the number of items corresponding to each of these feature values. The first experiment is denoted *unidirectional orientation*, the second *bidirectional orientation*, the third *unidirectional color*, and the fourth *bidirectional color*.

One goal of the *unidirectional* experiments (Experiments 1 and 3) was to obtain corroborating evidence for the hypothesis that tasks with heterogeneous distractors

may be harder than homogeneous ones, even when the distractors in the heterogeneous case are less similar to the target than in the homogeneous case (i.e., are less “confusable” with the target). To that end, all the distractor feature values in the unidirectional experiments lie on one side of the target’s feature value. That is, the feature values of all the distractors are either larger than the target’s feature value (Experiment 1, Figure 2a) or

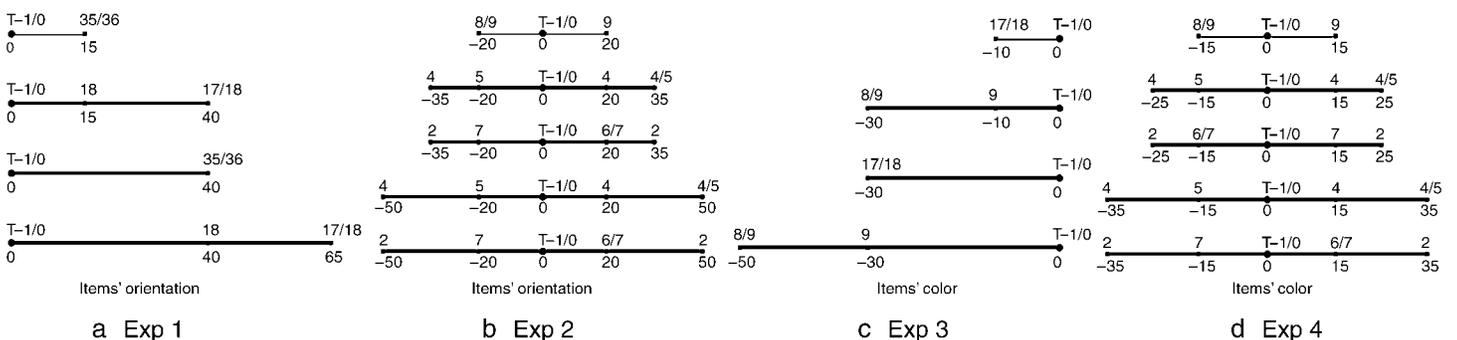


Figure 2. Feature values of the different items in each of the conditions of the four experiments. Experiments 1 and 2 are orientation search tasks (a, b). Experiments 3 and 4 are color search tasks (c, d). Experiments 1 and 3 have four conditions while Experiments 2 and 4 have five conditions. A single horizontal line represents one experimental condition, and the numbers below the points on this line describe the feature value (orientation or color) of the items in this condition. The target value is marked with a T, and the rest of the points represent distractors values. Above each point, the number of items with this value is indicated for both target-present and target-absent displays (e.g., 17/18 means that there are 17 such distractors in a target-present display and 18 such distractors in a target-absent display). See Table 3 for the corresponding colors used in Experiments 3 and 4 in the  $L^*u^*v^*$  and RGB color spaces.

smaller (Experiment 3, Figure 2c). The smallest target–distractor distance ( $d_T$ ) in conditions 1 and 2 of the unidirectional experiments was the same. However, whereas in condition 1 all the distractors were at  $d_T$  distance from the target (i.e., a homogeneous display), in condition 2 half of the distractors were at  $d_T$  distance from the target and the other half were at a greater distance from the target (i.e., a heterogeneous display). If a search through a heterogeneous display can be harder than a search through a homogeneous display, even though some of the heterogeneous distractors are less similar to the target than the distractors in the homogeneous case, condition 2 should be harder than condition 1.

Another goal of the unidirectional experiments was to show that search difficulty varies as a function of the distance between the target's and the distractors' feature values as compared to the distance between the feature values of the distractors themselves and not just as a function of the absolute value of either. That is, the goal was to show that the effect of distractor heterogeneity on search difficulty depends on the distance between the target and the most similar distractors. To test this hypothesis, in conditions 3 and 4, we kept the distance between the two types of distractors the same as the corresponding difference in conditions 1 and 2 but increased the target–distractor distance  $d_T$  (see Figures 2a and 2c). The specific values were set so that the target–distractor distance in condition 2 was smaller than the distance between the distractors and larger in condition 4. If the relative rather than the absolute values of these distances affects the efficiency of the search, there should be a considerable performance difference between conditions 1 and 2 but a smaller difference between conditions 3 and 4.

The cover and the saliency measures and the FLNN and the best-normal models can qualitatively predict both hypotheses, but the SDT model cannot. Thus, in addition to demonstrating that human search efficiency follows, qualitatively, these two hypotheses, we also quantitatively compared the abilities of the models to predict participant performance.

In the *bidirectional* experiments (Experiments 2 and 4), the distractors' feature values were on both sides of the target's feature value (and were always arranged symmetrically): In each display, half of the distractors' feature values were larger than the target's feature value and half were smaller (Figures 2b and 2d). They were designed this way because the cover measure and the FLNN model can predict performance differences between different symmetric conditions, while the saliency measure cannot predict performance differences for cases in which the distractors' feature values are symmetric around the target's feature value. In such cases, the saliency measure is 0 for all conditions. The bidirectional experiments tested whether human observers also experience such differences in difficulty and whether these differences follow our models' predictions. In particular, our models predict that the search should get harder as the target–

distractor distance decreases in comparison to the distance between the distractors themselves (condition 1 vs. conditions 2 and 3 vs. conditions 4 and 5) because more segments of length  $d_T$  are required to cover all the points representing the distractors in the feature space. Additionally, the FLNN model suggests that the search is harder when the feature value that is initially examined is the one most similar to the target. Since the FLNN model chooses this initial feature value randomly, it predicts that the search should be harder when there are more distractors that are similar to the target than distractors that differ from the target (condition 2 vs. 3 and condition 4 vs. 5; Figures 2b and 2d).

## Modeling overview

Below we compare the abilities of the various models to predict the results of these four experiments and those of two additional experiments reported in Rosenholtz (2001a). The predictive abilities of the cover measure are compared to that of the saliency measure. Because the relation between these two measures and observer accuracy is not explicit, we compare their predictive abilities by checking whether each measure can predict the relative difficulty of the search for each of the experimental conditions, and by comparing the correlation coefficients between their predictions and the experimental results. The FLNN model can directly predict search accuracy, and its predictive abilities were therefore compared to other models that can predict search accuracy, including a standard SDT model, the best-normal model, the RCref model, and the temporal-serial model. We have chosen these models because they represent the main quantitative approaches to human visual search at present. To quantitatively compare the predictive abilities of these various models, we used the reduced chi-square measure ( $\chi^2/df$ ) and the chi-square test ( $\chi^2$  test) (see Appendix B for details). We have chosen these tests because they allow us to compare models with a different number of parameters (Taylor, 1982). The implementations of the cover measure and the FLNN model are described in detail in Appendix A. The implementation of the saliency measure followed its description in Rosenholtz (1999). The implementations of the SDT model, the best-normal model, and the RCref model followed the description of these models in Rosenholtz (2001a), and the implementation of the temporal-serial model followed the description of the model in Eckstein (1998). For the modeling of performance in the orientation search experiments (Experiments 1 and 2), we considered the orientation distribution to be wrapped (Rosenholtz, 2001a). A line segment of orientation  $\alpha^\circ$  can be considered both as of orientation  $\alpha^\circ$  and of orientation  $(180-\alpha)^\circ$ . For the modeling of the color search experiments (Experiments 3 and 4), we considered the distributions to be non-wrapped.

Finally, we highlight the major differences between the various models:

- a. the saliency, cover, FLNN, and best-normal models but not the temporal-serial or the SDT-based models can account for the effects of heterogeneous displays like those used in the unidirectional experiments;
- b. the cover, FLNN, and SDT-based models but not the saliency, best-normal, or temporal-serial models can predict performance differences between the different symmetric conditions of the bidirectional experiments;
- c. both the FLNN and the temporal-serial models assume that capacity is limited: When display duration is limited, only  $k$  items out of the total number of items is considered. However, because the  $k$  items in the temporal-serial model are selected randomly, performance differences should only emerge when the number of items is different, whereas in the FLNN model the  $k$  items are chosen based on their feature values. Specifically, the FLNN selects the  $k$  items that are most dissimilar.
- d. Lastly, the final decision of the various models is based on different information. The SDT model uses the exact feature distribution of all the items in the display; the RCref model uses a distribution of the relative feature values (i.e., the differences between the various items) rather than the absolute feature values and the differences between the display items and a reference target; the best-normal model uses the normal distribution that best fits the distractors' true distribution; and the FLNN model uses the distribution of the  $k$  chosen items that typically represent the different similarity-based groups present in the display.

## Experiment 1: Unidirectional orientation

### Method

#### Observers

Five students (A.P., Y.B., D.A., V.S., and A.P.Z.) from the University of Haifa with normal or corrected-to-normal vision participated in this experiment; all were naive to the purpose of the study.

#### Stimuli and apparatus

The stimuli were presented on a 21-in. monitor of a PowerMac G4 computer (resolution: 1280 × 1024 85 Hz), using Vscope™ (Enns & Rensink, 1992). The search display consisted of 36 black line segments, each subtending a visual angle of 0.5° height × 0.1° width, presented on a white background. The lines were randomly scattered within a non-visible circle with a radius of 4° (Figure 1a). The target was always a horizontal 0° line segment, and it was present in the first or second interval equally often. The orientation of the distractor lines for each of the four conditions of this experiment was as follows (see also Figure 2a): In

condition 1, the orientation of all the distractors was 15°. In condition 2, half of the distractors had an orientation of 15° and the other half an orientation of 40°. In condition 3, all the distractors had an orientation of 40°. Finally, in condition 4, half of the distractors had an orientation of 40° and the other half an orientation of 65°. The fixation mark was a plus sign (0.5° width × 0.5° height) presented in the center of the screen, and a plus (0.33° × 0.33°) or a minus (0.33° × 0.1°) sign served as the feedback.

#### Procedure

An experimental trial included two temporal intervals. Each interval began with a 750-ms presentation of the fixation mark followed by a 500-ms search display. The observers were required to indicate whether the target appeared in the first or second interval. Immediately after the observers responded, the appropriate feedback sign was presented for 1 second. Each observer participated in 3 experimental sessions. A single session consisted of four blocks of 100 trials, each corresponding to one of the four experimental conditions. The order of the blocks within a session and the order of trials within a block were randomized. Overall, observers participated in 300 trials per experimental condition and 1200 trials in all.

