The roles of distractor noise and target certainty in search: A signal detection model

Nayantara Santhi* and Adam Reeves

Department of Psychology, Northeastern University,

360 Huntington Ave., Boston MA 02115

(Tel: 617-373-4708; email reeves@neu.edu)

*Now at: Harvard Medical School,

Division of Sleep Medicine, Brigham & Women's Hospital,

221 Longwood Ave., Boston MA 02115.

(Tel: 617-732-7531; email nsanthi@rics.bwh.harvard.edu ; reeves@neu.edu

This work partially fulfilled the requirements for a Ph.D in Experimental Psychology from Northeastern University for the first author (Santhi, 2000). The authors would like to thank Rhea Eskew, Todd Horowitz, Philip Smith, and an anonymous reviewer, for valuable input, and Stefano DeCaro for help with the software.

revised ms

ABSTRACT

Observers searched arrays of briefly-presented near-isoluminant colored disks for a single disk of known color (feature search) or unknown color (oddity search). Speed and accuracy were converted to a single, model-based measure of performance (Perf), in units of d'^2 per sec of latency. Perf decreased with set size in feature search and increased in oddity. In both types of search, grouping the distractors, making them homogeneous in color, and reducing their saturation, all increased Perf. These commonalities suggested an SDT-based model in which distractors increase noise in the same way in both types of search. However, in oddity, though not in feature search, distractors must be attended and so adding distractors also boosts the effective target contrast, overcoming the added noise. A model with two free parameters for noise and one for attention accounted for every combination except for oddity searches among heterogeneous grouped distractors.

Introduction

This paper reports experiments in which observers searched for a single target which was either known in advance ('feature search') or not known in advance ('oddity search'), set among various numbers of distractors. Previous research has shown that when a single known feature defines the target, performance is efficient, either independent of set size ('pop-out') or declining slightly with set size (Treisman, 1982; Farmer & Taylor, 1980; Duncan & Humphreys, 1989; Palmer, 1994; Palmer, Ames, & Lindsey, 1993; Palmer, Verghese & Pavel, 2000). Performance, however, <u>improves</u> with set size for target-unknown search, for example a search for an odd item (Bacon & Egeth, 1991; Bravo & Nakayama, 1992). Stimuli and tasks in oddity have typically been more complex than those in feature search; for example, Bravo and Nakayama's observers reported a missing corner of an oddly-colored search item presented among homogeneously colored distractors. Thus these contrasting effects of set size may involve stimulus complexity. Our hypothesis, however, is that feature and oddity searches differ primarily because of the knowledge of the signal available to the observer. Knowledge of the signal allows observers to optimize d' by matching their input analyzers to the signal. Filtering out too much and losing signal, or filtering too little and including excess noise, will lower d'. (See McCabe, Caelli, West, & Reeves, 2000, for a candidate spatio-chromatic matched filter, and Dosher & Lu, 2000, for tests of noise exclusion.) The current research tested this hypothesis using the same stimuli in both oddity and feature searches. We also tested its generality by varying several distractor properties known to affect performance.

Efficient feature search can result from a preattentive process operating in parallel over the visual field (Neisser, 1967; Gardner, 1973; Cave & Wolfe, 1990). Such a process may be modeled by signal-detection theory, or SDT (Green & Swets, 1966) with a single decision per trial (e.g. Palmer, Verghese & Pavel, 2000). In contrast, slow or inefficient searches (Triesman, 1982, 1986, 1991; Wolfe, 1994; Duncan & Humphreys, 1989) can result from multiple decisions on each trial, as evidenced by multiple fixations (Geisler & Chou, 1995) or attention shifts (Reeves & Sperling, 1986), or as demanded by difficult conjunction or disjunction searches (Eckstein, Thomas, Palmer, Shimozaki, & Steven, 2000). We attempted to elicit just one decision per trial by using brief displays and simple, disk-shaped targets defined by a single feature (chromaticity). Also, we kept targets and distractors clearly distinct in chromaticity to avoid the inefficient searches found with close colors (Green & Anderson, 1956; Carter & Carter, 1981; Carter, 1982; Nagy & Sanchez, 1990; Duncan, 1989). Thus we anticipated finding efficient searches to which we could apply an SDT model in which all items are processed in parallel, exhaustively on every trial. SDT has been applied to search numerous times since Palmer (1994), in particular to feature searches in color space (Nagy and Thomas, 2003), but our attempt to measure and characterize both feature and oddity searches within the same framework is new. The parallel SDT-based model we developed is non-standard in that it applies to both speed and accuracy, not to just one or the other, so we next remind the reader of both of these aspects of the Yes/No visual search task.

Speed and accuracy in visual search: Four measures (T , Tbias, d', and c).

Performance in visual search is generally measured with response time (RT) or errors, but not both. However, the observer's Yes/No decisions both occur over time and are prone to error (e.g. Ratcliff & Rouder, 1998). We therefore analyzed search using RT to infer processing time (T) and error rates to estimate sensitivity (d'). We also estimated the bias (Tbias) towards faster

Yes than No decisions, and the bias (c, in SDT) towards reporting Yes less often than No. To infer processing (observation + decision) time, we assume that

$$RT = T + RTo, \qquad (1)$$

where RTo is the 'residual' time to process the sensory input and generate a motor output, and is assumed not to be correlated with T. RTo was estimated from the median simple reaction time to targets presented with no distractors. (Technically Equation 1 relies on the additivity of means, but medians avoid contamination by outliers.) RT was estimated from the correct reaction time to targets with distractors (i.e., in search), being the average of the median RT on hit trials (RThit) and the median RT on correct rejection trials (RTcr). Equation 1 implies that the processing time, T, is the difference between two measurable quantities, RT and RTo, as long as RT > RTo. Our display durations were brief, typically less than T, but we assume displays are held in a visual information store (Sperling, 1960, 1963) so that processing continues after display termination until the time the motor command is given.

In self-terminating models, the differences between RThit and RTcr are of crucial importance, but in our model this is not so; all items are processed exhaustively whether the target is present or not. However, RTcr is usually slower than RThit, so it is important to know whether averaging them loses information. We therefore defined a *latency bias*, Tbias, equal to the slowness of correct No decisions relative to correct Yes decisions;

$$Tbias = (RTcr - RTo) / (RThit - RTo).$$

We have not encountered this measure before. To anticipate the results, "No" decisions were slower than "Yes" decisions in all the experiments reported here, but Tbias, which averaged to 1.7, hardly varied with set size or condition. Constancy of Tbias means that differences between correct Yes and No RTs can be characterized by one number. Knowing this, RThit and RTcr can be averaged to estimate processing time in Equation 1 without loss of further information.

To estimate sensitivity we computed d' as in Green & Swets (1966):

$$d' = z(p_{hit}) - z(p_{false\ alarm}), \qquad (2)$$

where z(p) is the z-score corresponding to p. Equation (2) assumes an equal-variance model of d', appropriately so since all the stimuli were about equally visible and the signal (target) stimulus was not added to the display but rather replaced one of the distractors.

We also computed the criterion c, where $c = -[z(p_{hit}) + z(p_{false\ alarm})]/2$, with $c>0$ reflecting a tendency to report 'No' more often than 'Yes'. To anticipate results, c was independent of set size in all our Experiments. Indeed c was close to optimal ($-0.2 < c < 0.2$) in every condition of every experiment, excepting one condition in Experiment 3.1 in which c averaged to -0.4 but was still flat across set size. This was as hoped for, since participants were told that there were as many present as absent trials, and that false alarms and misses were equally bad. As c was close to zero, we do not report the analysis of this measure here, but rather refer the reader to Santhi (2000) for details.

In developing the model we took advantage of the fact that d' and T convey all the information needed to describe the data. A more complex model would be required for data in which the bias measures, c and Tbias, varied systematically.

Signal and Noise in visual search: a phenomenological model

According to SDT, sensitivity (d') depends on the signal/noise ratio. We employed this definition to generate a phenomenological model in which sources of signal and noise are used to generate predictions of the signal/noise ratio. The feature searches, and fits to them, served to validate this (descriptive) model and our methods by replicating well-known effects. The oddity search data are new and permitted us to extend the model. The phenomenological model is not a processing model and the experiments were not designed to be analytical with respect to summation or maximum decision models, but we discuss these in an appendix (Appendix 3).

Signal. Signal strength depends on the target and on the observation period, T. In our experiments, target properties like size and shape were fixed [FOOTNOTE 1], so only target contrast, c_{con}, mattered. Target contrast certainly depends on target-field contrast (e.g., pink-grey, for a pink target flashed on a grey field) and possibly on target-distractor contrast (e.g., pink-blue, if a pink target is presented with blue distractors). In the model, the signal is

$$\text{signal} = T\, c_{con}, \quad \text{for } T < \text{Tcrit}. \tag{4}$$

Thus, doubling either the observation time or the contrast doubles the signal, implying complete integration of information during the critical period, Tcrit. We have not measured Tcrit, the duration over which integration is complete. However, published speed-accuracy trade-off (SATO) curves for uncued feature searches using Gabor patches of fixed luminance contrast show near-linear increases of d' with latency until d' saturates at 470 msec after the SATO curve leaves the baseline (Carrasco & McElree, 2001, Fig. 3, top; these authors used exponentials to fit the full data set since longer observation times produce diminishing returns.) Our near-isoluminant disks were presumably processed more slowly than their luminance-defined stimuli, so Tcrit for our stimuli may have been in excess of 470 ms. At any rate, all our estimates of T were less than 436 msec, so Equation 4 is adequate for our data.

If the target is known in advance, as in feature search, the visual system can optimize target detection by using a filter which processes only the neighborhood of the target in the feature (color) space. As our distractors were far from the target in hue space, only target-field contrast should affect feature search. To model this idea, we assume that in feature search, the effective target contrast, c_{con}, is just

$$c_{con} = c_{field}, \qquad (5),$$

where c_{field} is the purely local contrast between the target and the immediately surrounding grey field. In oddity, this relation will change to include the effects of attended-to distractors.

Noise. We define distractor-evoked noise, σ^2_E, to include both noise in the distractors themselves and noise generated in the visual system in response to the distractors. We lump together all other sources of noise, such as momentary fluctuations in the visual system (intrinsic noise) and random Poisson fluctuations in the display (Reeves, Wu, & Schirillo, 1997), and call this 'distractor-independent' noise, σ^2_I. This source of noise is necessary to account for imperfect performance in the absence of distractors. The total noise (σ^2_{tot}) is then the sum of the independent noise and the noise evoked by each of the m distractors, during the period T;

$$\sigma^2_{tot} = T\,(m\sigma^2_E + \sigma^2_I) \qquad (6).$$

Here, σ^2_E is the noise evoked by each distractor, σ^2_I is the distractor-independent noise, and all noise sources are assumed to be independent Gaussian random variables. The total noise from m distractors is $m\sigma^2_E$ because σ^2_E is assumed constant, independent of location in color space.

An alternative to the model just described is a Maximum model in which the observer responds to the greatest, rather than the sum, of the inputs. Although we are primarily concerned to present a phenomenological model, and we did not attempt any analytical experiments designed to compare summation with maximum, we nevertheless attempted to fit various Maximum models, as explained in Appendix 3. To anticipate, we found that summation fit better than any of the Maximum models, and as it is simpler, we discuss the model in that form.

However, we do not commit to either summation or maximum as describing the underlying process, and we do not imply that some other form of Maximum model might not work well.

Performance Index (Perf) and the fundamental model Equation

Together, Equations 3-6 imply that

$d' = Tc_{con} / \mathrm{SQRT}(T\,m\sigma^2_E + T\sigma^2_I)$. PRINTER: Please USE SQUARE-ROOT SIGN

Both counting and timing models imply that d' = T(s-n) / SQRT(Tn), where n is the noise and s-n is the difference between signal and noise (Green & Luce, 1973). Our model is of this form, specialized to visual search by including set size. Bringing d' and T to the left and squaring,

$$\mathrm{Perf} \equiv (d')^2/T = (c_{con})^2 / (m\sigma^2_E + \sigma^2_I) \qquad (7)$$

PRINTER: Please USE 3 EQUAL BARS FOR $\equiv$

The left side of this equation defines ($\equiv$) an index of performance, Perf, or $(d')^2/T$, in units of information $(d')^2$ per unit time. Perf combined the dependent variables of accuracy and response time in a rational manner; as d' increases or RT drops, Perf increases from a base of 0 (for random responding). The right-hand side contains only independent variables, related to the stimulus (c_{con} and m) and to the noise (σ^2_E and σ^2_I).

Perf has been derived before from various models in which information accumulates progressively up to the critical period, Tcrit (Swensson & Thomas, 1974). If the evidence provided by the stimulus is fixed, Perf is constant in these models, since the resulting speed-accuracy trade-off obeys, to a close approximation, the law that $(d')^2$ is proportional to T (Swensson & Thomas, 1974). We regard Perf as useful, given a target present/absent decision process which occurs over time, which is error prone, and which is subject to the restriction in

Equation 4 (that T < Tcrit). Indeed changes in RT do relate to changes in sensitivity in our data; e.g. the overall correlation between mean RT and $(d')^2$ across all conditions was -0.73.

Many researchers treat RT and accuracy as distinct measures of performance, but this can be highly misleading (e.g. Harris, Shaw, and Bates, 1979). The low error rates in RT-oriented studies of visual search are especially deceptive; for example, an easily overlooked drop in error rate from 5% to 1%, with no change in c, will double $(d')^2$, a large effect compared to the typical change in RT. The influence of serial models in visual search research has been strong, accounting for the tendency of many investigators to study intercept and slope effects on RT while discounting changes in error rate, even when analyzing efficient (thought-to-be parallel) search. However, if both RTs and errors vary systematically with set size, then the exact slope relating RT to set size has less meaning, and a combined measure such as Perf is preferable. A noticeable feature of Perf is that slow or inaccurate search would count as inefficient (i.e., low Perf), even if RT was flat over set size and would count as efficient in the guided search model (Wolfe, 1994).

Model predictions

We tested the model predictions for Perf expressed in Equation 7 by systematically varying set size (m), σ^2_E, and c_{con}, as explained below.

The Model's treatment of Heterogeneity and Grouping

Compared to homogeneously-colored distractors, heterogeneously-colored distractors impair feature search (Farmer & Taylor, 1980; Duncan & Humphreys, 1989; Triesman, 1982; Nagy, 1999). The model predicts this effect, as heterogeneity increases total distractor noise.

Grouping distractors counteracts their adverse effect on search (Carter, 1982; Triesman, 1982; Farmer and Taylor, 1980; Bundesen & Pedersen, 1983; Humphreys, Quinlan, & Riddoch, 1989). D'Zmura, Lennie, & Tiana (1997) even obtained efficient search for color-shape conjunctions when their stimuli could be grouped into figure and ground. Grouping is not entirely automatic, in that an attentionally-demanding secondary task can disrupt pre-attentive grouping (Ben-Av, Sagi, & Braun, 1992). Grouping tends to unitizes the distractors (Duncan, 1984, 1995; Bravo & Blake, 1990), so that the effective set-size (m) is the number of distinct groups in grouped displays, rather than the number of distractors as in ungrouped displays. The model predicts that the effects of heterogeneity and grouping depend on the products of effective set-size (m) and evoked noise (σ^2_E).

The Model's treatment of color space

The model uses terms involving hue contrast. These terms depend on the Euclidean distances (FOOTNOTE 2) between pairs of contrasting colors, not on their absolute locations in color space. Thus, had the experiments been run on a colored field to which the observer was fully adapted, the contrasts would change but the model would be otherwise unaffected. On the grey field that we used, the Euclidean distances of the disks from the grey field correlate with their saturations. Thus the model predicts that feature search will improve with more saturated targets, increasing c_{field}, and with less saturated distractors, decreasing σ^2_E.

