

Detecting conjunctions of color and form: Reassessing the serial search hypothesis

HAROLD PASHLER

University of California, San Diego, La Jolla, California

Treisman and Gelade (1980), who had subjects search for conjunctions of color and form, observed that the slope relating reaction times to number of items in the display was twice as steep for negative (target absent) responses as for positive responses. This result is widely accepted as evidence that a serial self-terminating search is necessary to detect conjunctions. In Experiment 1 of the present study, the subjects searched for a conjunction target (a green T) in displays of 2, 4, or 8 items. Positive and negative display-size slopes were essentially the same in this range. In Experiment 2, the subjects searched for conjunction targets in displays of 2 to 24 items. In displays of 2 to 8 items, the positive and negative slopes were again comparable and the 2:1 ratio was again rejected. In displays of 8 to 24 items, on the other hand, the negative slopes were approximately twice the positive slopes. The same result was obtained in Experiment 3, which was modeled after the conjunction search experiment of Egeth, Virzi, and Garbart (1984). In summary, a 2:1 slope ratio appears to describe the function quite well in the range above 8 items (or if display size is coarsely sampled, as in previous research), but not in the lower range, in which the true ratio is closer to 1:1. These nearly parallel functions in the lower range probably do not, however, reflect a serial exhaustive search: in Experiment 4, positive reaction times were substantially reduced when redundant targets were added to the display. An alternative model of conjunction search is proposed on the basis of these data. This model postulates a molar serial self-terminating search over relatively large clumps of items, and a parallel search within these clumps. It appears, then, that the nature of the attentional limits arising in the process of feature conjunction must be very different from that suggested by Treisman and Gelade.

In a series of recent papers, Treisman and her colleagues have proposed and elaborated a theory of visual attention termed *feature integration theory* (Treisman, 1985; Treisman & Gelade, 1980; Treisman & Schmidt, 1982). This theory proposes that certain perceptual operations can occur *preattentively*—without visual attention—whereas other perceptual operations require that the visual attention mechanism postulated by the theory play a special role. In the preattentive category are those processes that extract relatively low level features from the retinal array. The theory proposes that a wide range of different features are represented in separate (retinotopically organized) maps; these features include color and motion as well as orientation and various other simple elements of form. Registration in the preattentive system permits detection of the presence of the features and underlies segregation of objects by textural cues, but is not sufficient for awareness of the location of features and does not allow different features present in the same locations to be appropriately integrated for purposes of accurate perception or higher level object recognition. According to

Treisman's theory, the latter functions operate accurately only when the visual attention mechanism is focused on individual locations, one at a time, integrating the different features present at those locations.

The strong version of this theory makes the prediction that when a task requires accurate conjunction of different features (e.g., color and form), a subject must attend to the display elements one at a time. Support for this hypothesis comes primarily from visual search experiments reported by Treisman and Gelade (1980). Treisman and Gelade compared several different visual search tasks involving displays of colored letters. In some conditions (conjunction search), subjects searched for a target defined as a conjunction of color and form (e.g., a green T) among distractors sharing features of the target (e.g., brown Ts and green Xs). The subjects made speeded responses indicating the presence or absence of the target. The basic finding was a pattern of reaction times (RTs) in the conjunction search task that suggested a rapid serial self-terminating search. As display size increased from 1 to 30 or more, mean RTs increased linearly with display size. Furthermore, the slope of this increase was approximately twice as steep for negative responses (target absent) as for positive responses (target present). On the other hand, when subjects searched for a disjunction of color and form (e.g., blue or S), RTs did not increase much with display size, and the small increase that did occur was not strikingly linear. Treisman and Gelade also

The author is grateful to James C. Johnston, Mike Mozer, V. Ramachandran, Dave Rumelhart, and Manuel Sanches for useful discussions of this research, and to Jeff Miller for comments on the manuscript. Andrea Diem, Steve Rotblatt, and Manuel Sanches assisted in conducting and analyzing the experiments reported here. Correspondence should be addressed to the author at the Department of Psychology, C-009, University of California, San Diego, La Jolla, CA 92093.

showed that this pattern of results did not depend upon the discriminability of the colors and forms involved; manipulation of discriminability substantially changed the slopes, but left the linearity and 2:1 ratio of negative to positive slopes quite unaffected. In addition, Treisman and Gelade found the same pattern of performance when targets were composed of conjunctions of different features of form, with distractors selected to collectively include these same features (e.g., the letter E among Fs and Ls). Recently, Egeth, Virzi, and Garbart (1984) replicated Treisman and Gelade's basic observations about color/form conjunctions, although Egeth et al.'s results also raised the possibility that subjects may be able to serially scan only those items in the display that share the color of the target, rather than all the items in the display.