### Results and predictions

For all five participants, condition 2 was significantly harder than condition 1 ( $z$  test,  $p < 0.05$ ), demonstrating that a search through a heterogeneous display can be harder than a search through a homogeneous display even when half of the distractors in the heterogeneous case are less similar to the target than those in the homogeneous case (see Figure 3a). Moreover, in contrast to the observed difference between conditions 1 and 2, performance in conditions 3 and 4 was not significantly different for all 5 participants. This finding is consistent with the hypothesis that search efficiency depends on the relative rather than absolute values of target–distractor and distractor–distractor feature space distances. In addition, for 2 participants (A.P. and V.S.), condition 1 was significantly harder than conditions 3 and 4 ( $z$  test,  $p < 0.05$ ).

Most central for this study is the quantitative comparison of the abilities of the various measures and models to predict these results. Following the common procedure for models that include a parameter for internal noise, the various parameters were individually evaluated for each participant (e.g., Cameron, Tai, Eckstein, & Carrasco, 2004; Carrasco, Penpeci-Talgar, & Eckstein, 2000; Eckstein, 1998; Eckstein et al., 2000; Rosenholtz, 2001b). First, we checked the ability of the cover measure, the saliency measure, and the SDT model to predict the order of difficulty of the 4 experimental conditions. As can be seen in Table 1, the cover measure can predict the order of difficulty for all five participants. The saliency measure can predict the order for four out of five participants. The

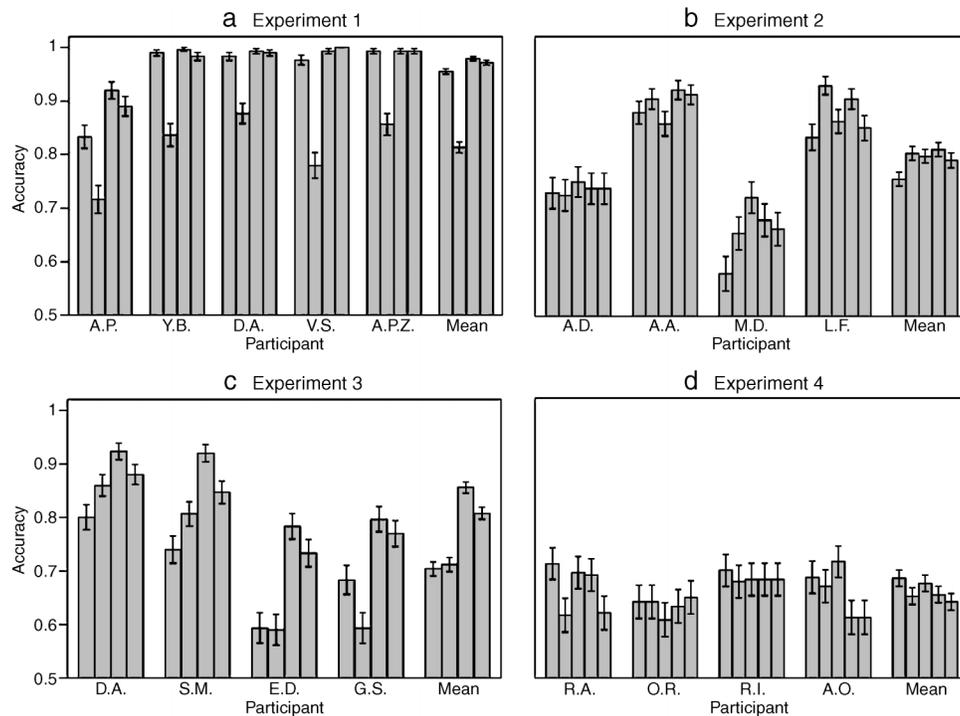


Figure 3. The accuracy of each participant in each experimental condition for the 4 experiments. The conditions are ordered by their index (e.g., the leftmost bar for each participant refers to condition 1). Mean accuracy across participants of each experiment is presented on the right. Error bars correspond to one standard error (SE).

standard SDT model cannot predict the order of difficulty for any of the participants because, regardless of the level of noise, it predicts that condition 1, in which the distractors are most similar to the target, should be the hardest.

Second, to achieve a more quantitative comparison between the cover and saliency measures, we calculated the correlation coefficient ( $r$ ) between the predictions of each model and participant's accuracy (Papoulis & Pillai, 2002). We report the correlation coefficient  $r$ , its significance,<sup>4</sup> and the resulting noise parameter ( $\sigma$ ) for each participant in the right side of Table 1. As can be seen in the table, the correlation coefficients of the cover measure were significant for all five participants. This suggests that the cover measure can quantitatively predict the performance of all five participants. In contrast, none of the correlation coefficients of the saliency measure were significant. Thus, although the saliency measure can predict the correct order of difficulty for most participants, its correlation coefficients are lower than those of the cover measure, and none reached statistical significance. This indicates that there is no good linear transformation from the saliency measure predictions to the accuracy of the participants, while there are some good linear transformations from the cover measure predictions to the accuracy of the participants. It is possible, of course, that there is a good non-linear transformation from the saliency measure to the observed data, and it should be interesting to examine whether indeed there is a different

transformation that will suggest better fitting for the saliency measure prediction. We leave this to future research.

The relationship between participant accuracy and the predictions of the FLNN, SDT, best-normal, RCref, and temporal-serial models are explicit, and we could therefore directly compare the experiment results to these predictions. Figure 4 depicts the observed accuracy of each participant plotted against the predictions of each of the models. If a model could perfectly predict participant accuracy, the various points should fall exactly on the diagonal line. Hence, the closer the points are to this diagonal line, the better the model's predictive abilities are. Moreover, because the number of parameters used in the various models is not equal, we used the reduced chi-square measure ( $\chi^2/df$ ) and the chi-square test ( $\chi^2$  test) (for details, see Appendix B) to compare their predictive abilities (Taylor, 1982). The reduced chi-square measure allows us to compare models that use a different number of parameters by assigning a better fitting grade to a model with fewer parameters that predicts the same results. The lower the value of the reduced chi-square is, the more accurate the prediction. The chi-square test determines whether the probability of obtaining a  $\chi^2$  value larger than the one measured is higher than 0.05, given that the data was supported by the model and taking into account the degrees of freedom ( $df$ ). If it is not, the model is rejected.<sup>5</sup> In Table 2, we report, for each combination of model and participant, the values of the model's parameters that

	Participant	Predictions of difficulty order			Correlation coefficients ( $r$ ) of accuracy vs. prediction and model parameters ( $\sigma$ )			
		SDT	Cover	Saliency	Cover		Saliency	
					$r$	$\sigma$	$r$	$\sigma$
Experiment 1	A.P.	–	+	+	0.999*	2.6	0.812	10.1
	Y.B.	–	+	+	0.998*	1.6	0.538	8.5
	D.A.	–	+	+	0.997*	1.8	0.572	9.8
	V.S.	–	+	+	0.999*	1.8	0.570	10.6
	A.P.Z.	–	+	~	1*	0	0.510	9.7
Experiment 2	A.D.	~	+	+	0.926*	17.8	–	–
	A.A.	~	–	–	0.903*	20.0	–	–
	M.D.	+	–	–	0.962*	19.9	–	–
	L.F.	~	–	–	0.873	16.3	–	–
Experiment 3	D.A.	–	+	~	0.862	11.0	0.900	8.0
	S.M.	–	+	~	0.883	11.0	0.945	7.4
	E.D.	~	+	+	0.993*	4.3	0.996*	8.7
	G.S.	–	+	~	0.997*	2.8	0.880	8.6
Experiment 4	R.A.	–	~	–	0.967*	15.1	–	–
	O.R.	~	+	+	0.956*	13.2	–	–
	R.I.	~	+	+	0.979*	16.7	–	–
	A.O.	–	+	–	0.951*	19.9	–	–
R-1	R.E.R.	–	+	+	0.910	1.5	0.894	8.0
	B.L.B.	–	+	~	0.918	8.6	0.903	10.2
R-2	J.A.K.	~	+	~	0.994*	14.6	0.998*	13.0
	J.O.E.	~	+	~	0.986*	13.0	0.982*	11.9

Table 1. The predictive abilities of the cover measure, the saliency measure, and the SDT-based model. The left side of the table reports, for each participant, whether the model can qualitatively predict the difficulty order of the experimental conditions. A “+” sign indicates that the model can predict the exact difficulty order; a “–” sign indicates that the model predicts at least two conditions in reverse order; and a “~” sign indicates that the model failed to predict the presence or absence of some differences in difficulty, but there is no reverse ordering. The right side of the table reports the best correlation coefficients ( $r$ ) between the participants’ accuracy and the models’ predictions. A “\*\*” next to the  $r$  coefficient indicates that it reached statistical significance.  $\sigma$  is the noise level corresponding to the best  $r$ . For Experiments 2 and 4,  $r$  cannot be calculated for the saliency measure, as it is constant for all conditions.

gave the best fit (i.e., resulted in the lowest  $\chi^2$  value), the reduced chi-square measure, and whether the model was rejected on the basis of the chi-square test. As can be seen

in Table 2, the reduced chi-square score of the FLNN model is the lowest for 4 out of the 5 participants, and it is only rejected by 1 out of the 5 chi-square tests. The

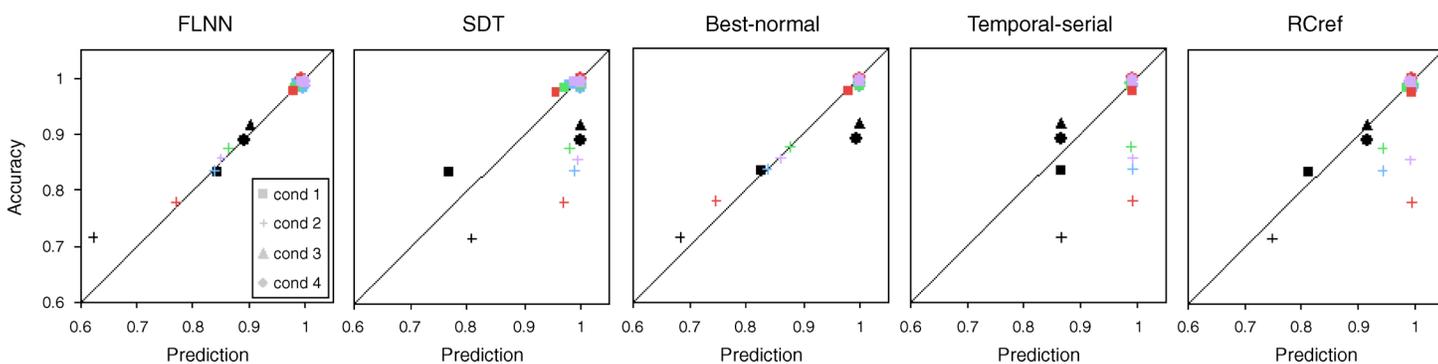


Figure 4. Participants’ accuracy in Experiment 1 vs. the models’ prediction. Points with the same color belong to the same participant. Different marker shapes refer to different conditions. If a model could perfectly predict participants’ accuracy, the various points should fall exactly on the diagonal line.