The Euclidean distance assumption may be challenged on at least four grounds. First, linearly separable stimuli produce more efficient search than do collinear ones, even if Euclidean distances are equated (D'Zmura, 1991). However, linearly separability speeds search only at short color distances, those near threshold (Bauer, Joliceour & Cowan, 1996). Since we used larger color distances we anticipated finding no linear separability effect, but we tested this anyway as any such effect would disqualify our model, and indeed there was no such effect (Experiment 2.2, control). Second, color space is categorized into color regions by color name, some of which are basic ('red') and some not ('aquamarine'). Any independent effect of color name on search would violate the Euclidean distance assumption. However, Smallman & Boynton (1990) found that search efficiency does not depend on whether the color names are basic. They also showed that Euclidean distances between colors controlled search times, in precise agreement with our model. Third, color asymmetries (e.g., search for red among pink distractors is faster than for pink among red) appear to violate any distance metric, but Rosenholtz (2001b) showed that the appearance of asymmetry disappears when the background is taken into account. Fourth, it is conceivable that search operates independently on the two cardinal directions of color space, but this is not so; search on a grey field can occur along any direction in color space (Nagy and Thomas, 2003).

General Method

We varied distractor heterogeneity (homogeneous vs. heterogeneous), distractor grouping (ungrouped vs. grouped), and distractor type (saturated vs.desaturated), in both types of search. Table 1 lists the different experiments with their experimental conditions. The last column of the table indicates the results in summary form.

Table 1: Summary of Experimental Conditions and Main Results

Exper-iment	Target certainty	Distractor Heterogeneity	spatial arrangement	Distractor saturation	Summary of Results*
1.1	Feature	Heterogeneous & Homogeneous	Ungrouped	Saturated	Hetero Ungr - Homog Ungr 0
1.2	Feature	Heterogeneous	Grouped & Ungrouped	Saturated	Hetero Ungr - Hetero Gr 0
1.3	Feature	Heterogeneous	Proximity vs. Sim+Prox.	Saturated	Hetero Prox - Hetero Sim+Prox 0
2.1	Feature	Homogeneous	Ungrouped	Desaturated	Homog Ungr Desat 0
2.2	Feature	Heterogeneous	Grouped & Ungrouped	Saturated & Desaturated	Hetero Ungr Sat - others 0
3.1	Oddity	Homogeneous	Ungrouped	Saturated & Desaturated	Homog Ungr 0
3.2	Oddity	Heterogeneous	Grouped & Ungrouped	Saturated & Desaturated	Hetero: all cases +

* Performance increases (+), does not change (0), or decreases (-), with set size.

Participants.

Participants were males and females between 18 and 24 years of age, with normal or corrected vision (Snellen acuity of 20/20 or better), and with normal color vision on the Ishihara color plates. They were naïve to the purposes of the experiments. Different participants were used in each experiment. They were recruited from the Northeastern University subject pool, and received credit for participation regardless of outcome.

Monitor Calibration.

Displays were presented on a ViewSonic CRT monitor (640 by 480 pixels; 8 bits), viewed from 60 cm away. The monitor was controlled by an IBM computer driven at 80 Hz and programmed in QBASIC run under DOS so that the software could be synchronized with each frame (Santhi, 2000). The monitor was lit indirectly by two shaded 40-Watt tungsten bulbs. The contrast and brightness control knobs were suitably adjusted and taped down. We recorded the levels of the red and green guns needed to match a monochromatic 580 nm yellow field when the blue gun was off, and the level of the blue gun which, when added to this yellow field, generated a grey (neither yellow nor blue) field. The program used these levels to generate both the grey field and the disks. The display was then calibrated with a Minolta (X,Y, Z) colorimeter. The X,Y,Z values for each stimulus were converted to (Y, u', v') CIE co-ordinates (Wyszecki and Stiles, 1982, p.165). The measured Y (luminance) values agreed with the results of a flicker photometry program run at 20 Hz on the two authors (both trichromats).

<u>Stimuli</u>

The stimuli were colored disks of 5 mm diameter (0.48 deg). The grey field (u'=0.19, v'=0.45) subtended 19 deg high by 25 deg wide. The field and disks had the same nominal luminance (Y=1.7 'L), so the stimuli would be identified by color, but as we did not adjust for individual differences or retinal location, isoluminance was not met exactly, and luminance cues are likely to have sharpened the edges of the disks. Disks were flashed briefly (for 167 or 213 ms) to nullify the effects of any eye movements. There were 64 possible disk locations, specified by an invisible polar grid of 16 lines which passed through the fixation point, and four centered rings. The rings were spaced 0.8 deg of visual angle apart, with radii of 1.6, 2.4, 3.2, and 4.0 deg of visual angle, as portrayed schematically in Fig. 1 (upper panel). All locations were extra-foveal. The centers of disks were at least 0.68 deg apart in the inner ring, and at least 0.92, 1.27, and 1.72 deg apart in successive rings, as indicated by short grey lines. A illustrative display is shown in the lower panel. All the disks served as distracters on target-absent trials. On target-present trials, the target replaced one randomly-chosen distractor, so the set size was the same on target-present as on target-absent trials.

Displays were grouped or ungrouped. In ungrouped displays, disk locations were chosen at random from the grid. In grouped displays, the location of the first disk in each group was chosen at random from one of the two middle rings, and the remaining elements in the group were placed in neighboring grid locations. There were always five groups. Thus with 40 distractors, the largest set size used, all eight neighbors were occupied, but with smaller set sizes, the displays were sparser.

INSERT FIGURES 1, 2 ABOUT HERE

Colors.

Five distractor colors were used in Experiments 1 and 2 (feature search). These colors were saturated (Figure 2, top left), unsaturated (top right), or linearly separable (lower left). Twelve colors were used in Experiment 3 (oddity) (Fig. 2, lower right panel). In oddity, distractors of both saturations were run together and saturation was analyzed post-hoc; linear separability was not tested. The color distance is the Euclidean distance in (u', v') space. For example, the pink target in feature search was at (0.31, 0.49) and the grey field was at (0.19, 0.45), giving the color distance $c_{field} = 0.13$. Distractor-field color distances (d_{field}) were in the ranges 0.11 to 0.13 for saturated distractors and 0.06 to 0.08 for desaturated ones. (Appendix 1, Table 1, gives all the color distances, and Table 2 gives all the stimulus coordinates.)

Even our most desaturated distractors had d_{field} values large enough to afford efficient search (Nagy & Sanchez, 1990), as confirmed in a pilot study with eight participants whose search RTs increased only for stimuli with $d_{field} < 0.05$. A further pilot study with 20 participants demonstrated that all the disks in each set were about equally reportable when briefly presented alone on the grey field (RTo in Appendix 1). Had some disks been easier than others, subjects might have processed them first, and a parallel model would not apply.

In both feature and oddity searches, distractors were either heterogeneous, with five different colors in each display, or homogeneous, with one color in each display. Colors in homogeneous displays were blocked, with colors fixed across an entire block of trials, or randomized, with colors varying from trial to trial. The target color was fixed in feature search and varied in oddity search. Using heterogeneous distractors in oddity seems counter-intuitive; it is explained in the methods for Experiment 3.

Procedure

The participant initiated each trial by a button press. A fixation symbol (+) then appeared at the center of the screen. The participant indicated readiness by a second button press. From 400 to 800 ms later, the fixation cross was replaced by a display for 200 ms. The time from onset of the display to onset of the button press (the RT) was recorded. Any error was flagged, and feedback (a beep) was provided on error trials. Responses were right or left mouse button presses, recorded with 2 ms precision using custom software.

Simple Reaction Time (RTo). Each participant was first run in 150 trials to collect simple RTs; the first 10 trials were discarded. Participants were told to depress their preferred mouse button as quickly as possible with their dominant hand, and to avoid errors (anticipations or failures to respond within 2.8 sec). A single target disk was presented on every trial, with no distractors. Simple RTs were collected from every participant in every Experiment (Appendix 2).

Search. Participants then received 30 trials of practice pressing the left and right mouse keys in response to the words Yes and No flashed on the screen, followed by a practice block of search trials, before search data (RTs and error rates) were collected. Search trials employed the same targets as simple RT trials, but distractors were also present. Trial type (target present or absent) and set size (m) were randomized and equally distributed within each display condition. Participants pressed the left mouse key to indicate target presence and the right key to indicate target absence. Feedback (a beep) was again provided on errors, which now included false alarms and misses as well as anticipations and failures to respond within 2.8 sec.

Analysis of the Search Data

Data were median hit RTs, median correct rejection RTs, d's, and criteria (c). These were normally averaged over participants and analyzed by t-tests or ANOVA, although in the saturation analysis in Experiment 3, RTs, hits and false alarm rates had to be pooled across participants before d' and Perf could be calculated reliably. Eccentricity (FOOTNOTE 1) had a marginal effect, and hemifield effects are small in color search (Pavlovskaya, Ring, Groswasser, Keren, & Hochstein, 2001), so we collapsed over both factors.

Experiments 1.1, 1.2 and 1.3

Experiment 1 studied the effects of distractor heterogeneity and grouping on feature search for a fixed (pink) target. According to the model, search should depend on set size, distractor heterogeneity, and grouping. Set size was varied in all Experiments. In Experiment 1.1, Distractor Heterogeneity, we compared search among homogeneous and heterogeneous distractors. In Experiment 1.2, Grouping, we compared search among grouped and ungrouped heterogeneous distractors. In Experiment 1.3, Similarity versus Proximity, we used a novel procedure to compare these two cues for grouping, the outcome being important for the model.

Experiment 1.1: Distractor Heterogeneity.

In Experiment 1.1, we measured feature search with ungrouped, saturated distractors, in 19 participants. Distractors were homogeneous randomized, in which colors were identical in each display but varied across trials, homogeneous blocked, in which colors were identical both within displays and across trials, or heterogeneous, in which distractors were multicolored within each display but identical across trials. The same five distractor colors were used in all three conditions. Distractor condition was blocked with the order counterbalanced across participants. Trial type (target present or absent) and set sizes (5, 10, or 15) were randomized and equally distributed within each distractor condition.

Results

INSERT FIGURE 3 HERE

Figure 3 shows median correct search RTs (circles and left-hand ordinates) for the three display conditions: homogeneous randomized (top), homogeneous blocked (middle), and heterogeneous (bottom). Hit RTs are shown by filled circles and correct rejection RTs by open circles in this and subsequent plots. Overall, search was considerably slower with heterogeneous than with homogeneous distractors [$F(2,18) = 53.8$, $p<.001$]. Slopes were flat (-1 to 2 ms/distractor) with homogeneous displays, but increased significantly (at 5 to 7 ms/distractor) with heterogeneous displays [$F(4,18) =16.9$, $p<.01$]. Sensitivity (open squares and right-hand ordinates) mirrored the pattern of the RTs; d' on heterogeneous trials was significantly lower than on homogeneous trials [$F(2,18) = 23.9$, $p<.001$]. The d' by set size functions were flat over set size for both types of homogeneous trials but declined significantly for the heterogeneous trials [$F(4,18) = 16.8$, $p<.001$].

These results are similar to those of other studies that have examined distractor heterogeneity (Farmer & Taylor, 1980; Duncan & Humphreys, 1989; Triesman, 1982); search was worse in heterogeneous trials than in homogeneous trials. This finding verifies the model assumption that noise is greater in heterogeneous displays than in homogeneous displays. Also, performance was about the same with blocked and randomized homogeneous distractors. It is true that Olds, Cowan, & Jolicoeur (1999) found a statistical difference between RTs in these two conditions in a similar feature search, but their difference was numerically small. Therefore, in subsequent experiments we only ran intermixed (randomized) trials.

Experiment 1.2 (Grouping)

Experiment 1.2 employed saturated, heterogeneous distractors, either grouped or ungrouped. Grouping heterogeneous distractors into columns counteracts their adverse effect on search (Farmer and Taylor, 1980). We eliminated any global (e.g., columnar) structure to maintain unpredictability of location (FOOTNOTE 3), but we anticipated that local grouping would still affect search, as it does for shapes (Enns & Kingstone, 1995). If a group of distractors is processed pre-attentively as a unit (Bravo & Blake, 1990; Triesman, 1982), then the effect on search should be mediated by a decrease in the effective set size. Set size was 5 on half the trials and 15 on the other half. Displays with 5 distractors were never grouped. Half the displays with 15 distractors were ungrouped and half were grouped into five clusters, each containing three identically-colored distractors.

Results

Data for 12 participants are shown in Figure 4. Data for set size 15 were from ungrouped displays (top panels) or from grouped displays (bottom panels). The data for set size 5 (always ungrouped) were arbitrarily split in half, with those for one half plotted in the top panels and the those for the other half in the bottom panels, to permit comparison with an equal number of trials in each set. The means of the median RTs are shown separately for hit and correct rejection trials (circles and left-hand ordinates). RTs increased with set size for ungrouped distractors by 45 ms for hit RTs, $t(11)=3.81$, $p <.01$, and by 55 ms for correct rejection RTs, $t(11)=6.44$, $p<.01$, giving rise to slopes of 4-5 ms/distractor. With 15 grouped distractors, however, there was no clear effect of set size on RT. A significant increase in hit RTs of 26 ms, $t(11)=3.50$, $p <.01$, was partly balanced by a non-significant drop in correct rejection RTs of 16 ms, $t(11)=1.68$, for slopes of -2 to +3 ms/distractor. Sensitivity (squares and right-hand ordinates) mirrored the pattern seen with mean RT, in that d' decreased with increasing set size for the ungrouped trials, from 3.6 to 3.0, $t(11)= 3.51$, $p<.01$, but d' did not change in the grouped trials (3.7; ns).

INSERT FIGURE 4 HERE

Experiment 1.2 verified that grouping identically-colored distractors counteracts the adverse effects of distractor heterogeneity. Search declined with set size from 5 to 15 ungrouped heterogeneous distractors, but was almost equally as good for 15 distractors placed in five groups as for five ungrouped distractors (unitization predicts exact equality). These results with our randomly-located displays are consistent with Farmer and Taylor's (1980) findings with row-column display. Hoverer, they leave open whether grouping by similarity or by proximity produces unitization. The next experiment tests this.

Experiment 1.3 (Similarity versus Proximity)

In this Experiment, we compared searches in which differently-colored distractors were spatially clustered (proximity) with searches in which identically-colored distractors were spatially clustered (similarity). If grouping depended only on proximity there should be no difference in performance between the two conditions.

Set size was again 5 or 15. Half the trials had displays of five ungrouped distractors. The remaining trials had 15 grouped distractors, presented in five clusters of three items each. Trial type and set size were randomized and equally distributed within each trial block. Grouping was a between-subjects variable. Each of 12 participants was assigned to be a member of a pair, one member being assigned to proximity and the other to similarity. Otherwise both members of a pair received identical spatial configurations and the same randomized trial sequence.

INSERT FIGURE 5 HERE

Results.

Data from one pair of participants were eliminated due to an unusually high error rate, leaving five useful pairs. Figure 5 shows RT and d' in proximity (top panel) and in similarity (bottom) in the usual format. Hit RTs (filled circles) increased with set size for both conditions, by 28 ms for proximity, $t(4)=3.17$, $p< .05$, and by 30 ms for similarity, $t(4)=2.14$, $p< 0.1$. Correct rejection RTs (open circles) increased with set size in proximity, by 70 ms, $t(4)=3.68$, $p< .05$, but dropped by 18 ms with set size in similarity, $t(4)=2.4$, $p<0.1$. Sensitivity (squares) was unaffected by set size (ns).

Experiment 1.2 showed that grouping aided search, since increasing set size increased RT by 50 ms in ungrouped displays but by only 5 ms in grouped displays. Experiment 1.3 showed that this grouping effect was not due to the spatial arrangement: the set size effects on d' and on RT in proximity grouping were the same as in ungrouped displays. Only in similarity grouping did the set-size effect diminish. Since grouping only aided search when identical items were clustered, and did so by almost entirely eliminating the set-size effect, we conclude that unitization occurred only when items shared both color and spatial proximity. This conclusion is important for identifying the effective set size (m) in the model equation.