Treisman and Gelade (1980) and Egeth et al. (1984) used relatively large displays (up to several dozen items). This feature of the experiments may have certain disadvantages. Large display sizes make the occurrence of eye movements likely (or even necessary). Thus, slopes computed over such a range may reflect the duration of processes not involved in attention shifts per se. Furthermore, in these experiments the display size variable was sampled rather coarsely over the lower portion of its range. For instance, Treisman and Gelade recorded RTs with display sizes of 1, 5, 15, and 25 in a typical study. A display consisting of 1 item may be a special case; therefore, Treisman and Gelade's study is not very informative about the lower range of the display size function.

Two recent studies provide evidence that this limitation may have led to misleading conclusions. Pashler and Badgio (1985), in the course of an investigation of the effects of stimulus degradation and display size in a number of different tasks, examined RTs in the search for a target defined as a conjunction of different elements of form (E among Fs and Ls, following Treisman & Gelade, 1980). Using displays of 2, 4, and 6 items, they observed slopes on negative trials that were somewhat larger than slopes on positive trials; however, the ratio of these slopes was approximately 1.5:1, rather than 2:1. In another recent report, Houck and Hoffman (1986), using stimuli composed of patches of square wave gratings, had subjects search for a conjunction of color and orientation. Over a range of display sizes from 1 to 8 they observed negative and positive slopes with a ratio of about 1:1.

Both of these recent studies sampled the lower range of the display-size function more thoroughly than had Treisman and Gelade (1980) and Egeth et al. (1984), and both call into question the 2:1 generalization. This may or may not reflect a difference in the performance observed in the different studies. It may be that the actual underlying display-size function is the same in all the studies discussed, and the coarse sampling remarked on earlier (e.g., in Treisman & Gelade's study) has simply obscured that fact. Alternatively, there may be a genuine difference in the actual pattern of the display-size function over the lower range as a consequence of method-

ological differences among the various studies. A variety of factors might be responsible for such a difference, several of which were noted by Houck and Hoffman (1986) (the possibilities are not mutually exclusive).

(1) Both of the more recent studies used circular displays, whereas Treisman and her colleagues and Egeth et al. used items scattered around large regions of the display.

(2) In Pashler and Badgio's (1985) experiment, the inclusion of the degradation manipulation (in mixed trials) could have affected subjects' strategies. (3) In Houck and Hoffman's study, the use of oriented grating patterns, rather than alphanumeric characters, might be responsible for the different results. (4) The results obtained by Pashler and Badgio and by Houck and Hoffman may have been dependent upon the fact that large displays (of more than 8 items) were not included in these experiments. Subjects may adopt different strategies for search of a given size display depending upon the range of display sizes they encounter in the experiment.

Houck and Hoffman (1986) pointed out an especially important possibility suggested by their 1:1 slope ratios: with smaller displays subjects may exhaustively search the displays, whereas with larger displays they may switch to a self-terminating search. Such a view could be fully compatible with Treisman's proposal about the basic role of visual attention in conjoining features, since it would still posit an item-by-item serial search. In the present paper I undertake a more systematic empirical examination of the basic color/form conjunction search task, with the goal of adjudicating among these various possibilities.

EXPERIMENT 1

In this experiment, the subjects searched for a green T among green Os and red Ts. Two factors were independently manipulated between trials: the presence or absence of the target, and the number of items in a circular display (2, 4, or 8).

Method

Subjects. Sixteen undergraduates at the University of California, San Diego participated as subjects in the experiment in partial fulfillment of a course requirement.