	Participant	FLNN			SDT		Best-normal		Temporal-serial		RCref	
		$\chi^2/df$	$\sigma$	k	$\chi^2/df$	$\sigma$	$\chi^2/df$	$\sigma$	$\chi^2/df$	k	$\chi^2/df$	$\sigma$
Experiment 1	A.P.	7.477	3.2	1.8	27.971	6.7	20.142	5.8	16.302	26.4	1.535*	3.6
	Y.B.	<b>2.395*</b>	2.7	2.6	19.491	3.6	2.940	1.9	18.785	35.4	11.270	0.4
	D.A.	<b>1.954*</b>	2.8	2.7	12.610	3.8	3.313	1.4	12.044	35.1	5.734	0.4
	V.S.	<b>1.376*</b>	2.7	2.4	23.859	4.1	1.510*	3.6	27.891	35.4	28.350	0.5
	A.P.Z.	<b>1.800*</b>	2.4	2.6	16.786	3.4	1.954*	1.6	14.962	35.3	14.888	0.5
Experiment 2	A.D.	1.553*	7.2	6.5	1.135*	8.0	0.150*	2.7	<b>0.115*</b>	8.5	0.280*	2.4
	A.A.	1.272*	5.9	10.0	1.007*	6.0	<b>0.987*</b>	0.6	1.5888*	14.4	5.781	1.3
	M.D.	3.491	8.2	6.6	2.510	9.6	3.606	6.5	2.820	5.8	<b>2.480</b>	3.8
	L.F.	<b>1.213*</b>	6.1	14.0	1.368*	6.2	3.497	0.8	4.015	13.9	1.860*	1.4
Experiment 3	D.A.	32.912	3.3	7.0	22.128	4.6	37.029	4.1	6.980	13.6	<b>4.750</b>	2.9
	S.M.	36.668	2.9	1.8	26.838	5.5	31.85	9.0	14.14	12.7	<b>7.786</b>	3.6
	E.D.	0.649*	3.5	1.5	2.853	15.2	2.274*	14.8	14.092	6.8	<b>0.074*</b>	10.4
	G.S.	<b>0.633*</b>	3.1	1.6	7.236	13.9	6.477	13.6	12.294	8.0	2.646	7.4
Experiment 4	R.A.	3.665	5.6	5.0	3.028	5.5	<b>1.805*</b>	8.5	2.189*	6.1	2.510	2.2
	O.R.	1.067*	6.0	5.0	<b>0.252*</b>	8.9	0.456*	16.0	0.263*	4.9	1.494*	3.7
	R.I.	1.069*	5.6	5.4	0.782*	7.0	0.155*	6.3	<b>0.075*</b>	6.7	2.919	2.0
	A.O.	1.759*	5.5	4.6	3.131	7.8	<b>1.372*</b>	9.8	2.327*	5.8	2.848	2.3
R-1	R.E.R.	<b>1.565*</b>	6.1	2.7	15.964	8.3	2.326*	7.5	9.364	32.4	2.687	2.7
	B.L.B.	<b>5.725</b>	0.8	1.6	15.025	14.9	9.099	13.9	12.666	18.7	8.483	9.7
R-2	J.A.K.	<b>0.467*</b>	9.8	2.4	2.276*	14.6	1.182*	14.3	0.471*	5.9	2.390*	3.6
	J.O.E.	<b>0.326*</b>	9.1	2.5	2.741	12.9	1.686*	12.6	<b>0.326*</b>	6.4	1.875*	3.0

Table 2. A quantitative comparison of the predictive abilities of the FLNN, SDT, best-normal, temporal-serial, and RCref models for all experiments. For each model and participant, we report the reduced chi-square ( $\chi^2/df$ ) value (lowest value in bold), the chi-square test ( $\chi^2$  test) result ("\*" next to the  $\chi^2/df$  value indicates that the model was not rejected), and the parameters that gave the best fit.

best-normal model is rejected by 3 out of 5 chi-square tests, the RCref by 4 out of 5 tests, and the SDT and temporal-serial models are rejected by all 5 tests. Thus, the predictions of the FLNN model are clearly the closest to human search performance in this experiment.

## Experiment 2: Bidirectional orientation

### Method

#### Observers

Four students (A.D., A.A., M.D., and L.F.) from the University of Haifa with normal or corrected-to-normal vision participated in this experiment; all were naive to the purpose of the study and did not participate in the other experiments.

#### Stimuli and apparatus

The stimuli and apparatus were identical to [Experiment 1](#) except that, to avoid floor effect, the search display consisted of 18 line segments ([Figure 1b](#)) and the orientation of the distractor lines presented in each of the 5 experimental conditions was as follows: In condition 1, half of the distractors had an orientation of  $20^\circ$  and the other half of  $-20^\circ$ . In condition 2, the orientation of the

distractors was more-or-less equally divided between  $-35^\circ$ ,  $-20^\circ$ ,  $20^\circ$ , and  $35^\circ$ . Condition 3 employed the same type of distractors as condition 2, but with more  $-20^\circ$  and  $20^\circ$  distractors and fewer  $-35^\circ$  and  $35^\circ$  distractors (see details in [Figure 2b](#)). In condition 4, the orientation of the distractors was more-or-less equally divided between  $-50^\circ$ ,  $-20^\circ$ ,  $20^\circ$ , and  $50^\circ$ . Finally, condition 5 employed the same type of distractors as condition 4, but with more  $-20^\circ$  and  $20^\circ$  distractors and fewer  $-50^\circ$  and  $50^\circ$  distractors.

#### Procedure

The procedure was identical to [Experiment 1](#) except that a single session consisted of five blocks of 80 trials, each corresponding to one of the five experimental conditions. Each observer participated in 3 such sessions, for a total of 240 trials per condition and 1200 trials for the entire experiment.

### Results and predictions

The statistical analysis revealed significant performance differences between some conditions for 3 out of the 4 participants (see [Figure 3b](#)). This finding demonstrates that accuracy might differ even with displays that include

distractors whose feature values lie symmetrically on both sides of the target's feature value (i.e., some feature values are larger and some are smaller than the target). The specific performance differences that reached statistical significance ( $z$  test,  $p < 0.05$ ) were different for the different participants: For A.A., condition 3 was significantly harder than conditions 4 and 5; for M.D., condition 1 was significantly harder than the other conditions; and for L.F., condition 2 was significantly easier than conditions 1, 3, and 5 while condition 4 was significantly easier than conditions 1 and 5.

As in [Experiment 1](#), we evaluated the predictive abilities of the various models by first comparing the success of the cover measure, the saliency measure, and the SDT model in predicting the relative difficulty of the 5 experimental conditions, and we calculated the correlation coefficient  $r$  and tested its significance for the cover and the saliency measures ([Table 1](#)). The predictive success of the SDT model depends on the similarity of the distractors to the target; the model could predict the order of difficulty for one participant. The cover measure cannot predict that the display in which the distance between the distractors is the smallest (condition 1) would be the hardest to search through, and it could only predict the order of difficulty for one participant for which condition 1 was not the hardest. Still, it succeeded in providing significant correlation coefficients for three out of the four participants. The saliency measure predicts a 0 saliency for all conditions because the target value and the mean of the distractors' values are equal. It is therefore not possible to calculate the  $r$  correlation coefficient for this model as it is not defined for constant vectors. In other words, no linear transformation can transform the constant values of the saliency measure to the participants' measured accuracies. Like the cover measure, it could only predict the difficulty order of one participant.

To quantitatively compare the predictive abilities of the FLNN, SDT, best-normal, RCref, and temporal-serial models, we plotted in [Figure 5](#) the observed accuracy of each participant against the predictions of each of the

models. Additionally, we calculated the reduced chi-square measure and performed the chi-square test for each of these models ([Table 2](#)). Both the SDT and FLNN models passed the chi-square test for three out of the four participants. The best-normal, RCref, and temporal-serial models passed the chi-square test for only two out of the four participants. Finally, none of these models stood out in terms of the lowest (best) reduced chi-square measure.

### Experiment 3: Unidirectional color

This experiment examined whether the results for the unidirectional orientation search ([Experiment 1](#)) could be replicated with another feature and whether our models can also predict performance for another feature. To that end, the conditions in this experiment were very similar to [Experiment 1](#) but the search-relevant feature was color rather than orientation.

#### Method

##### Observers

Four students (D.A., S.M., E.D., and G.S.) from the University of Haifa with normal or corrected-to-normal vision participated in this experiment; all were naive to the purpose of the study and did not participate in the other experiments.

##### Stimuli, apparatus, and procedure

The stimuli, apparatus, and procedure were identical to [Experiment 1](#) except for the following: The search elements consisted of 18 colored disks, with a diameter of  $1.2^\circ$ , presented on a black background ([Figure 1c](#)). The disks were randomly placed within a non-visible circle with a radius of  $4.5^\circ$ . A detailed description of the various colors employed in this experiment is given in [Figure 2c](#) and [Table 3](#). These colors were selected after we measured the appropriate  $u'$ ,  $v'$ , and  $cd/m^2$  values with a

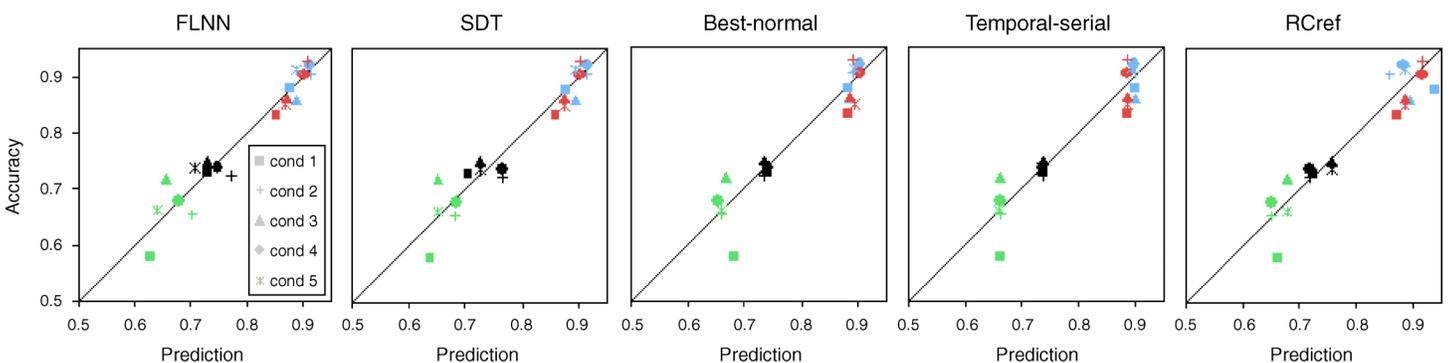


Figure 5. Participants' accuracy in [Experiment 2](#) vs. the models' prediction. Points with the same color belong to the same participant. Different marker shapes refer to different conditions. If a model could perfectly predict participants' accuracy, the various points should fall exactly on the diagonal line.