Experiments 2.1 & 2.2

In Experiment 2, we examined how <u>saturation</u> influences feature search. Desaturated colors have less contrast than saturated ones, and therefore should produce less noise (σ^2_E) than saturated ones. Therefore, overall performance with desaturated distractors should be better than with saturated ones. This prediction is not just a trivial consequence of lower visibility, as the pilot study showed that the desaturated distractors were as visible (in terms of simple RT and accuracy of detection) as the saturated ones. A further prediction is that set size should have a greater effect with saturated than with desaturated distractors, reflecting the product $m\sigma^2_E$ in Equation 7.

Experiment 2.1 (Desaturated Homogeneous)

Experiment 2.1 served to check that feature search was efficient with homogeneous desaturated distractors (as it had been with saturated ones in Experiment 1.1). Five new desaturated distractor colors (Fig. 2, upper right panel) were chosen. Their mean distance from the grey field in (u', v') space was 0.08 units, half that of the saturated set (0.16 units) used before. As in Experiment 1.1, the target was pink, and distractors were identical in color on each trial but varied in color across trials. Trial type and set size (5, 11, or 15) were randomized and equally distributed over blocks.

Results

Feature searches were obtained from 12 participants. RTs (not plotted) increased slightly (from 435 ms at a set size of 5 to 451 ms at a set size of 15 on hit trials, and from 466 ms to 479 ms on correct rejections) for a mean slope of +1.5 ms/item. There was a small decline in d' with set size, from 3.24 to 3.07. Neither change was statistically significant. Thus, a known target will pop-out from homogeneous distractors, regardless of saturation.

Experiment 2.2 (Saturated and Desaturated heterogeneous distractors)

The goal of Experiment 2.2 was to examine the effects of color saturation using heterogeneous distractors. A wider range of set sizes (1-40) was used than before. Thus, Experiment 2.2 included the heterogeneous ungrouped saturated display condition from Experiment 1.1, in which feature search performance dropped markedly with set size. We wanted to replicate this effect with a wider range of set sizes, as it illustrates a large failure of 'pop-out'. Recall that these same saturated distractors hardly affected performance in Experiment 2.1 when displays were homogeneous. The model predicts this difference because evoked noise (σ^2_E) depends on distractor-distractor contrast, which is strong in heterogeneous displays but disappears in homogeneous ones.

Experiment 2.2 also included a grouping manipulation, to determine if the unitization found in the grouped displays in Experiment 1 would occur with the larger set sizes. Thus there were four types of displays: ungrouped saturated, ungrouped desaturated, grouped saturated, and grouped desaturated. Distractors were always heterogeneous. In the grouped displays, the number of groups was constant (5), and set size was increased by increasing the number of items within each group. Thus with set size 20 there were 4 items in each group. The design was complete except that small set sizes (1, 3, and 5) were always ungrouped. Different participants were run in three different set-size ranges, small (1, 3, 5), medium (11, 15, 20), and large (25, 30, 40), so that all participants could be run within an hour.

Results.

There were 35 participants in total. With saturated ungrouped distractors (Fig. 6, top panel), RTs increased with set size, from 483 to 599 ms for RThit (closed circles), and from 550 to 705 ms for RTcr (open circles), for a mean slope of 2 ms/distractor. For desaturated ungrouped distractors (Fig. 6, bottom panel), RThit increased marginally, from 490 to 508 ms and RTcr from 531 to 534 ms, for a mean of 0.3 ms/distractor. With grouped distractors (Fig. 7), RTs changed little with set size, for either the saturated (top) or desaturated distractors (bottom).

The effect of saturation on RT was larger for ungrouped than for grouped displays [$F(1,11) = 35.6$, $p<.01$, for the medium set sizes, and $F(1,7) = 11.2$, $p<.01$, for the large set sizes]. This difference increased with set size, and the resulting saturation by set-size interactions were significant for small [$F(2,13) = 26.8$, $p<.01$] and medium set sizes [$F(2,11)= 10.1$, $p<.01$], though not for the large set size [$F(2,7) = 0.2$, ns].

With ungrouped displays, d' decreased rapidly (from 3.25 to 1.72; $p<.01$) for saturated distractors (squares in the top panel of Fig. 6), but hardly changed for desaturated ones (bottom panel). There was no change in d' with set size for the grouped displays in Fig. 7 (ns). Overall, the d's for saturated distractors were lower than the d's for desaturated distractors [$F(1,13) = 14.8$, $p <.01$, for the small set sizes; $F(1,11) = 89.1$, $p<.0001$, for the middle set sizes; and [$F(1,7) = 48.6$, $p <.01$ for the large set sizes]. Saturation had a bigger effect at higher set sizes, and the resulting saturation by set-size interactions were significant: $F(2,13) = 10.1$, $p<.01$ for the small set sizes, $F(2,11) =4.64$, $p <.05$ for the middle set sizes, and $F(2,7) = 9.3$, $p<.05$ for the large set sizes.

INSERT FIGURES 6 AND 7 HERE

The model predicted that search should be better with grouped than ungrouped displays and should only deteriorate with set size in the ungrouped case. This was confirmed by the RTs and d's (Fig.s 6 and 7), and will be quantified with Perf (below). The results also show that desaturated distractors afforded fairly efficient search, whether they were grouped or not. Recall that all the distractors, saturated and desaturated, were equally visible, so the saturation by grouping interaction is specific to visual search. It arises in the model from the noise term, $m\sigma^2_E$, which is a product of effective set size (controlled by grouping) and distractor variance, controlled by saturation.

However, we were concerned about a possible confound between saturation and linear separability, as linear separability can affect search -- at least for near-threshold stimuli (D'Zmura, 1991). The target color in Experiment 2.2 was linearly separable in (u', v') space from three of the five distractor colors in the saturated set but from all five in the desaturated set. We therefore replaced two of the distractor colors in the saturated set to ensure that the target was linearly separable from all five distractors (Fig. 2, lower left panel). This lowered mean distractor saturation from 0.16 in previous Experiments to 0.13. We checked just the middle set sizes (11, 15, 20) where performance had dropped so much in Experiment 2.2. Results for a further 11 participants showed a similar pattern as in Experiment 2.2, and model fits (see General Discussion) indicated that the entire data set can be explained by the change in saturation alone. This conclusion agrees with Bauer et al.'s (1996) claim that linear separability does not affect efficient search for color stimuli outside the threshold regime (FOOTNOTE 4).

Summarizing the results of Experiments 1 and 2, feature search was fairly flat across set size (pop-out), except when distractors were saturated, heterogeneous, and ungrouped. The deleterious effect of set size on feature search, so vivid in this condition, was eliminated when these same distractors were grouped by similarity. We conclude that distractors grouped by similarity generate the same noise as single distractors due to pre-attentive unitization. Saturation affected search efficiency, as predicted, since distance in hue space from the grey field determines contrast. The model also predicted the well-known result that distractor heterogeneity reduces search efficiency (Farmer & Taylor, 1980; Nagy, 1999; but see Nagy and Thomas, 2003, for an exception at small set sizes.)

INSERT FIGURE 8 HERE

Symbols in Fig. 8 show Perfs for the feature searches in Experiments 1 and 2. The ordinates range from Perf = 0, when search is impossible (d'=0) or infinitely slow, to Perf=60, which corresponds to an accurate, rapid search (e.g. a d' of 3.2 achieved in 0.17 sec). Smooth lines show fits of the model (Equation 7) with c_{con} equaling the target-field contrast and parameters σ^2_E and σ^2_I best-fit as explained in the General Discussion.

<u>Statistical analysis.</u> Statistical analysis of the Perfs is redundant, given the earlier analysis in terms of RT and d', but the Perfs permit comparisons across experiments so we also analyzed them. Results from the ANOVAs confirmed the patterns which can be seen by eye in Fig. 10. First, *heterogeneity* had a significant effect. Perf for heterogeneous distractors in Experiments 1.1 and 1.2 (squares in Fig. 8) was significantly lower [$F(2,18) = 30.7$, $p<.0001$] than Perf for homogeneous distractors (triangles). Perfs for heterogeneous distractors more than halved with set size, from Perf = 46 when m was 5 to Perf = 22 when m was 15 [$F(4,18) = 11.1$, $p<.0001$]. Second, *grouping* was important. Perf for the ungrouped heterogeneous distractors (open squares in Fig. 8) declined significantly [$F(1,11)=12.9$, $p<.01$] from 51 to 30 and was overall lower [$F(1,11)=21.7$, $p<.001$] than Perf for the grouped ones (open circles; mean 56), which was flat over set size. Similarly, Perf was lower for ungrouped (squares) than for grouped (circles) displays, at all set size ranges in Experiment 2.2 [$F(1,11)=20.1$, $p<.001$ for (11,15,20); $F(1,7) = 6.0$, $p<.05$ for (25,30,40)]. Third, *saturation* had a significant effect for all three set size ranges; $F(1,13) = 33.8$, $P<.01$ for small, $F(1,11) = 61.3$, $p<.0001$ for middle, and [$F(1,7) = 135$, $p<.0001$ for large.

Fig. 8 shows that saturation interacted with homogeneity. With homogeneous distractors, Perf was about the same whether the distractors were desaturated Perf (filled triangles) or saturated (open triangles). However, with heterogeneous distractors, Perf in the desaturated case (filled circles and squares) was much better than in the saturated case (open circles and squares). Perf dropped markedly with set size for the saturated ungrouped distractors (open squares), in contrast to the steady level seen with desaturated ungrouped distractors (filled squares). The model explains these various effects in terms of the product of effective set size and distractor contrast (see General Discussion).

Experiment 3.1 (Oddity, Homogeneous)

The goal of Experiment 3 was to study oddity search for the same stimuli as used in the feature searches. Homogeneous distractors were used in Experiment 3.1 and heterogeneous ones in Experiment 3.2. With homogeneous distractors, data were flat in feature search and were predicted to be flat in oddity also.

Stimuli consisted of 13 colors (Fig. 2, lower right panel). Distractors were ungrouped. Distractors were homogeneous on each trial, but both target and distractor colors varied unpredictably from trial to trial. Each of the 13 colors appeared at least once in each trial block, both as a target and as a distractor. Distractor colors formed two broad classes; saturated, with a mean color distance from the grey field (d_{field}) of 0.16, and desaturated, with a mean d_{field} of 0.08. Trials with saturated and desaturated distractors were equally likely to occur and were intermixed in each trial block. The display parameters were otherwise identical to those in the feature searches.

Set sizes were 11, 15 and 20, randomized and equally distributed across trials in a block. The experimental procedure described under general method section was followed, but participants now had to report whether or not they detected an oddly colored target in the display. They were told to report 'Yes' if there was an odd color and 'No' if not. They were told that the distractors would always have the same color on each trial, but that the color would change at random from trial to trial.

Results

Data (not plotted) were analyzed for set size effects. There were none. Not surprisingly, the oddity searches were slower than the feature searches, and oddity d's (mean 3.3) were slightly lower than the corresponding feature search d's (mean 3.6). However, both types of search produced flat d' functions (within 0.2) and flat RT functions (within +-1 msec/item).

Our experimental task was similar to the 'detection' task of Bravo & Nakayama (1992), in which the observers just reported the presence or absence of an odd color among ungrouped homogeneous distractors. Our results were similar to theirs, and similar to those of Bacon & Egeth (1991) for $m > 8$, although neither earlier study reported errors in sufficient detail for us to compute d' or Perf. We conclude from these earlier studies and Experiment 3.1 that search for an odd target is unaffected by set size when the distractors are homogeneous.

Experiment 3.2 (Oddity, Heterogeneous)

In this experiment, we tested whether oddity search would be affected by set size and grouping when distractors were heterogeneous. We believe that this is a first report of an oddity experiment using maximally simple stimuli (disks) and heterogeneous distractors. The distractor colors differed from one another but were from the same broad region of color space, so that it was possible to tell if an odd color from an opposite part of color space (the target) was present or not. Heterogeneity had a large effect on feature search, eliminating pop-out; does it also affect oddity ?

Stimuli and Procedure

Seven sets of distractor colors were sampled from the13 colors used in Experiment 3.1. Each set contained five colors from a neighboring region of color space. Each distractor set was paired with two target colors chosen from a region of color space opposite to the distractor colors. Display parameters were otherwise unchanged. Distractors were ungrouped or grouped by similarity. Distractor grouping and saturation were crossed to create four types of trials, which were intermixed at random. Different participants were run in two different set size ranges, medium (11, 15, 20), and large (25, 30, 40), as before to permit the use of naïve participants run for only one hour each.

To define 'odd' in a heterogeneous display, we initially showed participants a color wheel and explained that the distractor colors would always occupy neighboring locations on the wheel, while the target (if present) would come from the opposite side. Thus, they had to respond "Yes" to a blue disk among red, orange, yellow disks, but "No" to an orange disk among a set of red, yellow, and yellow-green disks. None of the participants in our study (all of whom had normal color vision) reported having difficulty following this instruction. However, as a control, Experiment 3.1 was re-run with these slightly more complex instructions, and results (for 10 other participants) did not change.

INSERT FIGURES 9 & 10 ABOUT HERE

Results

Data from 15 participants were first analyzed by ANOVAs with d', c and Perf calculated individually from all the trials. Trials were subsequently separated by distractor saturation to permit comparison with the previous experiments, but hits and false alarms then had to be pooled across participants before d', criterion and Perf could be calculated reliably, precluding ANOVA.

RTs decreased with set size, at 2-3 ms per distractor, for both ungrouped displays (Fig. 9) and grouped (Fig. 10) displays, as confirmed by ANOVA [$F(2,13)=16.8$, $p<.001$) with data collapsed over saturation. RTs decreased with set size for saturated distractors (top panel in each figure) and desaturated ones (bottom). Saturation appears not to have interacted with set size.

Grouping increased d' significantly [$F(1,13)=74.9$, $p<.001$, with data collapsed over saturation]. Finally, RTs were longer with ungrouped displays than with grouped displays.

There was a slight increase in d' with set size [$F(2,13) = 8.56$, $p<.05$], at about the same rate for ungrouped as for grouped distractors [$F(2,13)<1$], and at similar rates for both saturated and unsaturated distractors. In sum, these results show that even with simple heterogeneously colored disks, performance improves with set size in oddity searches. Moreover, grouping improved oddity search, as it had improved feature search, whereas the effects of saturation were fairly minimal.

Perf in Oddity searches

Perf in oddity searches is shown by symbols in Fig. 11; the lines will be explained later. In Experiment 3.1, Perf for homogeneous distractors (shown by triangles) was high and did not seem to change with set size. In this respect, oddity is like feature search (Fig. 8). In Experiment 3.2, Perf for the heterogeneous distractors increased with set size for both grouped distractors (circles) and ungrouped distractors (squares) [$F(2,13)= 7.60$, $p<.01$]. The improvement with set size did not interact with grouping ($p > 0.1$). In both these respects performance in oddity differs from that in feature search.

INSERT FIGURE 11 ABOUT HERE

Four control experiments

Comparing Figs 8 and 11, the major finding is of very divergent set-size effects in feature and oddity searches. We were concerned that the use of different set-size ranges might have contaminated this, our major result. In Experiment 2.2 (feature search), the drop in Perf over set size might possibly have arisen because participants in the small (1-5), middle (10-20), and large (25-40) ranges adopted successively worse strategies. We therefore ran a replication with six new participants, in which set sizes were 5, 15, and 40. Feature searches with ungrouped distractors were comparable to those already reported; Perf dropped rapidly with saturated distractors, from 20 to 4.3 to 1.1 as set size increased from 5 to 15 to 40, but Perf barely dropped with desaturated ones, from 47 to 42 to 39. Grouping had its usual effect; Perf was 3.5 times higher when 15 or 40 saturated distractors were grouped by similarity than when they were ungrouped.