Apparatus and Stimuli. The stimuli were presented on Princeton Graphics SR-12 RGB color monitors controlled by IBM-PC microcomputers (equipped with Sigma Design Color-400 boards for increased display resolution). Each display consisted of 2, 4, or 8 colored letters presented against a black background. The letters were presented on the perimeter of an imaginary approximate circle. There were eight possible positions on this circle. Displays of 8 elements occupied all of the positions, and displays of 2 and 4 occupied positions on one or two diagonals, respectively. The overall dimensions of the imaginary circle were 4.2 cm (height) x 4.9 cm (width). Based on a typical viewing distance of 50 cm, this approximated 4.8° x 5.6° visual angle. The letters in the display measured about .5 cm (width) x .4 cm (height) (.57° x .46° visual angle).

Displays were composed as follows. The diagonals to be filled were selected at random without constraint. These positions were

then filled, at random, half with red Ts and half with green Os. If the trial was positive, one of the green Os, selected at random, was replaced with a green T. In this way, the displays were constrained to consist of half green and half red elements.

Design. The experiment was divided into 12 blocks, each of which consisted of 60 trials. For each of the six target-present/target-absent x display-size conditions, 10 observations were made in each of the 12 blocks, for a total of 120 observations per session. The order of trials was independently randomized for each block.

Procedure. Each subject was given written instructions indicating that half of the displays would contain a target (a green T), and that the task was to press the "yes" or "no" key as appropriate. The instructions included the statement "Please try to respond as accurately and rapidly as possible on each trial. We will be recording both your speed and your accuracy." Prior to data collection, each subject worked through 96 practice trials, in four miniblocks of 24 trials each.

Each trial began with the presentation of a fixation point (a plus sign). The fixation point appeared at the center of the display for 1,200 msec. It then disappeared, and 50 msec later, the display appeared and remained present until response. The subjects indicated that a target was present or absent by pressing the N (for "yes") or B (for "no") key on the microcomputer keyboard. The subjects kept the index and middle fingers of their dominant hands on these two keys throughout the experiment. If the subject's response was incorrect, a 600-Hz tone immediately sounded for 750 msec to notify the subject of the error; if the response was correct, no tone sounded. The intertrial interval between a correct response and the appearance of the fixation point for the next trial was approximately 1.2 sec. At the end of each block there was a rest period, during which the subject's mean RT in the previous block was displayed. The subject was advised to rest as long as he or she wished, and to resume by pressing the space bar.

Results

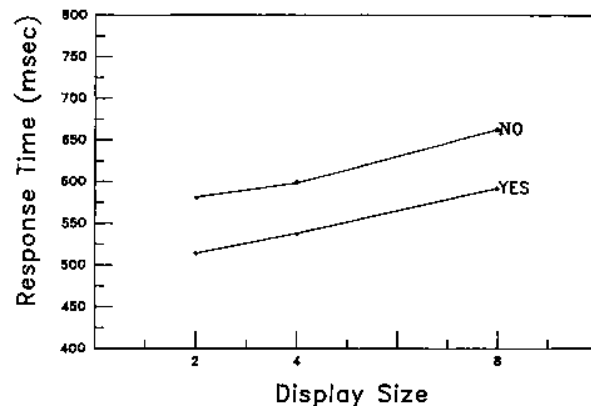
Figure 1 presents subjects' mean correct RTs in the six different conditions, and Table 1 presents the error rates for the same conditions. RTs less than 200 msec or greater than 1,400 msec were discarded. The overall slopes were 13.9 for negatives and 13.1 for positives.

An analysis of variance was performed on the RT data. The effect of target presence or absence was significant [$F(1,15) = 63.6, p < .001$], as was the effect of display size [$F(2,30) = 135.0, p < .001$]. The interaction of display size x target presence or absence was nonsignificant [$F(2,30) = 0.7, p > .50$]. This indicates no evidence that slopes are steeper for negatives than for positives, confirming the impression conveyed by the relatively parallel lines in the figure. Rather than accepting the null hypothesis, however, we would like to know whether the data actually rejected the 2:1 ratio hypothesis. For this purpose, the following analyses were performed. First, for each subject, two least squares slope estimates were obtained, one for negatives and one for positives, based on the subject's mean RTs for each display size. The medians of these estimates were 12.3 for negatives and 13.7 for positives. Next, for each subject, the ratio of the negative to the positive slope was computed. The mean observed ratio across subjects was 1.16 ($SD = .56$, median = 1.01). Seven of the 16 ratios were less than 1:1, and 11 of the 16 were less than 1.5:1. A sign test was performed to assess the hypothesis that the median of the ra-