Description	$u'$	$v'$	$cd/m^2 = Y$	R	G	B
Black	0	0	0	0	0	0
-50	0.138	0.422	22.1	0	185	158
-35	0.151	0.425	23.0	49	180	159
-30	0.1595	0.427	23.6	73	178	159
-25	0.162	0.426	23.25	81	177	159
-15	0.173	0.4265	24.1	115	172	160
-10	0.179	0.427	24.35	130	168	160
0 (target)	0.189	0.427	24.6	160	160	160
15	0.2055	0.424	24.25	204	143	161
25	0.211	0.421	23.85	218	135	162
35	0.223	0.4165	23.1	240	121	163

Table 3. The color values employed in Experiments 3 and 4.

Tektronix J18 LumaColor™ II Photometer. We used the  $L^*u'v'$  color space because it was designed so that the distances in the color space are linear with differences in color perception (C.I.E. 1978; for equations see, e.g., Travis, 1991). Figure 6a depicts the exact distances in  $u'v'$  space, when  $L^*$  is constant, for this experiment. We also tried to keep  $v'$  more-or-less constant and changed only  $u'$ . Hence, the distance in feature space between the search items (i.e., their feature value) was computed from the  $u'$ -value differences. Additionally, we chose a black background because its perceptual distance from all stimuli is approximately equal (Figure 6c). This exempted us from

taking it into account when considering the different models. The color of the target disk was always gray (0 feature value; Table 3), and it was present in the first or second interval equally often. The color of the distractor disks presented in each of the four conditions of this experiment was as follows (see also Figure 2c): In condition 1, all the distractors had the same greenish color (-10 feature value). In condition 2, half of the distractors had one greenish color (-10 feature value) and the other half had another greenish color (-30 feature value). In condition 3, all the distractors had the same greenish color (-30 feature value). Finally, in condition

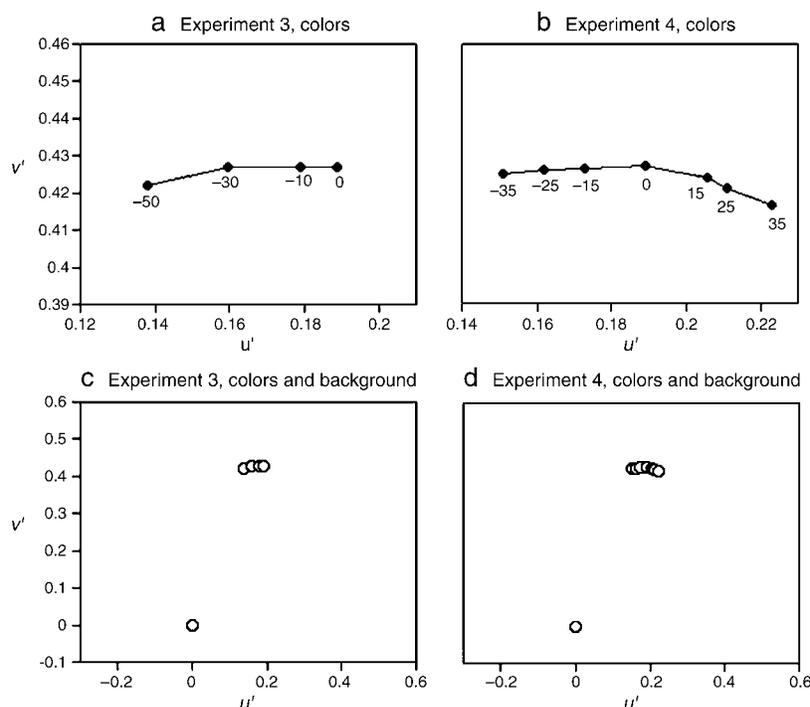


Figure 6. Panels a and b depict the colors used in Experiments 3 and 4 (respectively) in the  $u'v'$  space and demonstrate the feature-space distances. In panels c and d, the color black (0,0) is also plotted to demonstrate the approximately equal distances of all stimuli from the background.

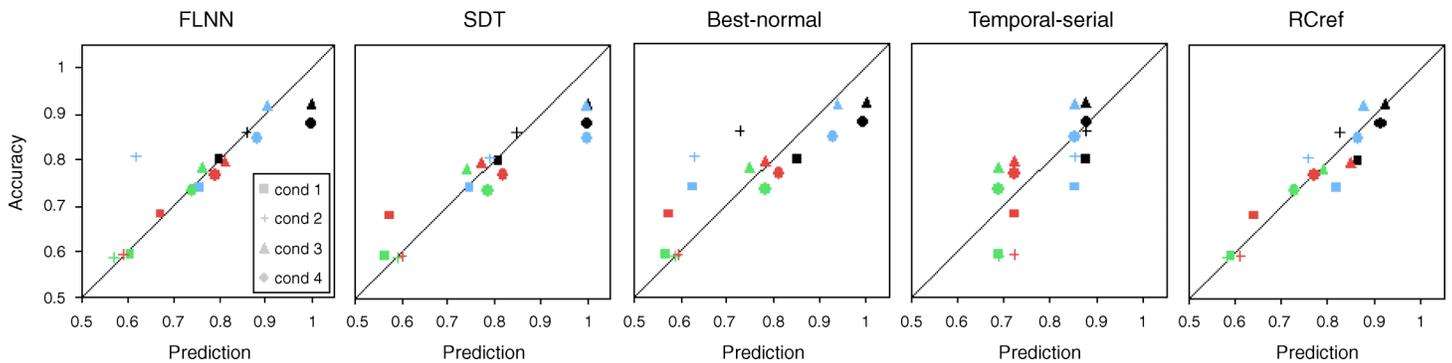


Figure 7. Participants' accuracy in [Experiment 3](#) vs. the models' prediction. Points with the same color belong to the same participant. Different marker shapes refer to different conditions. If a model could perfectly predict participants' accuracy, the various points should fall exactly on the diagonal line.

4, half of the distractors had one greenish color ( $-30$  feature value) and the other half had another greenish color ( $-50$  feature value).

### Results and predictions

As in [Experiment 1](#), for all four participants, the accuracy in condition 2 is lower than that in conditions 3 and 4 (see [Figure 3c](#)). Specifically, for all the participants, condition 3 was significantly easier than condition 2 ( $z$  test,  $p < 0.05$ ) and also significantly easier than condition 1. For 2 participants (E. D. and G.S.), condition 4 was also significantly easier than conditions 1 and 2. Unlike [Experiment 1](#), there were no consistent differences between conditions 1 and 2. For 2 participants (D.A. and S.M.) condition 1 was significantly harder than condition 2, and for one participant (G.S.) condition 2 was significantly harder than condition 1. Finally, for 2 participants (D.A. and S.M.), condition 4 was significantly harder than condition 3. Hence, the pattern of results found in this experiment is different than in [Experiment 1](#), where orientation is the search-relevant feature. This may suggest that searches that are based on different features are limited by different factors. Alternatively, the true distances between the various feature values that make up the search display might be different than those assumed here because the color space we used might not be a good enough match for the perceptual color space (e.g., Fairchild, 1998). We proceed to examine how well the models can predict these results.

As is evident in [Table 1](#), the cover measure is able to predict the order of difficulty for all four participants and passes the correlation coefficient test for two out of the four participants. The SDT model cannot predict the order of difficulty of all four participants because it cannot predict that condition 4, in which the distractors are less similar to the target, is harder than condition 3. The saliency model is able to predict relative difficulty for one out of the four participants and passes the correlation coefficient test for that one participant.

The observed accuracy of each participant is plotted against the predictions of each of the FLNN, SDT, best-normal, RCref, and temporal-serial models in [Figure 7](#), and the best-fit parameters, reduced chi-square measure and chi-square test of each combination are reported in [Table 2](#). The SDT and temporal-serial models did not pass any chi-square test, the best-normal and RCref models passed one out of four, and the FLNN passed two out of four tests.

## Experiment 4: Bidirectional color search

This experiment was designed to test whether the pattern of results found in [Experiment 2](#) would be replicated when the search-relevant feature is color and to evaluate the ability of our models to predict this pattern of results.

### Method

#### Observers

Four students (R.A., O.R., R.I., and A.O.) from the University of Haifa with normal or corrected-to-normal vision participated in this experiment; all were naive to the purpose of the study and did not participate in the other experiments.

#### Stimuli, apparatus, and procedure

The stimuli and apparatus were identical to [Experiment 3](#) except that here the color of the distractors presented in each of the 5 experimental conditions was as follows (see also [Figures 1d, 2d, 6b, 6d](#), and [Table 3](#)): In condition 1, half of the distractors had a greenish color ( $-15$  feature value) and the other half had a pinkish color ( $15$  feature value). In condition 2, the color of the distractors was more or less equally divided between 2 greenish colors and 2 pinkish colors (feature values of  $-25$ ,  $-15$ ,  $15$ ,  $25$ , respectively). Condition 3 employed the same type of distractors as condition 2, but with more  $-15$  and  $15$  distractors, and fewer  $-25$  and  $25$  distractors (see details in [Figure 2d](#)). In condition 4, the color of the distractors was more-or-less

equally divided between 2 greenish colors and 2 pinkish colors (feature values of  $-35$ ,  $-15$ ,  $15$ ,  $35$ , respectively). Finally, condition 5 employed the same type of distractors as condition 4, but with more  $-15$  and  $15$  distractors and fewer  $-35$  and  $35$  distractors. The procedure was identical to [Experiment 2](#).