We were similarly concerned that the oddity data might be contaminated by set-size range. Set-size ranges were bunched into medium (11, 15, 20) and large (25, 30, 40) with different groups of participants. We therefore repeated Experiment 3.2 with set sizes of 11, 20, and 40 to span the full range. Results for six new participants showed that Perf once again increased with set size, by 9 Perf units on average, in all four conditions (saturation crossed with grouping). We conclude that the major finding of the study, the divergent set-size effects in feature search and oddity, is not due to range effects.

We were also concerned that the difference between the feature search and oddity data might have been an artefact of the increase in the number of colors used in oddity. We therefore re-ran both the feature and the oddity searches with four practiced participants, using exactly the same distractors (the full set; Fig. 2, lower right panel) in both searches. The same divergent pattern of results occurred, so we concluded that the number of colors did not have an influence.

An additional finding is that saturation can be quite critical, especially in feature search. This is explained in the model in terms of Euclidean distance. If this is indeed the sole factor, then other salient properties of color space should have no effect. We tested one such property, that of color temperature. Eight naïve observers rated the color warmth of all 13 color stimuli (Fig. 2, lower right panel), presented briefly and in the same locations as in the other experiments. The mean ratings correlated remarkably well (r = 0.98) with the Judd corrected color temperature (Wysecki & Stiles, 1982), illustrating external validity. However, color temperature had a negligible effect on Perf, either in feature search (r=0.07, ns) in Experiment 2.2, or in oddity (r = -0.23, t (11)=0.93, ns) in Experiment 3.2.

Model fits to oddity

The solid lines in Fig. 11 show fits of the model to the ungrouped and homogeneous Perfs. The model fit these data quite well, accounting for 91% of the variance. However the model fit the grouped data poorly; these data are connected by broken dotted lines just to aid the eye. The model predicts that Perf should be flat when the distractors are grouped, due to unitization, but Perf increased with a slope of 0.21 per distractor. Moreover, the model correctly predicts that Perf should be better in grouping, but the mechanism (noise reduction due to unitization) under-predicted the magnitude of this effect. It appears that grouping also increases the effective contrast of the target. As we had no principled reason for this, we decided not to add post-hoc terms to the model, which, though descriptive, is not arbitrary. Hence the broken dotted lines are not theoretical.

The model fits to oddity are based on exactly the same noise parameters as for feature search. However, Equation 5, which specifies effective target contrast, is modified, as will now be explained. Detecting an oddly-colored target which is not known in advance requires that all the colors in the display be processed to determine if one is odd. Thus the filter, which in feature search could be concentrated on the known target, must encompass all of color space to avoid filtering out potential targets. With such a broadened filter, the target hue will contrast with the distractor hues as well as with the grey field, so target-distractor contrast will matter (see also Pashler, 1987). We therefore assume that the effective target contrast, c_{con}, is which was c_{field} (the target-field contrast) in feature search, is

$$c_{con} = A(c_{field} + m.D), \qquad \text{in oddity.} \qquad (5a)$$

Here, the total target-distractor contrast is mD, as the mean target-distractor contrast, D, was made independent of m, the set size, in the oddity Experiments. Factor $A < 1$ is the attention paid to each location in color space. In feature search, $A=1$ for the target color (and 0 elsewhere). Thus, the target in oddity will receive less attention than in feature search, and Perf will be lower overall. However, the effective target contrast will increase with set size, because of the term m.D. In this respect A acts as a gain control, and indeed, Blaser, Sperling, and Lu (1999), using apparent motion as a probe, showed that attending to a color increases gain.

In Equation 5a, the distractors contribute effective contrast to the target. We now discuss why the simple sum (m.D) is plausible. In our displays the targets and distractors were always 0.5 deg apart or more. Therefore the term D reflects not local but *long-distance* color contrast. In this respect D is akin to 'feature contrast' among orientations (Nothdurft, 1993). However, orientation contrast falls off with distance. In contrast, long-distance color contrast, when measured with briefly-presented stimuli like ours, is almost independent of distance over the range from 1.8 to 6.7 deg (Walraven, 1973). This range encompasses almost all our display elements, so all the distractors are weighted equally in the mD term, independently of distance. (We also modeled the data using a best-fit exponential weight on distance, but fits were not noticeably improved over the simpler Equation 5a.) Long-distance color contrast is often ignored but it is not negligible, for example, it altered the red/green ratio needed to match yellow by 40%

in the study of Walraven (1973), who used uniform inducers. Moreover, distant checkerboard inducers have almost twice the effect of uniform inducers (Shevell & Wei, 1998), the checks being more like our separated stimuli. However, long-distance color contrast is cortically mediated (Shevell & Wei, 2000), and as cortical color-sensitive cells with large receptive fields have expansive and compressive, as well as linear, contrast-response functions (Sclar, Maunsell, & Lennie, 1990), the summation in Equation 5a is likely to be approximate.

Equation 5a merely specifies that the effective contrast of the target in oddity grows in proportion to set size, which in turn amplifies the signal. It does not describe how the decision is made. Various possibilities are open. The simplest is that the observer says Yes if the sum of the contrasts exceeds a criterion. For $d'>0$, the target must, on average, gain more effective contrast from the distractors than any of the individual distractors, so that the sum is more likely to exceed the criterion on a target-present trial than on a target absent trial. Since the distractors are clustered in color space, distract-distractor contrast is much lower than distractor-target contrast in oddity, so this is possible. An alternative possibility (the Maximum model) is described in Appendix 3.

General Discussion

We first summarize the effects of (1) target certainty (feature versus oddity), (2) distractor heterogeneity, (3) spatial grouping, and (4) saturation, and then explain the model fits.

(1). Oddity was overall worse than feature search, but while performance decreased with set size in feature searches through heterogeneous displays, it increased in oddity search through the same displays. (Indeed, at the largest set size the Perf scores converged -- prior knowledge of the target became useless !) This difference demonstrates a pure task effect, since the same stimuli and conditions were used in both types of search. This is our main new result. It is explained in our phenomonological model because the subject must attend to the distractors in order to determine which is odd, and adding distractors increases effective target contrast. Appendix 3 explains how this effect might occur in a Maximum model.

(2). When distractors were homogeneous, the target popped out, giving fast RTs and high d's. In SDT models, pop-out occurs when distractors produce little evoked noise and the target is distinct, so the signal/noise ratio is high. Since the same distractors were used in both heterogeneous and homogeneous conditions, it might be thought that evoked noise should be the same in both conditions. However, the model postulates that evoked noise should be lower in a homogeneous than in a heterogeneous display since there is only one activate distractor location in color space when the display contains only one color. But how could this assumption of one active location be realized ? We suggest that the observer's color space acts like a feature map (Triesman, 1986, 1991) in which information about features accumulates over time. Noise will accumulate in such a map, more so from heterogeneously colored distractors than from homogeneously-colored ones (Shore & Klein, 2000.) Rosenholtz (2001a) rejected a similar SDT model in the orientation domain because increasing distractor heterogeneity reduced performance even when concomittent increases in distractor-target distances should have improved performance. Her results do not reject our model, however, in which evoked noise increases with the number of active locations in the feature map. An alternative explanation for pop-out in the homogeneous case is that the target, being salient, attracts attention exogenously and is thereby sped up sufficiently to be processed before the distractors. However, brief isoluminant stimuli like ours do not attract attention exogenously unless specifically cued (Lambert, Wells & Kean, 2003), so we retain the parallel processing assumption.

(3). Performance was better when the heterogeneous distractors were grouped by similarity than when they were ungrouped. The model accounts for this by assuming that similarity grouping unitizes the distractors and so reduces the effective set size, m, to the number of groups, thereby lowering the total evoked noise, $m\sigma^2_E$. (Grouping would not have reduced the other term, σ^2_E, as the colors were not changed.) What then explains the reduction in effective set size? Mere proximity had no effect (Experiment 1.3); only clustering by similarity reduced the noise, and it did so as if each group was unitized (Bravo & Blake, 1990; Duncan, 1995). This implies a global, preattentive grouping process which can spread widely if uninterrupted by other stimuli (Grossberg, Mingolla, and Ross, 1994), and which is sensitive to hue. Local spatial integration cannot explain grouping effects in our displays because briefly-presented stimuli do not interact laterally if they are more than about 0.5 deg apart (Gardner, 1973), and the elements in our groups were further apart than this. Indeed, in yet another pilot experiment, our grouped displays were reported to have as many elements as the ungrouped ones -- grouping did not smear the stimuli together.

(4). The <u>saturation</u> of the distractor colors also affected performance. In the model, evoked noise depends on the mean distance between the distractors in color space (controlling σ^2_E), which was smaller for desaturated than for saturated distractors. So performance was expected to be better with desaturated than with saturated displays, as was shown in Experiment 2.1 for feature search. This effect was reversed in oddity, however. In the model, increasing saturation increases σ^2_E in both types of search. Therefore, to obtain the reversal in oddity, saturation must increase the signal even more. This effect is accommodated in the model since increasing saturation increases D in the AmD term.

We note here that our model predicts the typical monotonic decline of Perf with set-size (m) in feature search. Sagi and Julesz (1987), however, reported a U-shaped function in a feature search for a line of a known orientation presented with a dense array of distractors, in contrast to the monotonic decline found with a sparse array. Their results could be explained by our model if their observers were able to filter out the distractors in the sparse array but not in the dense array, to which Equation 5a would apply.

Model Parameters and fits to Perf

In developing the model, we ignored any residual effects of eccentricity (FOOTNOTE 1), any trial-to-trial dependencies (e.g. Bravo and Nakayama, 1992), and any individual differences. With these simplifications, we were able to fit Equation 7. Inputs to the model were the Perfs for each condition, the effective set sizes (1 in homogeneous displays, m in ungrouped displays, and 5 when grouped), and the (u', v') hue co-ordinates (FOOTNOTE 5). The co-ordinates were used to calculate the target-field contrast, c_{field}, the mean target-distractor contrast D, and the mean distractor-field contrast d_{field} (respectively D and DD in Appendix 1, table 1). Display duration varied for trivial reasons from 180 to 210 ms across experiments, and fits were slightly but systematically improved when the 210 ms display was assumed to provide 210/180 = 1.17 times more signal than the 180 ms one.

The free parameters in the fits to the feature searches were σ^2_I and σ^2_E. The distractor-independent noise, σ^2_I, was best-fit to 0.00041. The distractor-evoked noise, σ^2_E, was best-fit to $0.052(d_{field}^4)$. The fourth power on d_{field}, the mean distractor-grey distance, was necessary because saturated distractors had so much more effect on feature search than desaturated ones, even though they were only twice as far away in (u', v') space. Since saturation is a power function of the Euclidean distance from white, with powers varying from 1.5 to 5 (Indow & Stevens, 1966), a fourth power is not unreasonable; it fitted much better than the second power and marginally better than the third power. We were also able to fit the Perfs obtained with linearly separable distractors (Fig. 2; the data are not shown), using the same parameters, in support of the Euclidean distance assumption in the model. The feature searches were fit well (see Fig. 8), with the model accounting for 94% of the variance as defined by {100 - SUM(Perf - PredictedPerf)^2 / SUM (Perf - MeanPerf)^2}.

In <u>oddity</u>, the noise terms σ^2_I and σ^2_E were clamped at the values fit to the feature searches, on the model's assumption that the same noise was generated in oddity. The attention parameter A was then best-fit to 0.39. The model fit the ungrouped oddity data well but missed the grouped data, as already explained; recall that grouped distractors improved Perf more than predicted, as if grouping the distractors somehow increased the extent to which they contrasted with the odd target. This effect remains unexplained. Thus, while the fit in Fig. 13 to the ungrouped (squares) and the homogeneous (triangles) distractors represents a test of the model, the fit to the grouped data (circles) does not. The three-parameter model, fit to all but the grouped oddity data, accounted for 91% of the variance in Perf.

Assuming parallel processing, the contrast between oddity and feature search suggests that attention controls a filter that determines the effective signal contrast in feature space. In this respect, the model emphasizes feature-based selection over object-based selection (e.g., Mounts & Melara, 1999). This role differs from that envisaged in serial models in which attention controls a series of decisions made on each trial about which item or dimension to process next.

The model in Equation 7 merely predicts the relative sizes of signal and noise when the number of distractors, and their properties, are varied. In this respect it is a phenomenological model, one which does not attempt to capture the underlying decision processes (Appendix 3). The model, though not unique (FOOTNOTE 6), is simple. Validation of the model, and especially of the utility of Perf, would require analysis of numerous experiments in which both d' and RT are reported for every set size. Unfortunately, few reports include both RT and accuracy in sufficient detail, and the few that do span different domains (color, orientation, texture).

In summary, performance was overall lower for unknown than known targets. Performance decreased as a function of set size when the target was known, but increased as a function of set size when the target was unknown, a novel finding for stimuli as simple as our colored disks. We varied several distractor attributes: heterogeneity, grouping, saturation, and linear separability. Homogeneous distractors of all types afforded efficient search for both known and unknown targets. With heterogeneous distractors, randomly positioned (ungrouped) saturated distractors greatly reduced efficiency in feature search, although grouped ones did not. A simple SDT-based model describes all these findings, except for performance in the grouped oddity case, with three well-motivated parameters.

FOOTNOTE 1. We used an annular, extra-foveal, display region (1.5 - 4.5 deg) in an attempt to minimize the effects of retinal eccentricity (Carrasco, Evert, Chang, & Katz, 1995). Indeed our eccentricity effects on RT and d' were small (Santhi, 2000). However, there are variations in color thresholds in this region (Stromeyer, Lee, & Eskew, 1992). Moreover, visual search speeds up beyond this region, e.g. search for a fixed-size Gabor patch is faster at 9 deg than at 4 deg (Carrasco, McElree, Denisova, & Giodrano, 2003). A complete search model should therefore include eccentricity.

FOOTNOTE 2. The Euclidean color distance in (u', v') space between stimuli A and B is $\{(u'_A - u'_B)^2 + (v'_A - v'_B)^2\}^{0.5}$. Carter and Carter (1981) found that three CIE color difference formulae were equally good predictors of search efficiency. Color distances in one of them, CIELUV, are proportional to those in (u', v') space at equiluminance, justifying our use of (u', v').

FOOTNOTE 3. Baldassi and Burr (2000) found that potential, but empty, distractor locations could also add noise in visual search. We anticipated that empty locations would not add noise in our experiments, as our distractor locations, unlike theirs, were not predictable. Moreover, Cepeda, Cave, Bichot & Kim (1998) found that a visual search for colored digits slowed RTs to a small probe flash at the locations of the distractors, but not between those locations, indicating that unoccupied distractor locations generated little noise in their displays. However, to see if there was a role for the (M-m) absent distractors, we let the total noise equal $[m\sigma^2_E + \sigma^2_I + (M-m).\sigma^2_U]$, where M is the maximum set size on a block of trials. Best fits of Equation 7 to the feature search data drove σ^2_U to zero.

FOOTNOTE 4. To account for effects of linear separability near threshold is tricky, because thresholds cannot be expressed accurately by linear mixtures of cone signals (Wyszecki and Stiles, 1982, p. 311). It is conceivable that a non-linear hue space such as D'Zmura's (1991) could account for the published data without invoking an effect of linear separability per se.

FOOTNOTE 5. Values of d' were capped at 3.4 to avoid outliers. Higher d's only occurred in feature search with homogeneous distractors, so capping does not affect the main body of the results. Distractor-distractor contrast, DD (Appendix 1, table 1) was only used as a flag in the model fits; that is, σ^2_E was set to 0 when the distractors were homogeneous (DD=0).

FOOTNOTE 6. Rapid serial scanning (at 10-12 ms/scan) can achieve similar effects (Sperling, 1963; Liss & Reeves, 1983), and can explain even shallower slopes if more than one item is picked up on each scan. Indeed, for many classes of parallel models, there exist equivalent serial (Townsend, 1971, 1990) or hybrid models (Harris, Shaw, and Bates, 1979).