Table 1
Percent Errors—Experiment 1

Display Size	Target Present	
	No	Yes
2	1.4	2.7
4	1.8	2.4
8	2.7	5.5

Figure 1. Mean response times in Experiment 1.



tios is actually equal to or greater than 2.0. In the sample, 2 out of 16 ratios were greater than or equal to 2.0; thus, the sign test rejected this hypothesis ($p < .002$). A second test was performed to evaluate the 2:1 hypothesis. For each subject, the slope for positives was doubled and subtracted from the slope for negatives. If the slopes exhibit a 2:1 ratio, then the expected value of this statistic is zero. The mean of this difference was -12.21, averaged over subjects. The hypothesis that the true mean is zero was rejected by t test [$t(15) = -4.97, p < .001$].

Discussion

The results of this experiment are quite clear: For a standard conjunction search task, detecting a green T among green Os and red Ts, the slopes relating response time to display size are essentially the same for positive and negative responses when display size is varied from 2 to 8 items. The 2:1 ratio of slopes for "no" and "yes" observed by Treisman and Gelade (1980) over a wider range of display sizes does not appear in this narrow range under the present experimental conditions.

Because the studies that have reported the 2:1 slope ratios have sampled the display size variable rather coarsely over this range, as noted above, the present results do not necessarily contradict these earlier data. On the other hand, subjects' strategies might depend upon the range of different display sizes encountered in a given experiment. Thus, the present results might have been obtained only because of the restricted range of display sizes used, which were all quite small compared to those used in previous studies. The following experiment, then, included both the display sizes of Experiment 1 (2, 4, and 8) and larger display sizes in the range employed in previous studies (16 and 24).

EXPERIMENTS 2A AND 2B

In Experiments 2A and 2B, two factors were independently manipulated between trials: the presence or absence of the target and the number of items in the display (2, 4, 8, 16, or 24). A minor variation in the stimuli was also introduced to check the generality of the findings: in Experiment 2A, the target and distractors were as in Experiment 1 (a green T among green Os and red Ts), whereas in Experiment 2B, subjects searched for a red O among red Ls and green Os. The two experiments did not otherwise differ.

Method

Subjects. Thirty-six undergraduates at the University of California, San Diego participated as subjects in the experiment in partial fulfillment of a course requirement. Eighteen participated in each of the two experiments.

Apparatus and Stimuli. The same equipment was used as in the previous experiment. Each display consisted of 2, 4, 8, 16, or 24 colored letters. The letters were scattered among the 24 positions on a grid composed of four rows of 6 positions. Displays of 24 items occupied all the positions, and displays of fewer items occupied positions chosen randomly from the 24 positions with no constraint. The overall dimensions of the grid were 4.1 cm (height) x 6.1 cm (width). Based on a typical viewing distance of 50 cm, this approximated $4.7^\circ \times 7.0^\circ$ visual angle. The letters in the display measured about .5 cm (width) x .4 cm (height) ($.57^\circ \times .46^\circ$ visual angle). In Experiment 2A, the target was a green T and the distractors were green Os and red Ts. In Experiment 2B, the target was a red O and the distractors were red Ls and green Os. As in Experiment 1, half of the items in all displays were red and half were green.

Design. The experiment was divided into 12 blocks, each of which consisted of 60 trials. For each of the 10 target-present/target-absent X display-size cells, 6 observations were made in each of the 12 blocks, for a total of 72 observations per subject.

Procedure. The procedure followed that of Experiment 1 exactly.

Results

Experiment 2A. Figure 2 presents subjects' mean correct RTs in the 10 different conditions, and Table 2 presents the error rates for the same conditions. RTs less than 200 msec or greater than 3,000 msec were discarded.