### Results and predictions

Significant performance differences ( $z$  test,  $p < 0.05$ ) were found between some conditions for 2 out of the 4 participants (see [Figure 3d](#)). Specifically, for R.A., conditions 2 and 5 were significantly harder than conditions 1, 3, and 4. For A.O., conditions 4 and 5 were significantly harder than conditions 1, 2, and 3. Thus, as in [Experiment 2](#), differences in accuracy were found even with displays that include distractors whose feature values lie symmetrically on both sides of the target's feature value. This finding does not agree with Bauer, Jolicoeur, and Cowan (1996), who suggest that all color search tasks should be equally hard whenever the target is not linearly separable from the distractors.

As mentioned above, the finding that there are differences in performance for different experimental conditions is not predicted by the saliency measure, which predicts equal difficulty for all the conditions for bidirectional (symmetric) displays. Indeed, it could only predict the order of difficulty for two out of the four participants ([Table 1](#)), as their performance was similar for all conditions. As explained in the discussion of the results of [Experiment 2](#), it is not possible to calculate the  $r$  correlation coefficient for this model when the display is bidirectional. The cover measure was able to predict the order of difficulty for three out of the four participants and passed the correlation coefficient test for all participants. The SDT model cannot predict the order of difficulty for any of the four participants, as it considers condition 1 to be the hardest, followed by 3, 5, 2, and finally 4. The predictive abilities of the FLNN, SDT, best-normal, RCref, and temporal-serial models can be evaluated from [Table 2](#) and [Figure 8](#). The best-normal and temporal-serial

models passed all four chi-square tests, the FLNN model passed three out of four chi-square tests, the SDT model passed two out of four chi-square tests, and the RCref model passed only one chi-square test.

### Previous orientation search experiments

In this section, we evaluate the ability of our models to predict the results of two additional orientation search experiments reported by Rosenholtz (2001b; denoted there [Experiments 1](#) and [2](#) and here R-1 and R-2, respectively) and compare their predictive abilities to those of the other models. Our aim here was to test whether our models are robust enough to predict performance in experiments that were carried out under different experimental conditions. We first report a summarized description of these experiments and then proceed to evaluate the models.

The target in Experiment R-1 was always a  $0^\circ$  horizontal line segment, and the distractors varied between the four experimental conditions as follows: In condition 1, all the distractors had an orientation of  $30^\circ$ . In condition 2, one-third of the distractors had an orientation of  $30^\circ$  and two-thirds had an orientation of  $50^\circ$ . In condition 3, one-third of the distractors had an orientation of  $30^\circ$ , one-third had an orientation of  $50^\circ$ , and one-third had an orientation of  $70^\circ$ . In condition 4, one-third of the distractors had an orientation of  $30^\circ$  and two-thirds had an orientation of  $70^\circ$ . The task was a 2IFC detection task with 36 items in each display. Experiment R-2 was identical to experiment R-1 apart from the fact that 8 elements appeared in each display (instead of 36). There were two participants in each of these experiments. For further details about the method used see Rosenholtz (2001a).

### Predictions

A comparison of the predictive ability of the cover, saliency, and SDT in these two experiments ([Table 1](#)) reveals that the cover measure successfully predicted the order of difficulty for all the participants of both

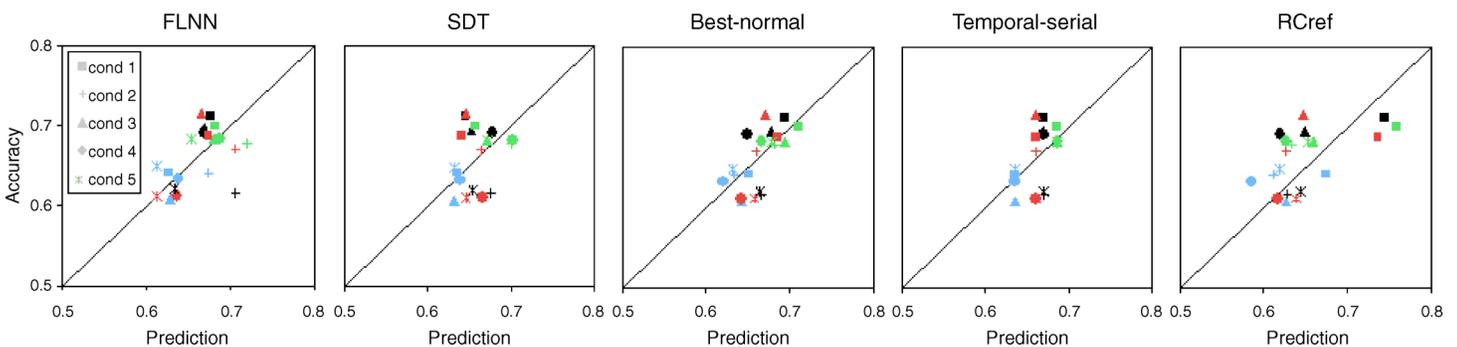


Figure 8. Participants' accuracy in [Experiment 4](#) vs. the models' prediction. Points with the same color belong to the same participant. Different marker shapes refer to different conditions. If a model could perfectly predict participants' accuracy, the various points should fall exactly on the diagonal line.

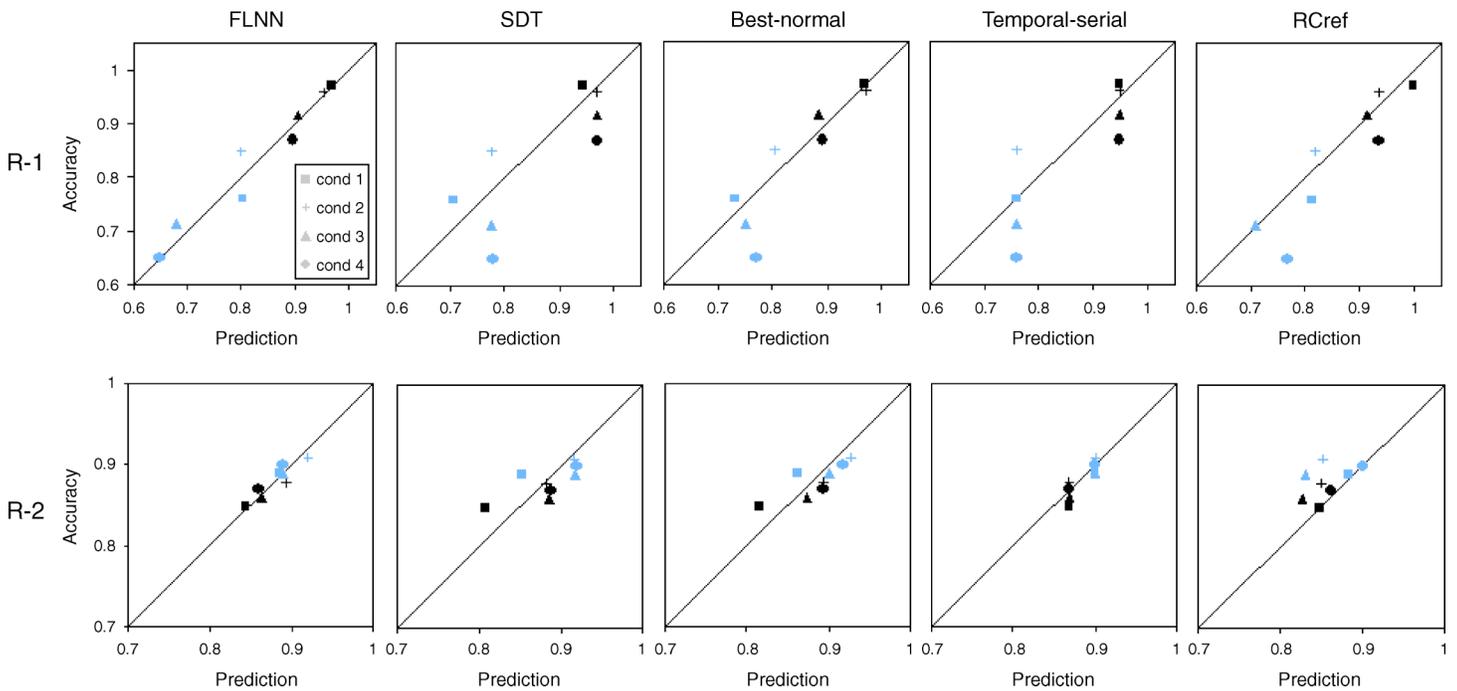


Figure 9. Participants' accuracy in Experiments R-1 and R-2 (Experiments 1 and 2, respectively, in Rosenholtz, 2001a) vs. the models' prediction. Points with the same color belong to the same participant. Different marker shapes refer to different conditions. If a model could perfectly predict participants' accuracy, the various points should fall exactly on the diagonal line.

experiments and passes both correlation coefficient tests in Experiment R-2, but neither in Experiment R-1. The saliency measure predicted the exact order of difficulty for one out of the two participants in Experiment R-1, and it passed the correlation coefficient test only for the participants in Experiment R-2. The SDT model could not predict the exact order of difficulty for any participant.

The observed accuracy of each participant is plotted against the predictions of each of the FLNN, SDT, best-normal, RCreff, and temporal-serial models in Figure 9, and the best-fit parameters, reduced chi-square measure, and chi-square test of each combination are reported in Table 2.<sup>6</sup> The FLNN passed the chi-square test for one participant in Experiment R-1 and for both participants in Experiment R-2. Additionally, it provided the lowest reduced chi-square measure for both experiments. The best-normal model also passed the chi-square test for one participant in Experiment R-1 and for both participants in Experiment R-2. The RCreff and the temporal-serial models passed the chi-square test for both participants in Experiments R-2 but for none in Experiment R-1. Finally, the SDT model passed only one chi-square test for one participant in Experiment R-2.

## General discussion

The main goal of this study was to test the ability of the cover measure and the FLNN model, suggested previously

for artificial intelligence (Avraham & Lindenbaum, 2005, 2006), to predict human performance in visual search tasks. These models were inspired by the finding that target–distractor similarity and similarity among distractors affect the efficiency of human visual search (e.g., Duncan & Humphreys, 1989). Specifically, these models can account for the findings that the difficulty of the search increases when the distractors are more similar to the target and less similar to each other. However, because the cover measure and the FLNN model were originally designed for computer vision, they assumed no difference between the displayed items and the observed input (i.e., there is no internal noise). To adapt these models to a human observer, we have extended them to consider the observed input as noisy. This allowed us to test the predictive abilities of these models for both orientation search (Experiments 1 and 2) and color search (Experiments 3 and 4) tasks and to compare them to the predictions of other prominent quantitative models of human visual search. As can be clearly seen in Table 4, our models better predicted human performance than did the other models in the comparison. The cover measure succeeded in accurately predicting the difficulty order for the largest number of cases (17 out of 21), and its correlation coefficients passed the significance test on the largest number of tests (16 out of 21). Similarly, the FLNN model passed the largest number of chi-square tests (15 out of 21), and it achieved the lowest reduced chi-square measure for the largest number of cases (10 out of 21).