References

Bacon, F, W., & Egeth, E. H.. (1991). Local processes in preattentive feature detection. Journal of Experimental Psychology: Human Perception and Performance, 17, 77-90.

Baldassi, S., & Burr, D. (2000). Feature-based integration of orientation signals in visual search. Vision Research, 40, 1293-1304.

Bauer, B., Jolicoeur, P., & Cowan, B.W. (1996a). Visual search for color targets that are or are not linearly separable from distractors. Vision Research, 36, 1439-1465.

Ben-Av, M. B., Sagi, D., & Braun, J. (1992). Visual attention and perceptual grouping. Perception and Psychophysics 52, 277-294.

Blaser, E., Sperling, G., & Lu, Z.L. (1999). Measuring the amplification of attention. Proceedings of the National Academy of Sciences USA, 96, 11681-11686.

Bravo, J. M., & Nakayama, K. (1992). The role of attention in different visual tasks. Perception and Psychophysics, 51, 465-572.

Bravo, M., & Blake, R. (1990). Preattentive vision and perceptual groups. Perception, 19, 515-522.

Bundesen, C., & Pedersen, L. F. (1983). Color segregation and visual search. Perception and Psychophysics, 33, 487-493.

Carrasco, M., Evert, D. L., Chang, I., & Katz, S. M. (1995). The eccentricity effect: Target eccentricity affects performance on conjunction searches. Perception and Psychophysics, 57, 1241-1261.

Carrasco, M. & McElree, B. (2001). Covert attention accelerates the rate of visual information processing. Proceedings of the National Academy of Sciences, 98, 5363-5367.

Perception and Psychophysics, 57, 1241-1261.

Carasco, M., McElree, B., Denisova, K., & Giodrano, A.-M. (2003). Speed of visual processing increases with eccentricity. Nature Neuroscience, in press (on-line).

Carter, R. C. (1982). Visual search with color. Journal of Experimental Psychology: Human Perception and Performance, 8, 127-136.

Carter, E, C., & Carter, R, C., (1981). Color and Conspicuousness. Journal of the Optical Society of America, 71, 723-729.

Cave, K. R., & Wolfe, J. M. (1990). Modeling the role of parallel processing in visual search. Cognitive Psychology, 22, 225-271.

Cepeda, N. J., Cave, K. R., Bichot, N. P., & Kim, M. S. (1998). Spatial selection via feature-driven inhibition of distractor locations. Perception and Psychophysics, 60, 727-746.

Dosher, B. A., & Lu, Z. L. (2000). Mechanisms of perceptual attention in precuing of location. Vision-Research, 40, 1269-1292.

Duncan, J. (1984). Selective attention and the organization of visual information. Journal of Experimental Psychology: General, 113, 501-517

Duncan, J. (1995). Target and non-target grouping in visual search. Perception and Psychophysics, 57, 117-120.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. Psychological Review, 96, 433-458.

D'Zmura, M. (1991). Color in visual search. Vision Research, 31, 951-966.

D'Zmura, M., Lennie, P., & Tiana, C. (1997) Color search and visual field segregation. Perception and Psychophysics, 59, 381-388.

Eckstein, M. P., Thomas, J. P., Palmer, J., Shimozaki, S., & Steven, 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 and Psychophysics, 62, 425-451.

Enns J.T., & Kingstone, A. (1995). Access to global and local properties in visual search for compound stimuli. Psychological Science, 6, 283-291.

Farmer, E. W., & Taylor, R. M. (1980). Visual search through color displays: Effects of target -field similarity and field uniformity. Perception and Psychophysics, 27, 267-272.

Gardner, G. T. (1973). Evidence for independent parallel channels in tachistoscopic perception. Cognitive Psychology, 4, 130-155.

Geisler, W. S., & Chou, K.-L. (1995). Separation of low-level and high-level factors in complex tasks: Visual Search. Psychological Review, 102(2), 356-378.

Green, D.M., & Swets, J.A., (1966). Signal detection theory and psychophysics. New York: Kreiger.

Green, D. M. & Luce, R. D. (1973). Speed-accuracy trade off in auditory detection. In: Attention and Performance IV, Ed. S. Kornblum. New York: Academic Press.

Green, B. F. & Anderson, L. K. (1956). Color coding in a visual search task. Journal of Experimental Psychology, 51, 19-24.

Grossberg, S., Mingolla, E., & Ross, W. D. (1994). A neural theory of attentive visual search: Interactions of boundary, surface, spatial and object representations. Psychological Review, 101, 470-489.

Harris, J. R., Shaw, M. L., & Bates, M. (1979). Visual search in multielement arrays with and without gaps. Perception and Psychophysics, 26, 69-84.

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.

Indow, T., & Stevens, S. S. (1966). Scaling of saturation and hue. Perception and Psychophysics, 1, 253-271.

Lambert, A., Wells, I., & Kean, M. (2003). Do isoluminant color changes capture attention ? Perception and Psychophysics, 65, 495-507.

Liss, P., & Reeves, A. (1983). Interruption of dot processing by a backward mask. Perception, 12, 513-529.

McCabe, A.., Caelli T., West G., and Reeves A. (2000). Theory of spatiochromatic image encoding and feature extraction. Journal of the Optical Society of America A, 17, 1744-1754.

Mollon, J. D., & Cavonius, C. R. (1986). The discriminability of colours on c.r.t. displays. Journal of the Institute of Electronics and Radio Engineering, 56, 107-110.

Mounts, J. R. W., & Melara, R. D. (1999). Attentional selection of objects or features; Evidence from a modified search task. Perception and Psychophysics 61, 322-341.

Nagy, A. L., & Sanchez, R. R. (1990). Critical color differences determined with a visual search task. Journal of the Optical Society of America, A, 7, 1209-1217.

Nagy, A. L. (1999). Interactions between achromatic and chromatic mechanisms in visual search. Vision Research 39, 3253-3326.

Nagy, A.L., & Thomas, G. (2003) Distractor heterogeneity, attention, and color in visual search. Vision Research 43, 1541-1552.

Nothdurft, H.-C. (1993). The role of features in preattentive vision: Comparison of orientation, motion and color cues. Vision Research 33, 1937-1958.

Olds, E. S., , W. B., & Jolicoeur, P. (1999a). Stimulus-determined discrimination mechanisms for color search. Perception and Psychophysics, 61, 1038-1045.

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.

Palmer, J., Verghese, P., & Pavel, M. (2000). The psychophysics of visual search. Vision Research, 40, 1227-1268.

Palmer, J., Ames, C. T., Lindsey, D. T. (1993). Measuring the effect of attention on simple visual search. Journal of Experimental-Psychology: Human Perception and Performance, 19, 108-130.

Pashler, H. (1987). Target-distractor discriminability in visual search. Perception and Psychophysics, 41, 285-302.

Pavlovskaya, M., Ring, H., Groswasser, Z., Keren, O., & Hochstein, S. (2001). Visual search in peripheral vision: learning effects and set-size dependence. Spatial Vision, 14, 151-173.

Ratcliff, R., & Rouder, J. N., (1998). Modeling response times for two-choice decisions. Psychological Science, 9, 347-356.

Reeves A., & Sperling, G. (1986). Attention gating in short-term visual memory. Psychological Review, 93, 180-206.

Reeves A., Wu S., & Schirillo J. (1997). The effect of photon noise on the detection of white flashes. Vision Research, 38, 2639-2647.

Rosenholtz, R. (2001a). Visual search for orientation among heterogeneous distractors: Experimental results and implications for signal detection theory models. Journal of

Experimental Psychology: Human Perception and Performance, 27, 985-999.

Rosenholtz, R. (2001b). Search asymmetries? What search asymmetries ? Perception and Psychophysics 63, 476-489.

Sagi, D., & Julesz, B. (1987). Short-range limitation on detection of feature differences. Spatial Vision, 2, 39-49.

Santhi, N. (2000). The role of distractor coherence and target certainty in feature search: a signal detection approach. Unpublished dissertation, Boston MA: Northeastern University.

Sclar, G., Maunsell, J.H.R., & Lennie, P. (1990). Coding of image contrast in central visual pathways of the macaque monkey. Vision Research, 30, 1-10.

Shevell, S. H., & Wei, J. (1998). Chromatic induction: border contrast or adaptation to surrounding light ? Vision Research, 38, 1561-1566.

Shevell, S. H., & Wei, J. (2000). A central mechanism of chromatic contrast. Vision Research, 40, 3173-3180.

Shore, D. I., & Klein R. M. (2000). On the manifestations of memory in visual search. Spatial Vision, 14, 59-75.

Smallman, H. S., & Boynton, R. M. (1990). Segregation of basic color in an information display. Journal of the Optical Society of America A, 7, 1985-1994.

Sperling, G. (1960). The information available in brief visual presentations. Psychological Monographs: 74 (11, Whole No. 498), 1-29.

Sperling, G. (1963). A model for visual memory tasks. Human Factors, 5, 19-31.

Stromeyer, C.F., Lee, J., & Eskew, R.T. (1992) Peripheral chromatic sensitivity for flashes: A post-receptoral red-green asymmetry. Vision Research, 32, 1865-1873.

Swensson, R. G., & Thomas R E. (1974). Fixed and optional stopping models for two-choice discrimination times. Journal of Mathematical Psychology, 15, 282-291.

Townsend, J. T. (1971). A note on the identification of parallel and serial processes. Perception and Psychophysics, 10, 161-163.

Townsend, J. T. (1990). Serial and parallel processing: Sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished. Psychological Science, 1, 46-54.

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.

Treisman, A. (1986). Features and objects in visual processing. Scientific American, 255, 114B-125.

Treisman, A. (1991). Search, similarity, and integration of features between and within dimensions. Journal of Experimental Psychology: Human Perception and Performance, 17, 652-676.

Walraven, J. (1973). Spatial characteristics of chromatic induction; the segregation of lateral effects from straylight artefacts. Vision Research 13, 1739-1753.

Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. Psychonomic Bulletin and Review, 1, 202-238.

Wyszecki G., Stiles, W. S. (1982). Color Science. New York, Wiley.

Appendix 1: Color Distances and simple RTs

Table 1: Color Distances in (u’, v’) space

Experiment	Condition	d_{field}	c_{field}	D	DD
1, 2	Saturated	0.16	0.13	0.15	0.22
2	DeSaturated	0.08	0.13	0.13	0.14
3	Saturated	0.17	0.12	0.25	0.14
3	DeSaturated	0.09	0.13	0.15	0.09

Table 2: Color Coordinates (u’, v’) and RT0.

experiment #	color region	u'	v'	saturation	RT0
1.1-1.3	Blue 1	0.19	0.33	0.12	252
1.1-1.3	Purple 1	0.32	0.38	0.17	244
1.1-1.3	Red 1	0.45	0.52	0.27	224
1.1-1.3	Green 1	0.13	0.56	0.12	261
1.1-1.3	Yellow 1	0.23	0.55	0.11	245
1.1-1.3	Pink 1	0.31	0.49	0.13	231
2.1-2.3	Blue 2	0.27	0.53	0.12	283
2.1-2.3	Green 2	0.23	0.41	0.06	288
2.1-2.3	Red 1	0.16	0.44	0.03	265
2.1-2.3	Purple 2	0.14	0.51	0.09	266
2.1-2.3	Yellow 2	0.21	0.54	0.09	264
2.1-2.3	Pink 1	0.31	0.49	0.13	256
2.2 only	Blue 1	0.19	0.33	0.12	247
2.2 only	Purple 2	0.27	0.31	0.16	283
2.2 only	Yellow 2	0.23	0.55	0.11	235
2.2 only	Green 1	0.13	0.56	0.12	271
2.2 only	Green 3	0.15	0.56	0.11	270
2.2 only	Pink 1	0.31	0.49	0.13	256
3.1-3.2	Blue 1	0.19	0.33	0.12	315
3.1-3.2	Purple 2	0.27	0.31	0.16	283
3.1-3.2	Purple 3	0.35	0.38	0.17	280
3.1-3.2	Pink 3	0.31	0.43	0.13	281
3.1-3.2	Purple 4	0.23	0.41	0.06	285
3.1-3.2	Blue 4	0.16	0.44	0.03	288
3.1-3.2	Green 4	0.13	0.51	0.09	274
3.1-3.2	Green 1	0.13	0.56	0.12	271
3.1-3.2	Yellow 2	0.21	0.54	0.09	277
3.1-3.2	Yellow 3	0.23	0.53	0.09	301
3.1-3.2	Orange 1	0.27	0.53	0.12	298
3.1-3.2	Pink 1	0.31	0.49	0.13	269
3.1-3.2	Red 1	0.45	0.52	0.27	271

Appendix 2: Analysis of simple RT.

The use of Perf requires invariance of RTo across color and eccentricity. We analyzed the RTo's to determine whether these invariances held. We also report mean RTo across Experiment.

Color. In Experiments 1, 2, and 3, the standard deviations of the RTo's over colors were 12 to 15 ms. Although these are not high, they mask some larger variations; for example, in Experiment 1, the difference between the slowest color, green, and the fastest, red, was 37 ms (Table 2). Ideally the most variant colors would have been shifted in color space to bring their RTo's closer to the others, but the color distances had to be kept appropriate.

Eccentricity. In measuring simple RTs, targets were presented in the inner ring ('near'), at 1.6 deg, and the outer ring ('far'), at 4 deg, randomly left and right of fixation. There was no eccentricity effect for pink, the target color in feature search; near and far RTo's were always within 5 ms. Averaging over the distractors, the far RTo's were slower than the near RTo's by 9, 12, and 8 ms in Experiments 1, 2, and 3 respectively. When we plotted the difference between near and far RTo's against the dominant wavelength for each color, averaged across experiments, we found the 'green weak' phenomenon (Stromeyer, Lee, & Eskew, 1992): that is, greens and yellows were slower in the 'far' locations by 15 ms on average, whereas blues, pinks and reds were slower by only 6 ms. These small interactions were ignored in data fitting and in the model.

Experiments. When the stimulus duration was 207 ms, mean RTo's were 230 ms (Experiment 1.1) and 234 ms (Experiment 2.1). In all other Experiments the stimulus duration was 170 ms, and mean RTo's were slightly longer: 251 and 252 ms in Experiments 1.2 and 1.3, 263 ms in Experiment 2.2, and 284 and 270 ms in Experiments 3.1 and 3.2. The slowing from Experiment 1 to 3 may reflect the addition of new colors. In calculating T, the appropriate RTo was used for each experiment, so these differences dropped out.

Appendix 3: the Maximum model.

The model (Equation 5) is phenomonalistic; i.e., it predicts the relative amounts of signal and noise but does not providing a probabilistic account of the underlying decision-making process. Possibly color contrasts are literally *summed* over space as in the model equation, to create a single one-dimensional decision variable. Alternatively, our data and model might be approximated by a form of *maximum* rule in which the observer is presumed to decide Yes or No by comparing the maximum of all m stimuli to a criterion (e.g., Palmer, Verghese & Pavel, 2000; Palmer, Ames, & Lindsey, 1993). In a *tour de force* of fitting feature, conjunction, and disjunction searches with various multi-dimensional SDT models, Eckstein, Thomas, Palmer, Shimozaki, & Steven (2000) found that the max-linear and max-min models did well. We tried fitting various versions of a Maximum SDT model to both feature and oddity searches, but as these models only predict d', we ignored RTs.

Feature searches. We used Monte-Carlo methods to estimate the distribution of the maximum of m noise sources. On each trial, m random variables (rv's) were created. The rv's were Gaussians centered on 0 and truncated at +-3 sigma. These rv's represented m distractors on the 50% of noise-alone trials. On the 50% of signal trials, one rv was increased by d, representing activity generated by a target. If the maximum of the rv's exceeded a criterion c_p, the response was defined as 'Yes' (giving a hit or false alarm); otherwise it was 'No'. For each (m, d, c_p) combination, we ran 5000 trials to obtain stable estimates of Phit and Pfa. (Histograms showed that as m increased, the noise distribution moved to the right and decreased in width, as expected.) Phit and Pfa were used to calculate d' = Z(Phit) - Z(Pfa) and c = -[Z(Phit) + Z(Pfa)]/2. The input variables (m, d, c_p) can now be related to the output variables ((m, d', c). When m =1, by definition d' = d, and setting c_p=d'/2 by definition eliminates bias (c=0).