An examination of Figure 2 suggests that a clear discontinuity in the relative slopes occurs between display sizes 8 and 16. Therefore, the ranges below and above 8 will be examined separately (in this and subsequent experiments). Obviously, this is not meant to suggest that some substantive change occurs *precisely* at display size 8.

For display sizes ranging from 2 to 8, slopes were computed over the mean RTs (averaged over subjects). These slopes were 16.4 for negatives and 13.3 for positives (a ratio of 1.24:1). They appear, then, to differ by much less than a 2:1 ratio. As in Experiment 1, the 2:1 hypothesis was tested in two ways. As a preliminary, for each subject, least squares slope estimates were computed separately for negatives and positives, for display sizes of 2, 4, and 8 only. The median estimates were 10.1 for negatives and 11.3 for positives. Then the ratio of the negative to the positive slope was computed separately

Table 2
Percent Errors-Experiment 2A

Display Size	Target Present	
	No	Yes
2	3.8	3.1
4	2.5	2.8
8	2.1	4.7
16	2.6	6.6

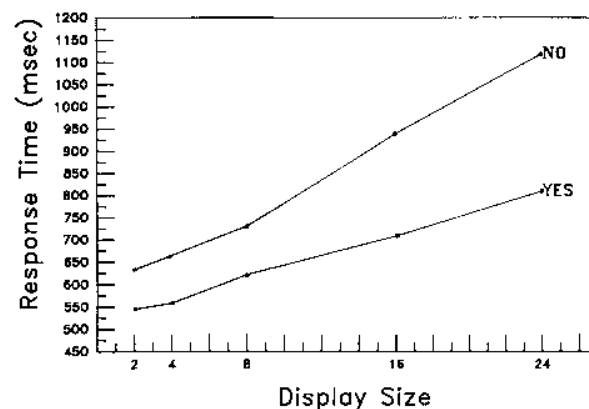


Figure 2. Mean response times in Experiment 2A.

for each subject. The median of these ratios was 0.99. Nine of the 18 ratios were less than 1:1, and 14 of the 18 ratios were less than 1.5:1. A sign test was employed to evaluate the null hypothesis that the true median of these ratios is greater than or equal to 2.0. Three of the 18 medians were greater than or equal to 2.0, rejecting the hypothesis ($p < .003$). The second analysis examined the difference between the negative slope and twice the positive slope, computed for each subject. The mean of this statistic, across subjects, was -10.1, which differed from the zero expected on the basis of the null hypothesis [$t(17) = -3.92, p < .001$]. The results for larger display sizes are discussed below.

Experiment 2B. Figure 3 presents subjects' mean correct RTs in the 10 different conditions, and Table 3 presents the error rates in these conditions. RTs less than 200 msec or greater than 3,000 msec were discarded. Data from one subject were discarded due to extraordinary error rates.

For display sizes 2-8, slopes were computed over the mean RTs (averaged over subjects), and these were 19.4 for negatives and 13.2 for positives (a ratio of 1.47:1). As in the previous experiments, for each subject, least squares slope estimates were computed separately for negatives and positives. The median estimates were 10.9 for negatives and 14.2 for positives. The ratio of the negative to the positive slope was computed separately for each subject. For 2 subjects, the observed slope for positive trials was less than zero. For these subjects, the ratio of slopes was not defined, but was simply assigned an arbitrary maximal value. Thus computed, the median of the slope ratios was 1.18. Seven of the 17 subjects' ratios fell

Table 3
Percent Errors-Experiment 2B

Display Size	Target Present	
	No	Yes
2	1.8	1.4
4	0.8	2.4
8	0.9	1.8
16	1.1	3.6
24	0.5	7.5

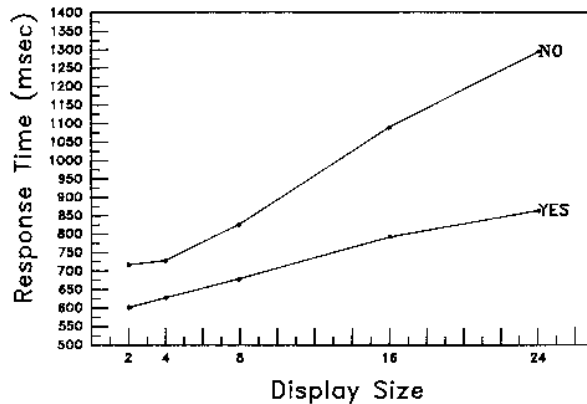


Figure 3. Mean response times in Experiment 2B.