	No. of participants	Order prediction		Correlation coefficient significance		Lowest $\chi^2/df$					Passed $\chi^2$ test				
		Cov	Sal	Cov	Sal	FLNN	SDT	B-N	T-S	RCref	FLNN	SDT	B-N	T-S	RCref
Experiment 1	5	5	4	5	0	4	0	0	0	1	4	0	2	0	1
Experiment 2	4	1	1	3	0	1	0	1	1	1	3	3	2	2	2
Experiment 3	4	4	1	2	1	1	0	0	0	3	2	0	1	0	1
Experiment 4	4	3	2	4	0	0	1	2	1	0	3	2	4	4	1
R-1	2	2	1	0	0	2	0	0	0	0	1	0	1	0	0
R-2	2	2	0	2	2	2	0	0	1	0	2	1	2	2	2
Total	21	17	9	16	3	10	1	3	3	5	15	6	12	8	7

Table 4. Summary of the various models' predictive abilities. From left to right: *Order prediction*—the number of participants for which the cover (Cov) and saliency (Sal) measures were able to predict the order of difficulty; *correlation coefficient significance*—the number of participants for which the predictions of these models achieved significant correlation coefficients; *lowest  $\chi^2/df$* —the number of participants for which the FLNN, the SDT, the best-normal (B-N), the temporal-serial (T-S), and the RCref models achieved the fit that gave the lowest reduced chi-square (relative to the other models); *passed  $\chi^2$  test*—the number of participants for which each of these models achieved a fit that was not rejected by the chi-square test.

The finding that our models' predictions were the closest to human performance suggests that these models were able to capture certain aspects of the processes underlying human visual search. Specifically, like the saliency and best-normal models, our models seem to capture the dependency of search efficiency on the relation between the target–distractor distance and the distance between the distractors themselves, but our models also highlight the importance of grouping-by-similarity processes. Several previous studies have dealt with the role of grouping processes in visual search (e.g., Bravo & Blake, 1990; Bundesen & Pedersen, 1983; Driver, McLeod, & Dienes, 1992; Farmer & Taylor, 1980; Humphreys, Quinlan, & Riddoch, 1989; Kahneman & Henik, 1977; Santhi & Reeves, 2004; Treisman, 1982). Most relevant to the present study is Duncan and Humphreys' (1989) similarity theory. The similarity theory suggests that each element in the display is assigned a weight indicating the strength of its activation as it competes with other elements over limited resources, and that the activations of strongly grouped elements tend to rise or fall together. Duncan (1995) refers to this tendency to treat grouped elements together as “weight linkage” and suggests that it can account for various visual search outcomes that seem to imply the involvement of perceptual grouping (e.g., Bundesen & Pedersen, 1983; Driver et al., 1992; Farmer & Taylor, 1980; Kahneman & Henik, 1977). The FLNN can be viewed as a computational implementation of such a “weight linkage” mechanism. Moreover, the fact that the FLNN could predict human performance relatively well suggests that the elements that are treated as a group are not necessarily identical, and that the degree of within-group heterogeneity depends on the target–distractor distance.

Another important difference between the FLNN and the best-normal and SDT-based models is the assumption that capacity is limited. In other words, when pre-

sentation time is limited and the number of items in the display is relatively large, not all the items can be processed to a similar degree. Although this assumption of the FLNN model seems to imply that visual search is a serial process, we do not think that this is the only possible instantiation of the model. The FLNN model is also consistent with a parallel search process. For example, consider a search process in which all the items are processed in parallel but capacity is limited and therefore the processing cannot be optimal for all items. To maximize performance, the items are assigned priorities, and the processing of items with high priority is facilitated. This parallel process will match the spirit of the FLNN model if priorities are assigned according to feature–space distances and if the priority assignment is dynamic. In that case, once an item is rejected, the priorities of other, sufficiently similar items are significantly reduced and the highest priority is assigned to the most dissimilar item. Such a dynamic change of priorities could be mediated by lateral interactions known to occur in the visual cortex (such as the connections between neurons whose preferred stimulus is similar) and some feedback connections (e.g., Gilbert, 1998). Whereas a serial process is probably more plausible in reference to overt attention—attentional processes that involve eye movements—a parallel process is more relevant in reference to covert attention—attentional processes that do not involve eye movements. Because the duration of the search display in this study (500 ms) allowed eye movements, the search was most likely a mixture of overt and covert attentional processes. In the future, we plan to tease these apart by restricting display duration on the one hand, and on the other hand, testing whether FLNN can also predict eye movements in a search task.

Although our models' predictions were the closest to human performance, they were nevertheless rejected in some cases. One possible reason might be that they work on the assumption that the feature spaces are linear. For the orientation search experiments, the models assume that the perceived difference between different orientations is proportional to the physical difference in degrees. This assumption might not be fully accurate. It was demonstrated, for instance, that orientation sensitivity is better around the principle meridians than around the oblique ones (the oblique effect; e.g., Andrews, 1967; Bouma & Andriessen, 1968). Additionally, Orban, Vandebussche, and Vogels (1984) measured the difference threshold (JND) for 15 different orientations and found that the JND increases as a function of obliquity, from the principal orientation up to 20° obliquity, and then levels off. In a similar manner, for the color search experiments, our models assume that differences in the  $L^*u^*v^*$  color space are proportional to the differences recognized by human observers. Although this assumption is commonly employed, there has been some recent criticism about the accuracy of this color system (e.g., Fairchild, 1998). The deviation from linear relations between the feature space employed by the models and the actual perceived feature space might be larger for color than orientation. If so, this can explain the fact that our models' predictions were better in the latter case. Hence, because the models' predictions are based on distances in feature space between the various search items, the predictions can only be as accurate as the feature space assumed by the models. The further our knowledge regarding the perceived feature space of various features advances, the more the accurate predictions provided by the models might be.

The fact that our models are rejected in some cases may also imply that they do not capture all the factors mediating human search performance. Clearly, the processes underlying human performance in visual search tasks are much more complex than those suggested by our models. For instance, our models do not take into consideration the spatial position of the stimuli, yet it was shown that targets appearing at peripheral locations are detected more slowly and less accurately than those appearing near the central fixation point (the eccentricity effect; Carrasco et al., 1995; Carrasco & Frieder, 1997). Another spatial consideration that may have more weight in improving the models predictions is the spatial proximity between different stimuli: Our models suggest that processes of grouping by feature similarity play an important role in visual search but since there is no coding of spatial position, the possible effects of proximity (Wertheimer, 1923) are currently not taken into account. In future work, we intend to modify the models tested here to consider the spatial location of the items and the spatial relations between them and test whether this significantly improves the models' predictive abilities.

## Additional visual search findings

Additional well-known phenomena in the literature of visual search, such as the set-size effect and search asymmetries, have not been dealt with in this paper. This section discusses how our models are related to these two phenomena and to the effect of distractor heterogeneity in an orthogonal direction. To fully account for these phenomena, the models must be extended to the case of two-dimensional feature spaces because their study usually involved items that vary on more than one dimension. Below we speculate about the manner in which our models can be extended to deal with two-dimensional spaces, but we leave the actual implementation of these extended models for future studies.

Regardless of the dimensionality, the FLNN and the cover models need only a relative measure of similarity between each pair of display elements as input. For one-dimensional cases, it is reasonable to assume that an appropriate similarity measure for two items is the difference in their feature values. However, when dealing with two dimensions or more, the appropriate similarity measure is less obvious. For instance, one has to first establish whether or not the two dimensions are separable (Garner, 1974). One possibility is to use the Euclidean distance as a measure of dissimilarity (Avraham & Lindenbaum, 2006): If  $d_1$  is the feature-wise distance between a pair of items in one dimension and  $d_2$  is the distance in the other dimension, then the Euclidean distance is  $\sqrt{d_1^2 + d_2^2}$ . Of course, this Euclidean distance is relevant only if empirical evidence demonstrates that the feature-wise distance of each dimension corresponds to the true perceptual distance. Moreover, intensive empirical work is needed to discover whether different weights are required for the different dimensions. In short, the exact implementation of our models for multidimensional cases may vary considerably, depending on the specific dimensions involved.

One common finding is a considerable set-size effect typically found for a search through items that differ on two dimensions (conjunction search) but not for a search through homogenous distractors (feature search; e.g., Carrasco & Yeshurun, 1998; Eckstein, 1998; Triesman & Gelade, 1980; Treisman & Paterson, 1984). Although a full test of our models' quantitative predictions with two-dimensional displays is beyond the scope of this study, qualitatively, the cover and the FLNN models can account for such findings. As described above, when the set-size increases,  $d_T$  decreases, and the corresponding cover increases. It can be shown that this effect is more pronounced with two dimensions.<sup>7</sup> We made a preliminary attempt to quantitatively account for the set-size effect by testing the ability of the FLNN model to predict the results reported in Eckstein (1998). Eckstein found a substantial set-size effect with a search for a target defined by a conjunction of orientation and contrast. There was no set-size effect with a search for a target defined by one

feature (orientation or contrast). He further demonstrated that an SDT-based model, but not the temporal-serial model, can account for these findings. Here, as a pilot test, the FLNN two-dimensional implementation assumes the Euclidean distance as the dissimilarity measure and uses three parameters:  $k$ , and one noise level for each of the two dimensions,  $\sigma_1$  and  $\sigma_2$ . For simplicity, the weights for the two dimensions were assumed to be equal. Figure 10 presents the FLNN predictions for the three participants of the original study. As can be seen, FLNN can predict that, for conjunction search, overall performance will be worse than for feature search; it can also predict the steeper slope as a function of set size in conjunction search. In future work, the various assumptions we employed here should be tested and a fourth parameter may be added to express the relative difference in the weights of the two dimensions. We believe that this would improve the predictions reported here. Also note that for this preliminary test the experimental results were estimated from the figures in Eckstein’s paper and therefore may not be accurate.

Another well-known finding in visual search is that of search asymmetry—a search for a specific target among specific distractors may be harder when their feature values are switched (e.g., Driver et al., 1992; Treisman & Gormican, 1988). Rosenholtz (2001b) suggested that many of the search asymmetries were in fact due to an asymmetrical experimental design. Our models can qualitatively account for such asymmetries, and they do so in a manner very similar to the saliency measure (Rosenholtz, 2001b; Rosenholtz, Nagy, & Bell, 2004). As an example, consider Ivry and Cohen’s (1992) finding that a search for a slow, horizontally oscillating target among fast, oscillating distractors is harder than the reverse case. Figures 11a and 11b depict the two-dimensional velocity space corresponding to these two cases, respectively (Rosenholtz, 2001b). FLNN would typically pick a

distractor as its initial item because there were several distractors and only one target. To reach the slow target (Figures 11a), a third step is required because, regardless of the distractor initially chosen, the farthest item is also a distractor. To reach the fast target (Figures 11b), only two steps are required because the target would be the farthest (feature-wise) from the first chosen distractor regardless of its identity.