However, when m>1, d' and c are joint functions of d and c_p. Thus, for each pair (d, m), we had to search though the Monte Carlo outputs to find the value of c_p that would ensure c ~ 0 (in practice, to ensure that $-0.1 < c < +0.1$) to conform to our data. With c_p chosen this way, we found by inspection that $d' = d/m^b$, with the power $b = 0.5 - (0.67)\log d_{10}(d)$, for $1<m<40$. (This is a good approximation, the error in d' being less than 0.16.) This formula predicts that d' will drop at a constant rate with m when d is fixed. In our feature searches, d =d' = 3.2 when m=1, with little variation, so b?is constant at 0.162. Our data do not agree with such constancy; d' dropped off at different rates from 3.2 at m=1, depending on the conditions.

<u>Oddity searches.</u> We also simulated a maximum model for oddity using Monte-Carlo. In The visual system is assumed to find the centroid of all the items on a trial. If any of the item-centroid color distances are greater than a fixed criterion (k), the system reports Yes (an odd item is present) and otherwise No. This model predicts d' if noise is included, because on each trial, k may be exceeded (or not) whether an odd item is present or absent. To model the noise, items (1, .,i.,. m) on a model trial have co-ordinates (u'_1+E, u'_2+E,. , u'_i+E, ... u'_m+E) and (v'_1+E, v'_2+E,.., v'_i+E, ,..v'_m+E), where each noise term, E, is an independently-chosen rv. The rv's, whose parent distributions were Gaussian with mean 0 and standard deviation s, were truncated at +- 3 standard deviations. In addition, every potential coordinate was checked to see if it lay inside the color space, and if not, its error (E) was re-sampled. (Color space has a curved boundary, but the model used a quadrilateral approximation accurate to 0.02 units in both u' an v'.) Since the boundary is one-sided, the resulting distributions were unsymmetrical. Runs in which the re-sampling rate was excessive, exceeding 20%, were excluded. The model was also run with each v' coordinate scaled by 1.4, as (u', v') color space is distorted in this manner for distances inferred from the scaling of reaction times (Cavonius and Mollon, 1986). After

obtaining a set of co-ordinates on each trial, the model calculated the centroid (u^o, v^o) for that trial and reported Yes if $\max[\mathrm{dist}\{(u^o,v^o),(u'_i+\varepsilon,v'_i+\varepsilon)\}] > k$, where dist {.} is the Euclidean distance function, and max[.] is taken over the 1..i..m items. Five hundred Monte-Carlo trials were run at each set size m = 5, 10, 20, 40, and various ranges of k and s. We defined "reasonable" d's as lying between 0 and 4, and recorded only those (m, k, s) combinations which produced reasonable d's. Two sets of colors were employed in modeling: green-yellow-orange distractors with blue-purple targets (set 1), and blue-purple distractors with yellow-green targets (set 2). In the case of set 1 colors, using $0.01<s<0.05$ for both targets and distractors ensured relatively few resamples and reasonable d's (as defined). With set 2 colors, $0.10< s<0.17$ for the purple distractors and $s<0.05$ for the yellow-green targets produced reasonable d's (using the same s for both did not). The result was that for each value of s and m, model d' increased linearly with k, but the slope varied with both m and s. We therefore chose two criterial values of k, high and low, and interpolated the corresponding values of d' from each linear function. Values of k were stepped finely enough to obtain at least 10 reasonable values of d' on each plot, so interpolation was straightforward. For set 1 (purple targets), the value of d' declined with set size (m) for low criteria, but sometimes increased for high criteria. To illustrate, for set 1 and s=0.04, the interpolated value of d' for k=0.161 declined from 1.6 to 0.3 as m was increased from 5 to 40, but increased from 2.6 to 3.0 over the same range when k=0.164. For set 2 (purple distractors), d' declined with set size both at high and low criteria. E.g., for the low criterion, k=0.11, d' dropped from 2.0 at m=5 to 1.2 at m=40 when s=0.10 and from 1.6 to 0.6 when s=0.16. For the high criterion, k=0.12, d' dropped from 2.7 at m=5 to 2.1 at m=40 when s=0.10 and from 2.4 to 1.4 when s=0.16. Thus d' <u>decreased</u> with set size, not increased, for values of s and k which generate model d's commensurate with the data d's for heterogeneous distractors (e.g. d' = 1.9 at m=11 to d' = 2.5 for m=40 for desaturated distractors).

In a randomized design, in which set size was unpredictable, one might expect that the criterion would be independent of set size. However, a reviewer suggested an alternative account in which k is automatically scaled by the standard deviation of the color co-ordinates. The s.d.s do not change with set size on target absent trials but they decrease with set size (by 3% for u' and by 25% for v') on target present trials when the additional distractors tend to wash out the aberrant target chromaticity. However, scaling k by the geometric mean of these s.d.s was too drastic. After scaling, if k and s were chosen so that d' was 2.0 or below for m = 5 or 10, d' increased, yes, but to infinity (recall, the noise was truncated) for m = 40. This was so for every value of s which created reasonable d's. Thus we can reject both these simple versions of the maximum model. A more sophisticated version in which the criterion is only partially scaled by the standard deviations might be promising.

FIGURE LEGENDS

Fig. 1. Top panel; all possible stimuli, located on rings at 1.6, 2.4, 3.2, and 4.0 deg of visual angle, with grey bars marking neighboring distances of 0.68, 0.92, 1.27, and 1.72 deg in successive rings. Bottom panel: the fixation cross and an ungrouped display of 10 elements.

Fig. 2. Stimuli in CIE (u', v') space, within the monitor gamut (open squares). The cross denotes the grey field and the triangle the target in feature search. Top panels show saturated (left) and desaturated (right) stimuli in feature search. The bottom left panel shows the linearly separable colors. The bottom right panel shows all the stimuli used in oddity.

Fig. 3. RTs (circles; refer to left ordinates) and d's (squares; refer to right ordinates) for feature search in Experiment 1.1. Distractors were ungrouped. They were homogeneous in color on each trial (top and middle panels), or heterogeneous in color (bottom).

Fig. 4. RTs (circles; left ordinates) and d's (squares; right ordinates) for feature search in Experiment 1.2. Distractors were heterogeneous. They were ungrouped at set size 5, and, at set size 15, either ungrouped (top panel) or placed into 5 groups (bottom).

Fig. 5. RTs (circles; left ordinates) and d's (squares; right ordinates) for feature search in Experiment 1.3. Distractors were heterogeneous, and grouped by <u>proximity</u> (top panel), with each group comprising 3 differently colored disks, or grouped by <u>similarity</u> (bottom), with disks in each group having the same color.

Fig. 6. RTs (circles; left ordinates) and d's (squares; right ordinates) for feature searches in Experiment 2.2 with ungrouped, heterogeneous distractors, either saturated (top panel) or desaturated (bottom).

Fig. 7. As in Fig.6, for grouped, heterogeneous distractors.

Fig. 8. Symbols: Perf in feature searches, versus set size. Homogeneous (HO) distractors (triangles) 'pop out'. With heterogeneous (HE) distractors, Perf is less good, depending on distractor grouping (Gr, circles, or Ungr, squares) and saturation (sat, open symbols, or desat, filled symbols). Lines: model best-fits.

Fig. 9. RTs (circles; left ordinates) and d's (squares; right ordinates) for oddity search in Experiment 3.2. Distractors were ungrouped and heterogeneous, and either saturated (top panel) or desaturated (bottom).

Fig. 10. As in Figure 9, for grouped, heterogeneous distractors.

Fig. 11. Symbols: Perf in oddity searches. Homogeneous (HO) distractors (triangles) show 'pop-out'. With heterogeneous (HE) distractors, Perf improves with additional distractors, depending on distractor grouping (Gr, circles, or Ungr, squares) and saturation (sat, open symbols, or desat, filled symbols). Straight Lines: model fits. Dotted lines are not model-based, but are there to aid the eye.

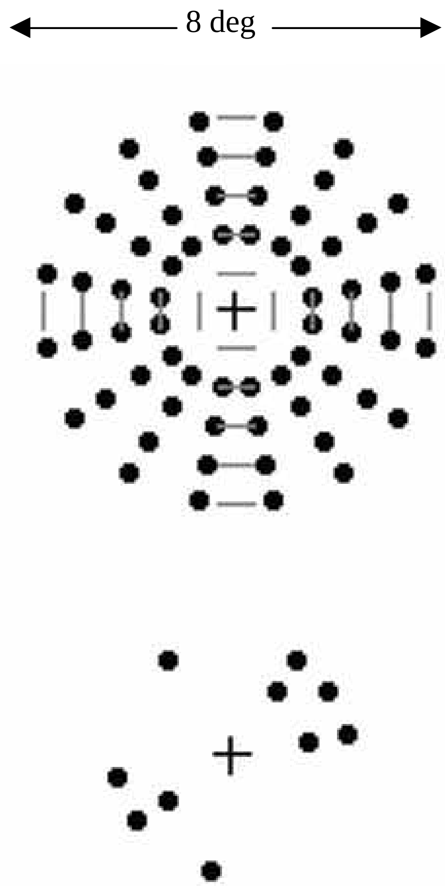


fig. 1

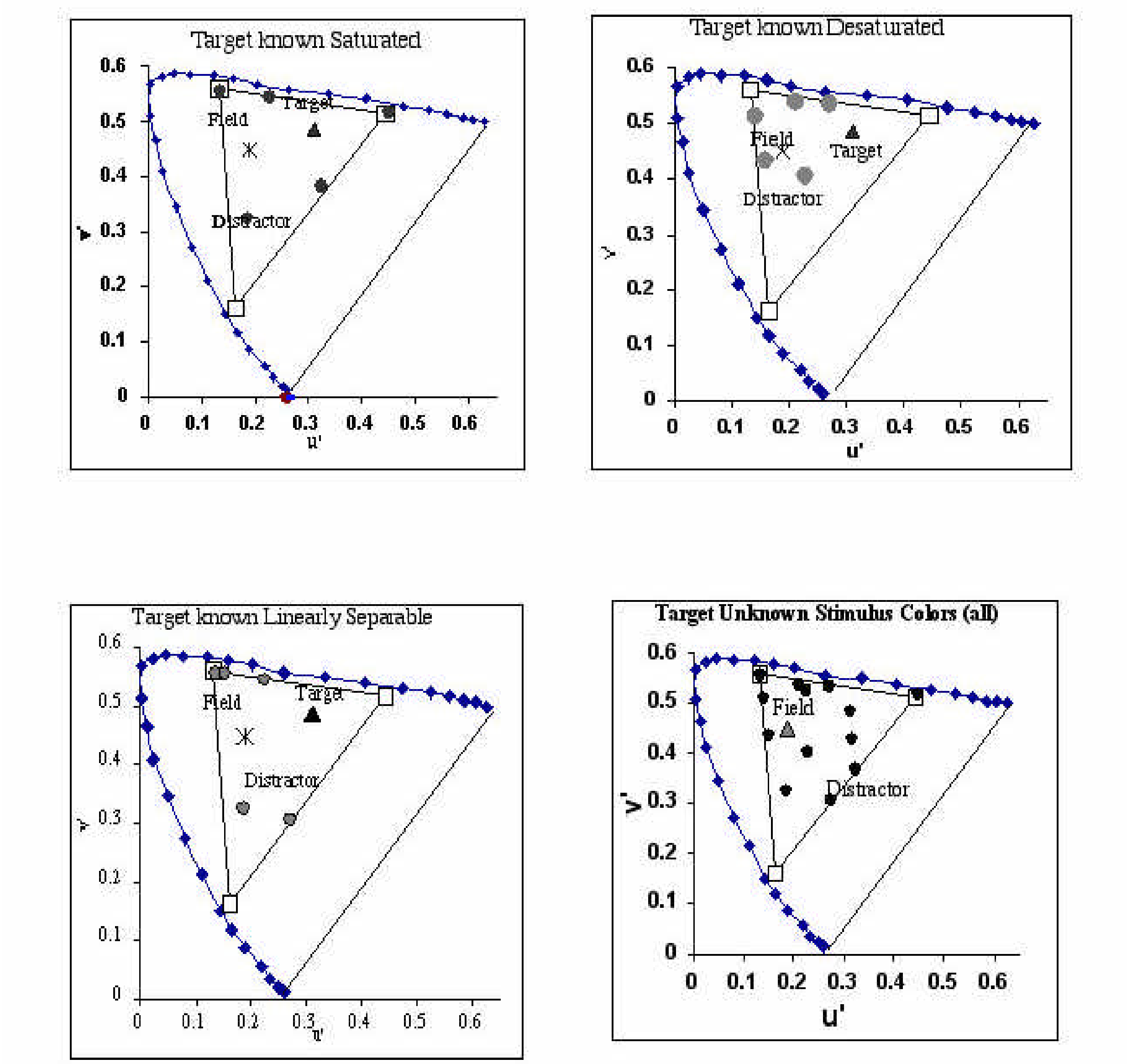

Target known Saturated
Target known Desaturated
Target known Linearly Separable
Target Unknown Stimulus Colors (all)
Field
Target
Distractor
v'
u'
0
0.1
0.2
0.3
0.4
0.5
0.6

fig. 2

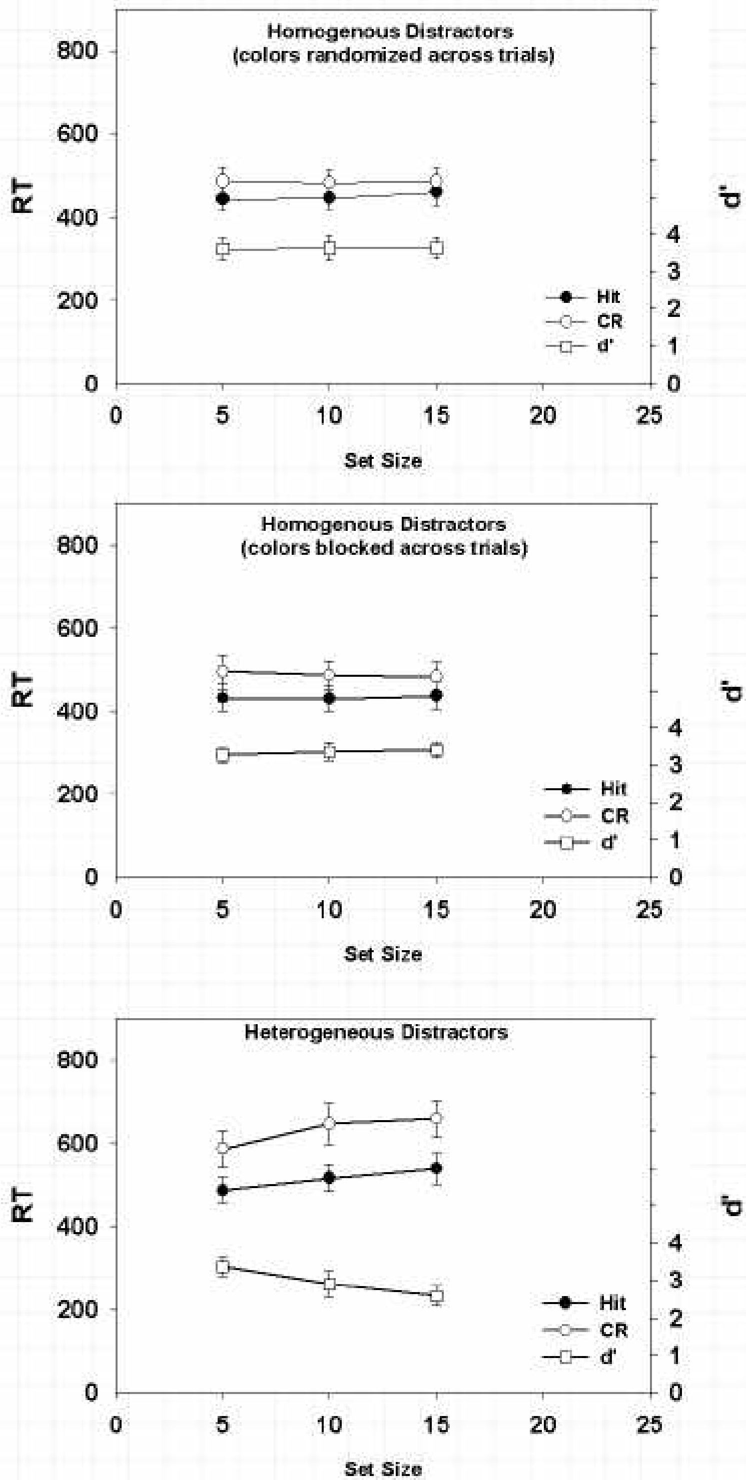

Homogenous Distractors
(colors randomized across trials)
RT
d'
Hit
CR
d'
Set Size
Homogenous Distractors
(colors blocked across trials)
RT
d'
Hit
CR
d'
Set Size
Heterogeneous Distractors
RT
d'
Hit
CR
d'
Set Size