below 1.0, and 13 of the 17 fell below 1.5:1. A sign test was used to evaluate the null hypothesis that the true median of these ratios is greater than or equal to 2.0. Four of the 17 medians were greater than or equal to 2.0 (this included the data of the 2 subjects with positive trial slopes less than zero), rejecting the 2:1 hypothesis ($p < .05$). Due to the cases of negative slopes for positive RTs, the test for differences used in the previous experiments could not be applied. The results for larger display sizes are discussed below.

Discussion

The results of Experiments 2A and 2B in the range of display sizes from 2 to 8 replicate those of Experiment 1 fairly satisfactorily, despite the differences in display configuration and the inclusion of much larger display sizes among the mixed trials. In neither case is the ratio of negative to positive slopes, computed over the mean RTs across subjects, very close to 2.0 (the ratio is 1.24 in Experiment 2A and 1.47 in Experiment 2B). In both Experiments 2A and 2B, the 2:1 hypothesis was statistically rejected.

Over the range of display sizes from 8 to 24, however, the pattern is quite different, and is generally compatible with the 2:1 hypothesis. For Experiment 2A, the slopes over this range are 24.3 and 11.7 msec per item for negatives and positives, respectively. For Experiment 2B, the slopes are 29.2 and 11.5, respectively. Furthermore, the *overall* display-size function is grossly compatible with previous published results. For instance, suppose we analyze the complete display-size range (2-24) in the manner of Treisman and Gelade (1980), but without deleting any of the display sizes. For Experiment 2A, the overall

slopes are 22.6 for negatives and 12.1 for positives, with the linear function accounting for 99.5% and 99.7% of the variance, respectively. For Experiment 2B, the overall slopes are 27.6 for negatives and 12.2 for positives, with the linear function accounting for 99.2% and 99.3% of the variance, respectively. Thus, even with all the display sizes included, an examination of linearity and the ratio of the best fitting slopes following Treisman and Gelade (1980) would lead to the conclusion that both of the present experiments support the serial self-terminating model. In each case, however, the RT data from display sizes 2-8 actually refute such an interpretation.

In summary, then, the gross character of the display-size function found in Experiments 2A and 2B is quite consistent with that found in previous work, even though a fine-grained analysis of the slopes contradicts the self-terminating search model. However, the actual slopes observed in the present experiments are shallower than those found in any of Treisman and Gelade's (1980) conjunction search conditions. Perhaps the most relevant experiment for comparison is Treisman and Gelade's Experiment 2, which included two conditions of target/background discriminability in a color-letter conjunction search. The lowest slopes obtained (in the higher of the two discriminability conditions) were 39.5 for negatives and 20.5 for positives. Why are the slopes found in the present experiments not much more than half the values of those observed by Treisman and Gelade? There are several possible reasons. First, the items in the present experiment were presented somewhat less peripherally than those in Treisman and Gelade's experiments. The likelihood that this factor played a role is lessened by the fact that Treisman and Gelade took precautions to ensure that display size and eccentricity were not confounded. A second possibility is that Treisman and Gelade's subjects for some reason elected a more conservative speed-accuracy tradeoff. The average percentages of errors in the present experiments (4.06% and 2.18% in Experiments 2A and 2B, respectively) are not much different from the 2.5 % reported by Treisman and Gelade in their high-discriminability condition. It is difficult to know whether Treisman and Gelade's subjects may have been unusually conservative for some reason, but the written instructions in the present experiments, together with the auditory warning sounded in the case of errors, would seem to provide at least as much accuracy emphasis as is typical in studies of this sort. A third possibility is that the visual quality of the displays may have been higher in the present experiments, constituting a "very high-discriminability condition" in terms of Treisman and Gelade's Experiment 2; informally speaking, the letters were crisp and the colors highly saturated.