Another “search asymmetry” example is Nagy and Cone’s (1996) finding that a more saturated target among less saturated distractors is easier to find than the reverse case. If the dark gray background is included as one of the relevant feature values (Figures 11c and 11d; Rosenholtz, 2001b; Rosenholtz et al., 2004), then, given the relatively large part of the display taken up by the background, it is likely that FLNN will choose it as the initial item. For the more saturated target (Figure 11c), the farthest item from the background is the target, while for the less saturated target (Figure 11d) it is a distractor.

Finally, Nagy and colleagues have carefully studied the effects of distractor heterogeneity on color search (e.g., Nagy, Neriani, & Young, 2005; Nagy & Thomas, 2003). One of their interesting findings is with regard to distractor heterogeneity in a direction orthogonal to that differentiating the target and the distractors. They found that performance was poorer in the heterogeneous case than in a homogenous case when the target color varied from trial to trial, (Nagy et al., 2005), but not when the target color was constant (Nagy & Thomas, 2003). Our models can qualitatively account for this finding if we assume that, when the color of the target was known, the participants could assign a low weight to the orthogonal dimension, but when the color of the target was unknown, the participants had to assign a higher weight to the other dimension, and heterogeneity hurt performance. These findings suggest therefore that the extension of our models to multidimensional spaces will have to consider a

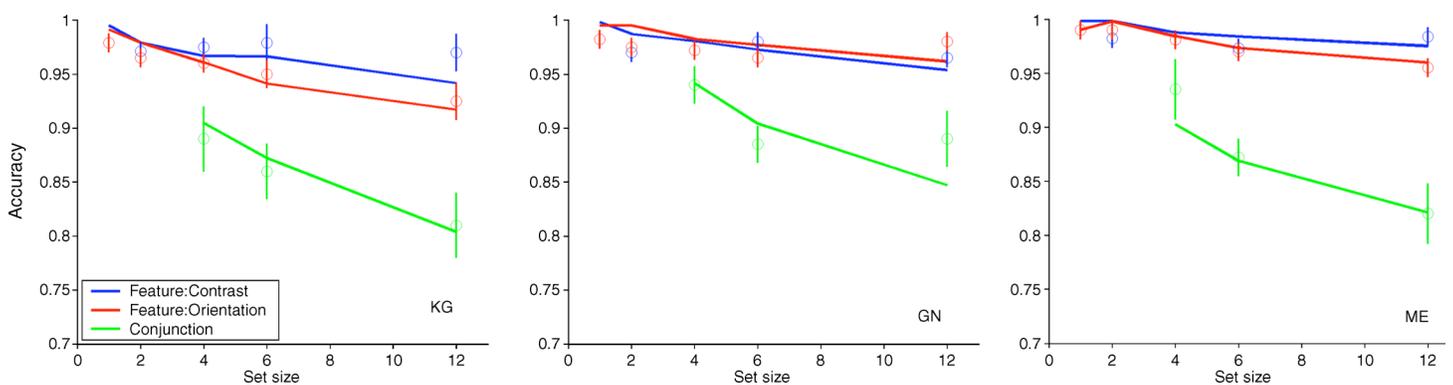


Figure 10. Estimated accuracy of the three participants in Eckstein (1998) as a function of set size for three search types: contrast feature search, orientation feature search, and conjunction search (orientation and contrast). The lines correspond to the best fit of the FLNN model based on a pilot extension of the model to a two-dimensional case.

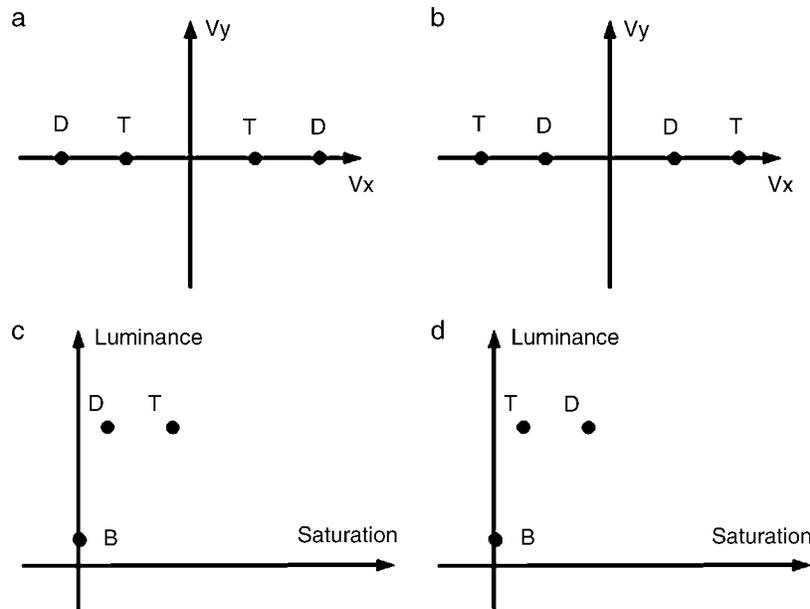


Figure 11. Two-dimensional feature spaces of (a) a search for a slow, horizontally oscillating target among fast, horizontally oscillating distractors; (b) a search for a fast, horizontally oscillating target among slow, horizontally oscillating distractors; (c) a search for a more saturated target among less saturated distractors; and (d) a search for a more saturated target among less saturated distractors. The target values are marked with a T, the distractors values are marked with a D, and when relevant, the background value is marked with a B.

dynamic assignment of weights, depending on task requirements and participant expectations.

## Conclusions

This study evaluated the ability of two novel visual search models to predict human performance in orientation and color search tasks. Our models' predictions were closer to human performance than those of other prominent quantitative models of visual search. These findings suggest that the distances in feature space between the various items in the search display, and specifically the relationship between target–distractor distance and the distance between the distractors themselves, are an important factor in search performance. These findings further suggest that grouping-by-similarity processes play a central role in visual search.

## Appendix A: A detailed description of the models

### Calculating the cover measure when internal noise is considered

The original cover measure (Avraham & Lindenbaum, 2005, 2006) was calculated as follows: First, the smallest difference between the target's feature value and a

distractor's feature value was measured and denoted  $d_T$ . Then, the cover measure is the number of segments of length  $d_T$  required to cover all the points representing the distractors in the feature-space. This original measure does not consider the observer's internal noise. Here we suggest a method for estimating the average cover measure when normally distributed internal noise is added to the observations. Given the feature values associated with the stimuli of a specific experimental condition and the internal noise variance  $\sigma^2$ , the cover measure is estimated as follows:

1. An observed target-present display is randomly generated by picking the feature value of each element from the normal distribution it belongs to (with the mean being its true displayed value and  $\sigma$  being a parameter denoting the level of noise).
2. The cover value is calculated for the specific generated case (using the original cover definition).
3. This is repeated  $N$  times, resulting in  $N$  Cover values.
4. The suggested prediction is the average over those calculated values.

All the predictions reported in this paper used  $N = 1000$ .

### The FLNN model: Original formulation

The original FLNN algorithm (Avraham & Lindenbaum, 2005, 2006) starts by choosing one of the displayed elements randomly. If the currently selected element is

not the target, another element is selected. This would be the element that is farthest (feature-wise) from all previously selected items. Let  $x_1, \dots, x_n$  be the feature values corresponding to the displayed elements and let  $x_{s1}, \dots, x_{sm}$  ( $m < n$ ) be the feature values corresponding to the already selected elements. For each element  $x_i$  that has not yet been selected, the algorithm finds the closest distance to a selected item's feature value:  $d_{\min}(x_i) = \min_{j=1, \dots, m} |x_i - x_{sj}|$ , and the next selected element is the one for which  $d_{\min}$  is maximum. This procedure is repeated until the target is found.

### The FLNN model: An extended version accounting for internal noise and limited time

The extended FLNN performance in a 2IFC experiment is found by applying the FLNN algorithm on stimuli from a target-present display and a target-absent display for a limited number of steps. The prediction is calculated in the following way:

1. An observed target-present display is randomly generated by picking the feature value of each element from the normal distribution to which it belongs (with the mean being its true displayed value and  $\sigma$  being a parameter denoting the level of noise).
2. In a similar way, an observed target-absent display is generated.
3. It is assumed that in the limited presentation time of the display, only  $k$  elements can be processed. Therefore, the FLNN algorithm is simulated for  $k$  steps on each of the two generated “displays,” and we get  $k$  selected elements for each display.
4. The algorithm identifies the display that contains the stimulus that is most similar (in terms of feature-values) to the true target value (without noise), out of the  $2k$  selected stimuli, as the target-present display.
5. If the algorithm points to the target-present generated display, this is counted as a success; otherwise, it is a failure.
6. All the above steps are repeated many times (10,000 in our case) and the ratio of success is the model's prediction.

Note that this model has two parameters for each participant:  $k$  (the number of display elements examined by the observer in the limited presentation time) and  $\sigma$  (the internal noise level of the observer). We allow non-integer values for  $k$ . If  $N < k < N + 1$ , where  $N$  is an integer, then  $N$  elements are considered in some trials and  $N + 1$  elements are considered in the other trials. For example, if  $k$  equals 5.8, then in 20% of the trials 5 items are considered and in 80% of the trials 6 items are considered.

## Appendix B

### Chi-square test and reduced chi-square measure of fit

We employed the reduced chi-square ( $\chi^2/df$ ) measure because it enables the comparison between models that use a different number of parameters (Taylor, 1982). A model with more parameters gets a lower fitting grade than a model with fewer parameters that gives the same prediction.

$$\chi^2/df = \frac{1}{c-p} \sum_{i=1}^c \frac{(\text{Acc}_i - \text{Prediction}_i)^2}{SE_i^2}, \quad (\text{B1})$$

where  $c$  is the number of conditions,  $p$  is the number of model parameters,  $\text{Acc}_i$  is the accuracy of the participant on condition  $i$ ,  $\text{Prediction}_i$  is the prediction of the model for condition  $i$ , and  $SE_i$  is the standard error of the participant's accuracy for the  $i$ th condition. The lower the  $\chi^2/df$  value is, the better the model can predict the results. The chi-square test reports whether the probability of obtaining a  $\chi^2$  larger than the one measured, given that the data followed the model and taking into account the degrees of freedom  $df$ , is higher than 0.05. If it is not, the model is rejected.