Fig 3

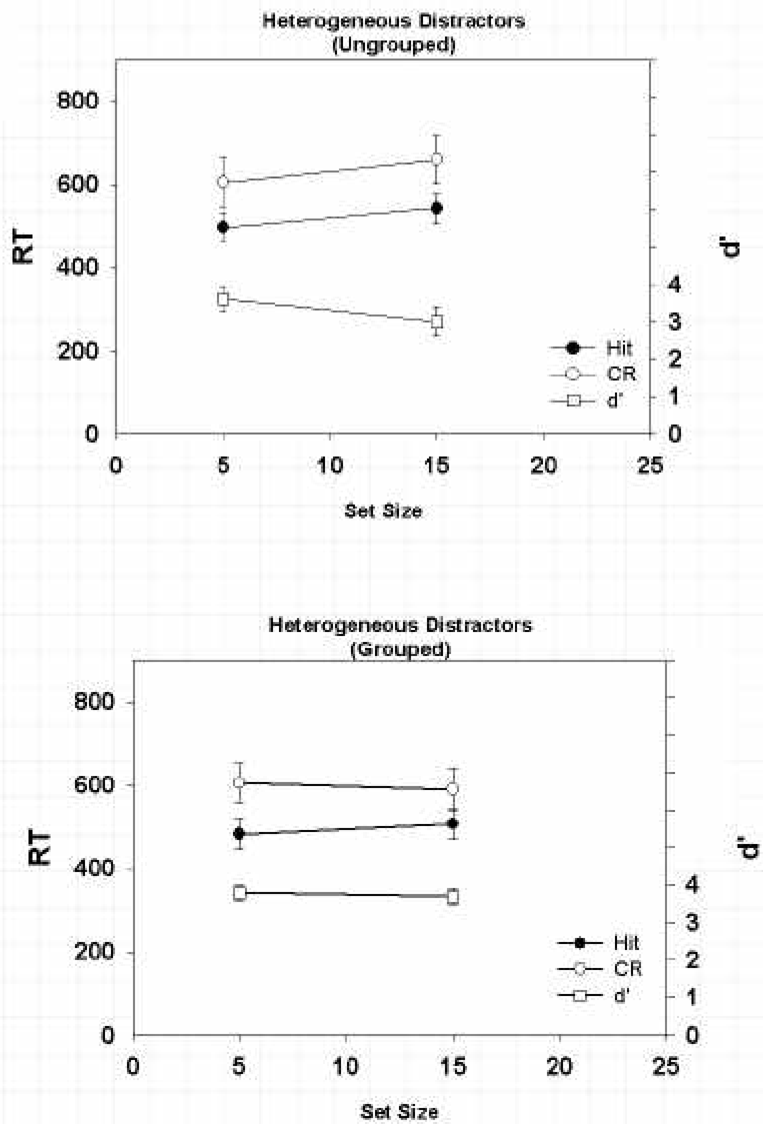
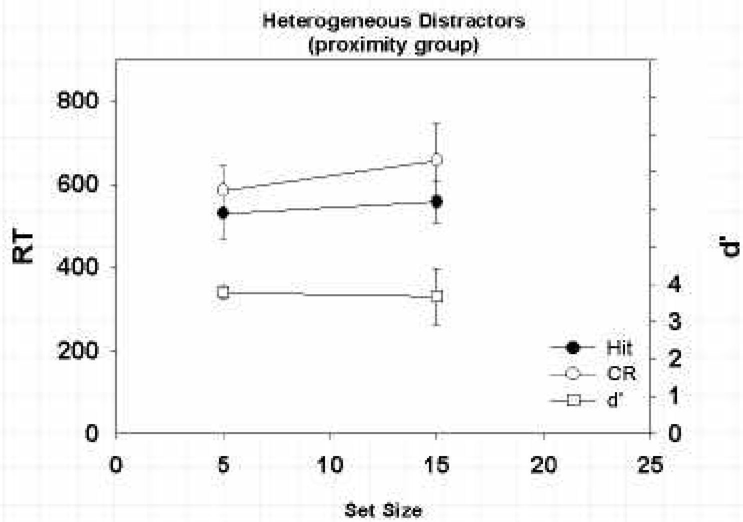

Heterogeneous Distractors
(Ungrouped)
RT
d'
800
600
400
200
0
4
3
2
1
0
Hit
CR
d'
0
5
10
15
20
25
Set Size
Heterogeneous Distractors
(Grouped)
RT
d'
800
600
400
200
0
4
3
2
1
0
Hit
CR
d'
0
5
10
15
20
25
Set Size

Fig 4


Heterogeneous Distractors
(proximity group)
RT
d'
Hit
CR
d'
Set Size
0
200
400
600
800
0
1
2
3
4
0
5
10
15
20
25

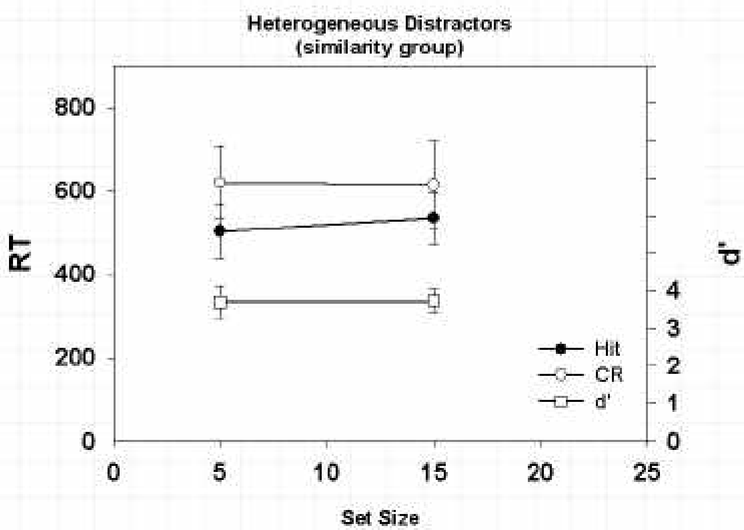
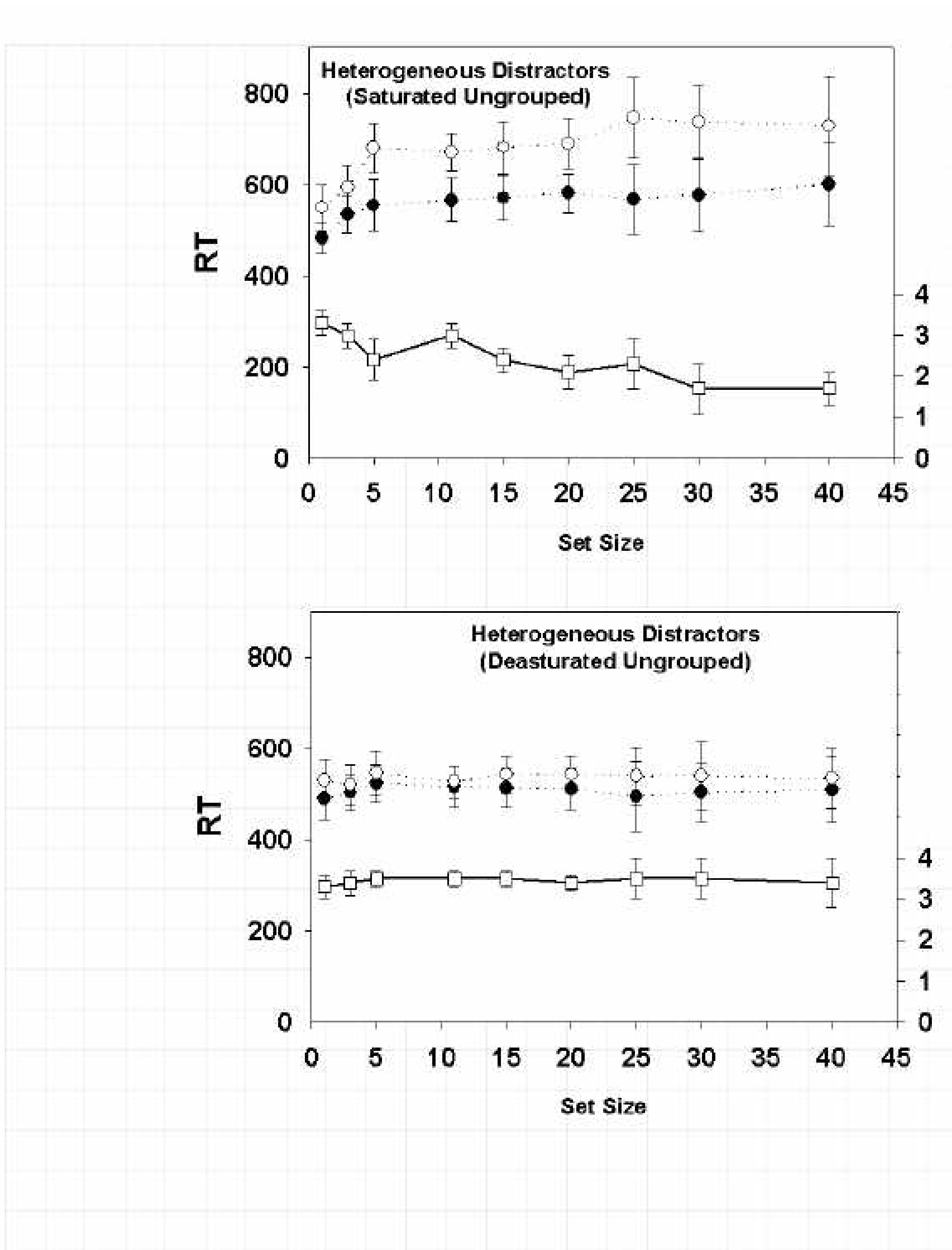

Heterogeneous Distractors
(similarity group)
RT
d'
Hit
CR
d'
Set Size
0
200
400
600
800
0
1
2
3
4
0
5
10
15
20
25

Fig 6



Heterogeneous Distractors
(Saturated Ungrouped)
RT
d'
Set Size
Heterogeneous Distractors
(Deasturated Ungrouped)
RT
d'
Set Size

Fig 6

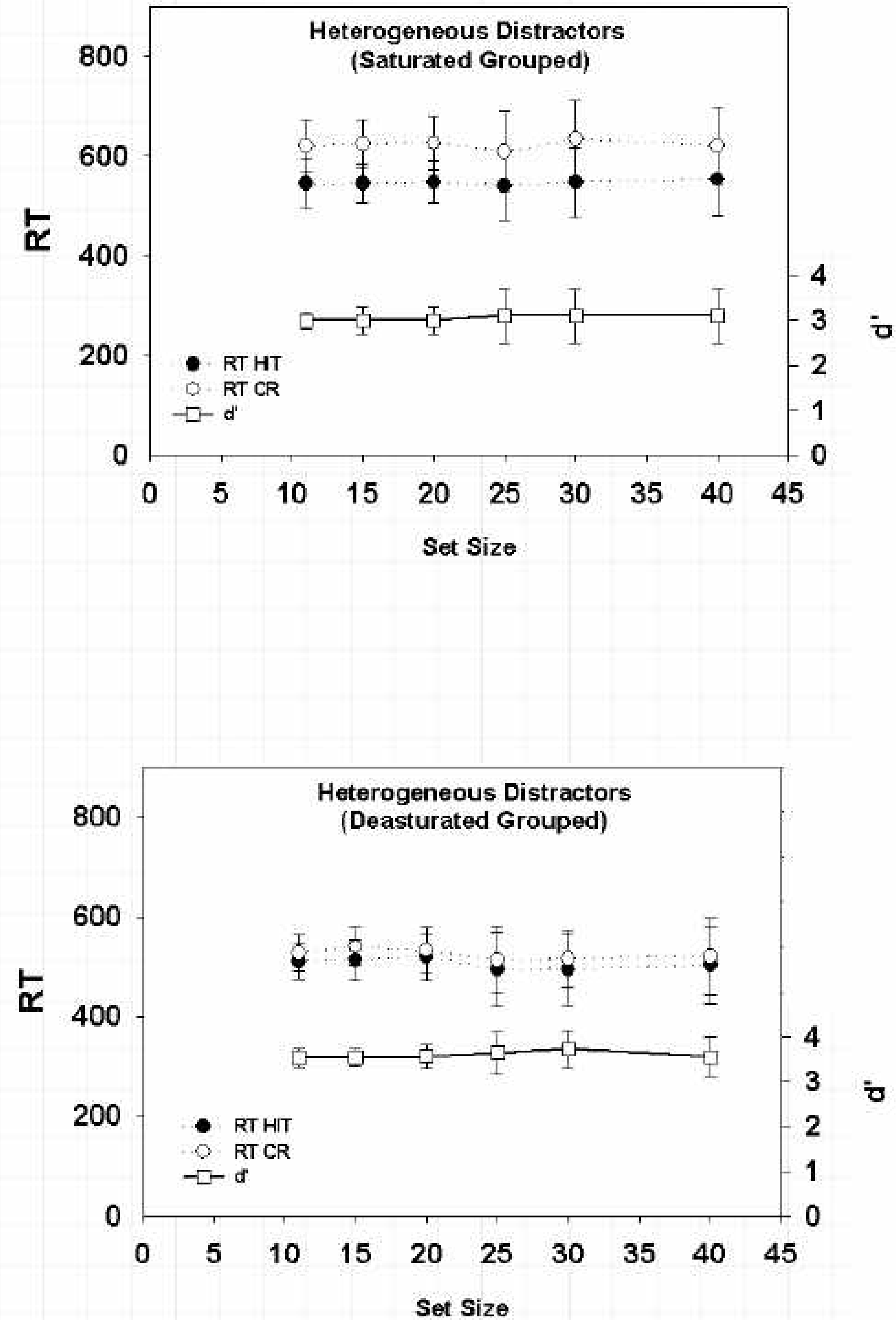

Heterogeneous Distractors
(Saturated Grouped)
RT
800
600
400
200
0
d'
4
3
2
1
0
0 5 10 15 20 25 30 35 40 45
Set Size
RT HIT
RT CR
d'
Heterogeneous Distractors
(Deasturated Grouped)
RT
800
600
400
200
0
d'
4
3
2
1
0
0 5 10 15 20 25 30 35 40 45
Set Size
RT HIT
RT CR
d'

Fig 7

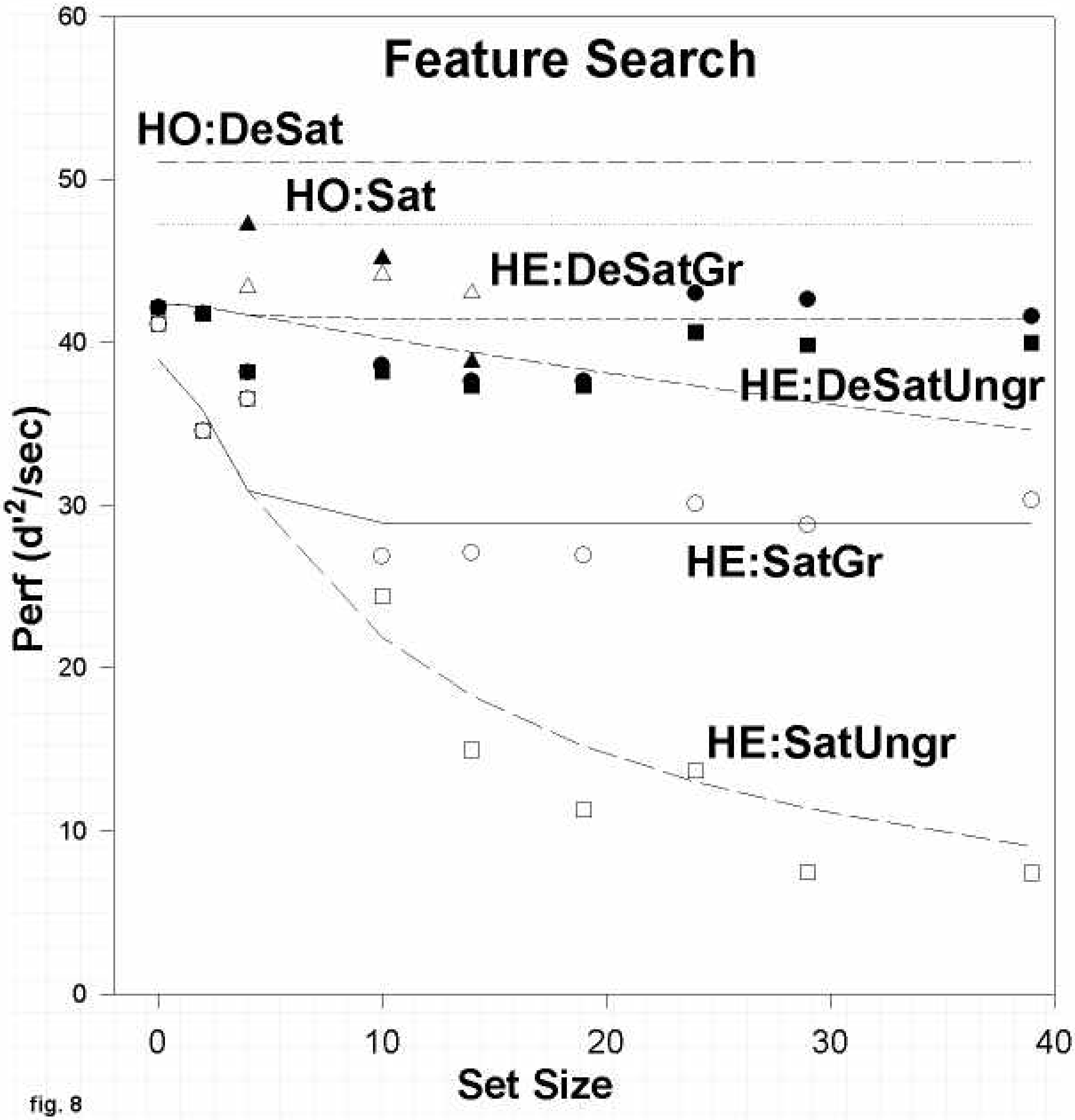

Feature Search
HO:DeSat
HO:Sat
HE:DeSatGr
HE:DeSatUngr
HE:SatGr
HE:SatUngr
Perf (d'2/sec)
Set Size
0
10
20
30
40
50
60
0
10
20
30
40

fig. 8

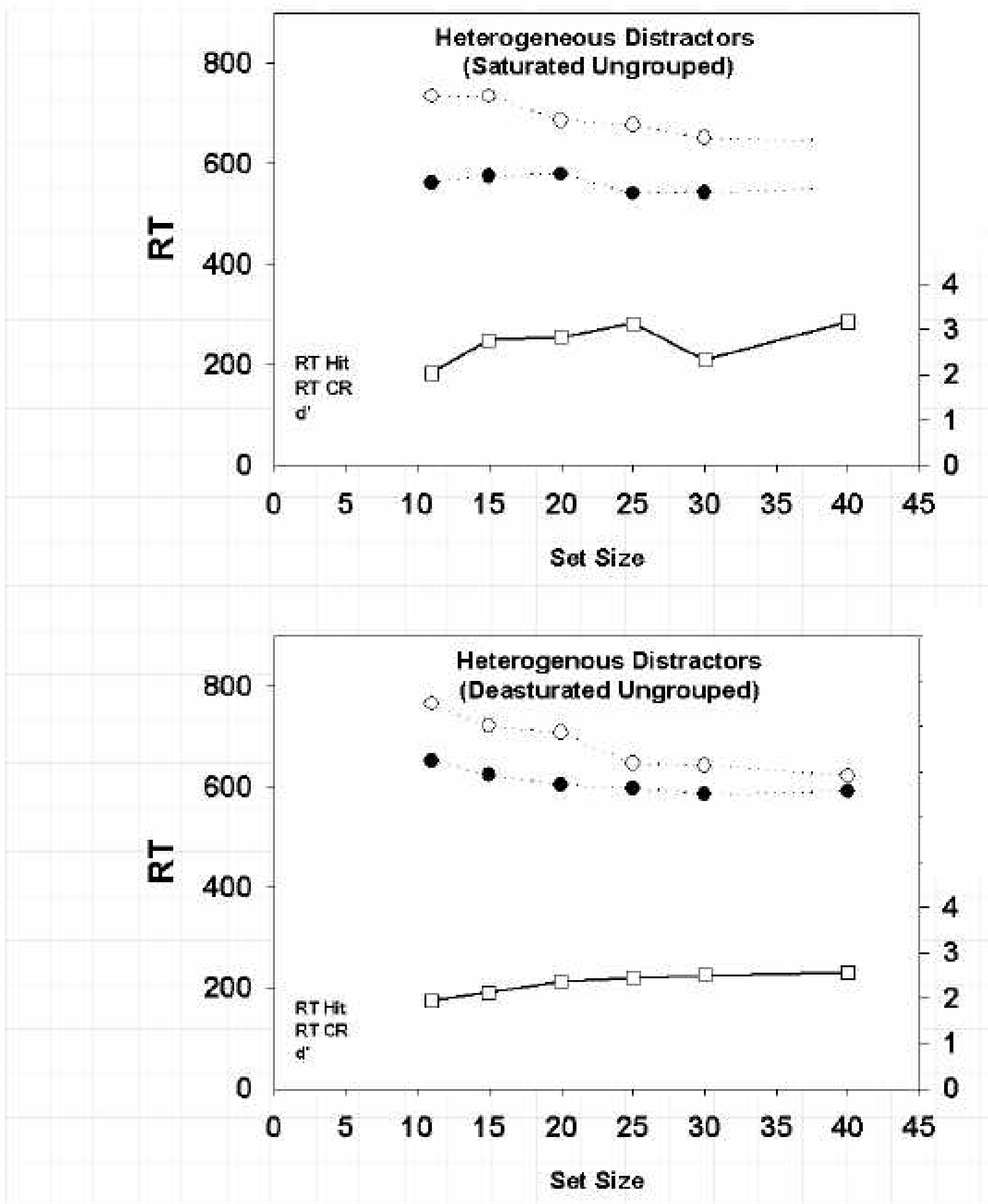

Heterogeneous Distractors
(Saturated Ungrouped)
RT
d'
RT Hit
RT CR
d'
Set Size
Heterogenous Distractors
(Deasturated Ungrouped)
RT
d'
RT Hit
RT CR
d'
Set Size

Fig 9

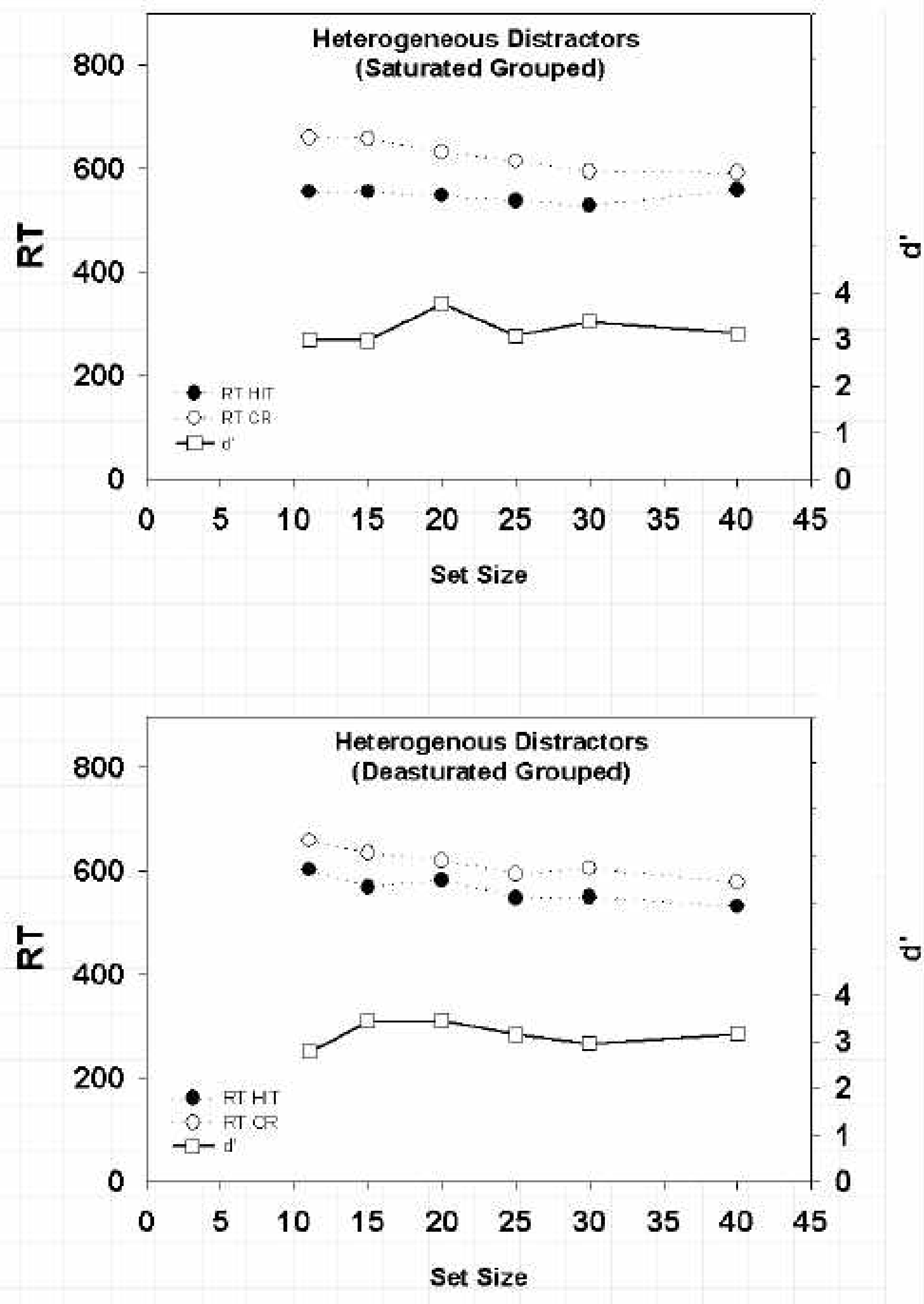

Heterogeneous Distractors
(Saturated Grouped)
RT
d'
RT HIT
RT CR
d'
Set Size
Heterogenous Distractors
(Deasturated Grouped)
RT
d'
RT HIT
RT CR
d'
Set Size

Fig 10

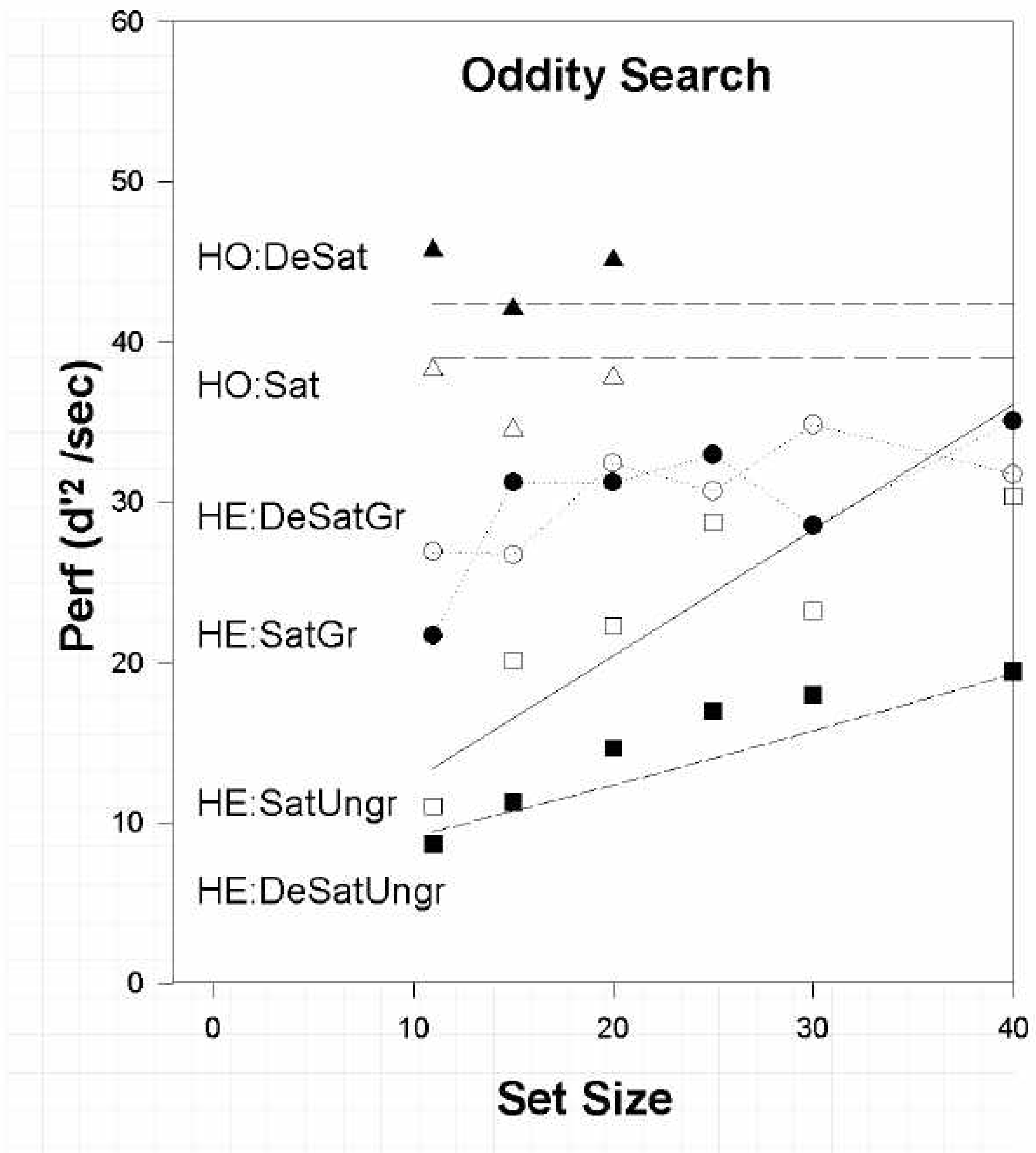

Oddity Search
Perf (d'2 /sec)
Set Size
HO:DeSat
HO:Sat
HE:DeSatGr
HE:SatGr
HE:SatUngr
HE:DeSatUngr
0
10
20
30
40
50
60
0
10
20
30
40

fig. 11