## Acknowledgments

We thank the anonymous reviewers for their helpful remarks. We thank Ruth Rosenholtz for her helpful comments on an earlier draft of the manuscript and for sharing her data sets.

Commercial relationships: none.

Corresponding author: Tamar Avraham.

Email: tammya@cs.technion.ac.il.

Address: Computer Science Department, Technion, Technion City, Haifa 32000, Israel.

## Footnotes

<sup>1</sup>Originally, the saliency measure was suggested also for multidimensional feature spaces, in which case it refers to covariance rather than to standard deviation. For further details see Rosenholtz (1999).

Our implementation of the RCref model followed Rosenholtz (2001a). Thus, an item was compared to the reference target on 30% of the times.

<sup>3</sup>Originally, the cover measure was defined for situations in which there can be more than one target in the display and for the more general case of multidimensional search (Avraham & Lindenbaum, 2005, 2006). In the context of this paper, the simplified definition described above suffices.

<sup>4</sup>The significance test of the correlation coefficient examines whether the probability of getting the observed correlation coefficient by chance is  $<0.05$ .

<sup>5</sup>The chi-square test penalizes FLNN for using two parameters (as opposed to the other models, which use only one) by demanding a closest fit between the prediction and the data. This motivated us to check whether the two parameters ( $\sigma$  and  $k$ ) covary. The optimization surfaces suggested that the parameters are not statistically dependent and that good fits require a combination of both parameters.

<sup>6</sup>There are minor differences between the predictions of the SDT, best-normal and RCref reported here, and those reported in Rosenholtz (2001a) because we have chosen the fits that minimize chi-square, whereas Rosenholtz minimized the sum of square differences.

<sup>7</sup>For one-dimensional feature space, we defined the cover as the number of  $d_T$ -long segments required to cover all the points in the feature space representing the distractors. For two dimensions, the cover is defined as the number of disks with diameter  $d_T$  required to cover the distractor points (see Avraham & Lindenbaum, 2006). If  $d_T$  decreases by 50% due to the increase in set size, then, for a one-dimensional case, a segment of length 2 is divided into two segments of length 1. Yet, for a two-dimensional case, even 4 disks of diameter 1 cannot cover a disk of diameter 2.

## References

- Andrews, D. P. (1967). Perception of contour orientation in the central fovea. Part 1: Short lines. *Vision Research*, 7, 957–997. [PubMed]
- Avraham, T., & Lindenbaum, M. (2005). Inherent limitations of visual search and the role of inner-scene similarity. In *Proceedings of WAPCV04—2nd International Workshop on Attention and Performance in Computational Vision. Lecture Notes in Computer Science* (vol. 3368, pp. 16–28). Berlin: Springer-Verlag.
- Avraham, T., & Lindenbaum, M. (2006). Attention—based dynamic visual search using inner-scene similarity: Algorithms and bounds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28, 251–264. [PubMed]
- Bauer, B., Jolicoeur, P., & Cowan, W. B. (1996). Visual search for colour targets that are or are not linearly separable from distractors. *Vision Research*, 36, 1439–1465. [PubMed]
- Bergen, J. R., & Julesz, B. (1983). Rapid discrimination of visual patterns. *IEEE Transactions on Systems, Man and Cybernetics*, 13, 857–863.
- Bouma, H., & Andriessen, J. J. (1968). Perceived orientation of isolated line segments. *Vision Research*, 8, 493–507. [PubMed]
- Bravo, M., & Blake, R. (1990). Preattentive vision and perceptual groups. *Perception*, 19, 515–522. [PubMed]
- Bundesen, C., & Pedersen, L. F. (1983). Color segregation and visual search. *Perception & Psychophysics*, 33, 487–493. [PubMed]
- Cameron, E. L., Tai, J. C., Eckstein, M. P., & Carrasco, M. (2004). Signal detection theory applied to three visual search tasks—Identification, yes/no detection and localization. *Spatial Vision*, 17, 295–325. [PubMed]
- Carrasco, M., Evert, D. L., Chang, I., & Katz, S. M. (1995). The eccentricity effect: Target eccentricity affects performance on conjunction searches. *Perception & Psychophysics*, 57, 1241–1261. [PubMed]
- Carrasco, M., & Frieder, K. S. (1997). Cortical magnification neutralizes the eccentricity effect in visual search. *Vision Research*, 37, 63–82. [PubMed]
- Carrasco, M., McLean, T. L., Katz, S. M., & Frieder, K. S. (1998). Feature asymmetries in visual search: Effects of display duration, target eccentricity, orientation and spatial frequency. *Vision Research*, 38, 347–374. [PubMed]
- Carrasco, M., Penpeci-Talgar, C., & Eckstein, M. (2000). Spatial covert attention increases contrast sensitivity across the CSF: Support for signal enhancement. *Vision Research*, 40, 1203–1215. [PubMed]
- Carrasco, M., & Yeshurun, Y. (1998). The contribution of covert attention to the set-size and eccentricity effects in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 673–692. [PubMed]
- Driver, J., McLeod, P., & Dienes, Z. (1992). Motion coherence and conjunction search: Implications for guided search theory. *Perception & Psychophysics*, 51, 79–85. [PubMed]
- Duncan, J. (1995). Target and non-target grouping in visual search. *Perception & Psychophysics*, 57, 117–120. [PubMed]
- Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, 96, 433–458. [PubMed]
- Eckstein, M. P. (1998). The lower visual search efficiency for conjunctions due to noise and not serial attentional processing. *Psychological Science*, 9, 111–118.

- 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*, 425–451. [PubMed]
- Enns, J. T., & Rensink, R. A. (1992). *VScope (TM) software and manual: Vision testing software for the Macintosh*. University of British Columbia, Vancouver, Canada: Micropsych Software.
- Fairchild, M. D. (1998). *Color appearance models*. Massachusetts: Addison Wesley Longman, Inc.
- Farmer, E. W., & Taylor, R. M. (1980). Visual search through color displays: Effects of target-background similarity and background uniformity. *Perception & Psychophysics*, *27*, 267–272. [PubMed]
- Garner, W. R. (1974). *The processing of information and structure*. New York: Wiley.
- Geisler, W. S., & Chou, K. L. (1995). Separation of low-level and high-level factors in complex tasks: Visual search. *Psychological Review*, *102*, 356–378. [PubMed]
- Gilbert, C. D. (1998). Adult cortical dynamics. *Physiological Review*, *78*, 467–485. [PubMed] [Article]
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Krieger.
- Humphreys, G. W., Quinlan, P. T., & Riddoch, M. J. (1989). Grouping processes in visual search: Effects with single and combined-feature targets. *Journal of Experimental Psychology: General*, *118*, 258–279. [PubMed]
- Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*, 1254–1259.
- Ivry, R., & Cohen, A. (1992). Asymmetry in visual search for target defined by differences in movement speed. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 1045–1057. [PubMed]
- Kahneman, D., & Henik, A. (1977). Effects of visual grouping on immediate recall and selective attention. In S. Dornic (Ed.), *Attention and performance VI* (pp. 307–332). Hillsdale, NJ: Erlbaum.
- Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology*, *4*, 219–227. [PubMed]
- Nagy, A. & Cone, S. M. (1996). Asymmetries in simple feature searches for color. *Vision Research*, *36*, 2837–2847. [PubMed]
- Nagy, A. L., Neriani, K. E., & Young, T. L. (2005). Effects of target and distractor heterogeneity on search for a color target. *Vision Research*, *45*, 1885–1899. [PubMed]
- Nagy, A. L., & Thomas, G. (2003). Distractor heterogeneity, attention, and color in visual search. *Vision Research*, *43*, 1541–1552. [PubMed]
- Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.
- Orban, G. A., Vandenbussche, E., & Vogels, R. (1984). Human orientation discrimination tested with long stimuli. *Vision Research*, *24*, 121–128. [PubMed]
- Palmer, J. (1994). Set-size effects in visual search: The effect of attention is independent of the stimulus for simple tasks. *Vision Research*, *34*, 1703–1721. [PubMed]
- Palmer, J., Ames, C. T., & Lindsey, D. T. (1993). Measuring effects of attention on simple visual search. *Journal of Experimental Psychology*, *19*, 108–130. [PubMed]
- Palmer, J., Verghese, P., & Pavel, M. (2000). The psychophysics of visual search. *Vision Research*, *40*, 1227–1268. [PubMed]
- Papoulis, A., & Pillai, S. U. (2002). *Probability, random variables, and stochastic processes* (4th ed.). New York: McGraw-Hill.
- Rosenholtz, R. (1999). A simple model predicts a number of motion popout phenomena. *Vision Research*, *39*, 3157–3163. [PubMed]
- Rosenholtz, R. (2001a). Search asymmetries? What search asymmetries? *Perception & Psychophysics*, *63*, 476–489. [PubMed]
- Rosenholtz, R. (2001b). Visual search orientation among heterogeneous distractors: Experimental results and implications for signal-detection theory models of search. *Journal of Experimental Psychology*, *27*, 985–999. [PubMed]
- Rosenholtz, R., Nagy, A. L., & Bell, N. R. (2004). The effects of background color on asymmetries in color search. *Journal of Vision*, *4*(3):9, 224–240, <http://journalofvision.org/4/3/9/>, doi:10.1167/4.3.9. [PubMed] [Article]
- Santhi, N., & Reeves, A. (2004). The roles of distractor noise and target certainty in search: A signal detection model. *Vision Research*, *44*, 1235–1256. [PubMed]
- Taylor, J. R. (1982). *An introduction to error analysis*. Mill Valley, CA: University Science Books, Oxford University Press.
- Travis, D. (1991). *Effective colour displays: Theory and practice*. London, UK: Academic Press.
- Treisman, A. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception and Performance*, *8*, 194–214. [PubMed]
- Treisman, A. M., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, *12*, 97–136. [PubMed]

- Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, *95*, 15–48. [[PubMed](#)]
- Treisman, A., & Paterson, R. (1984). Emergent features, attention, and object perception. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 12–31. [[PubMed](#)]
- Tsotsos, J. K., Culhane, S. M., Wai, W. Y. K., Lai, Y., Davis, N., & Nuflo, F. J. (1995). Modeling visual attention via selective tuning. *Artificial Intelligence*, *78*, 507–545.
- Vergheese, P., & Nakayama, K. (1994). Stimulus discriminability in visual search. *Vision Research*, *18*, 2453–2467. [[PubMed](#)]
- Wertheimer, M. (1923). Untersuchungen zur lehre von der gestalt. *Psychologische Forschung*, *4*, 301–350.
- Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin and Review*, *1*, 202–238.