EXPERIMENT 3

As noted above, in Experiments 1 and 2 I observed lower slopes than those found in some previous studies dealing with conjunction search. It seemed worthwhile

to include an experiment modeled closely after one of these previously published experiments, to be sure that the present procedures did not differ in some unknown way from other authors' procedures. The experiment chosen for (approximate) replication was Egeth et al.'s (1984) Experiment 2. This experiment was regarded by its authors as a satisfactory replication of the basic results of Treisman and Gelade (1980), with observed slopes of 32 for negatives and 21 for positives. The Egeth et al. experiment was chosen as a model because Egeth et al. used letters placed randomly in positions of a grid, a procedure for which the equipment used in the present experiment is well suited. There were a few differences, however. Egeth et al. used display sizes of 5, 15, and 25. In order to maintain the more fine-grained analysis of the lower display sizes, the present experiment used display sizes of 2, 4, 8, 16, and 24, as in Experiments 2A and 2B. Egeth et al. ran subjects for 12 practice trials, followed by 72 experimental trials. The present experiment included three phases: a practice set of 12 trials; Part A, which consisted of 80 trials in a single block and which was designed to approximate the levels of practice observed by Egeth et al.; and Part B, which consisted of another 720 trials of exactly the same task, divided into 12 blocks.

Method

Subjects. Twenty-five undergraduates at the University of California, San Diego participated as subjects in the experiment in partial fulfillment of a course requirement.

Apparatus and Stimuli. The same equipment was used as in the previous experiments. Each display consisted of 2, 4, 8, 16, or 24 colored letters. The letters were scattered among the 36 positions on a grid composed of six rows of 6 positions. The positions occupied by any given display were chosen at random without constraint. The outer dimensions of the display were 6.1 cm (width) x 6.7 cm (height). The target was a red O, and the distractors were red Ns and green Os. As in the previous experiments, displays were constrained to consist of half red and half green items.

Design. The first postpractice portion of the experiment—Part A—consisted of 80 trials in a single block, 8 in each of the 10 target-present/target-absent x display-size conditions. Part B was divided into 12 blocks, each of which consisted of 60 trials. For each of the 10 target-present/target-absent x display-size conditions, 6 observations were made in each of the 12 blocks, for a total of 72 observations per subject.

Procedure. The procedure followed that of Experiment 1, with the following exceptions. The subject first did 12 practice trials, described as practice that "did not count." Both Part A and Part B were described as part of the experiment proper, and no explanation was given for the division of the experiment into two parts.

Results

Part A. Figure 4 presents subjects' mean correct RTs in the 10 different conditions, and Table 4 presents the error rates for the same conditions. The data from 1 of the 25 subjects were not included in the analyses because a portion of one of his data files was inadvertently lost. RTs less than 200 msec or greater than 3,000 msec were discarded. First the complete data set (i.e., from display

Display Size	Target	Present
	No	Yes
2	4.7	3.1
4	5.7	2.1
8	1.0	0.0
16	1.0	7.8
24	2.1	12.5

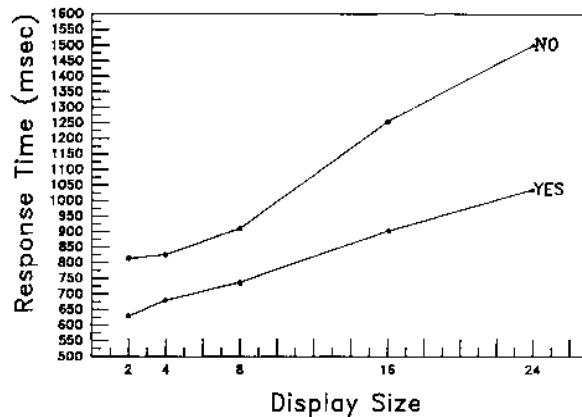


Figure 4. Mean response times in Experiment 3, Part A.

size 2 to display size 24) was analyzed. The overall slopes were 33.2 for negatives and 18.5 for positives, with the linear trend accounting for 98.5% and 99.8% of the variance, respectively.

For the range of display sizes from 2 to 8, slopes were computed over the mean RTs, over subjects. These slopes were 17.2 for negatives and 17.7 for positives. This shows no hint of a 2:1 ratio of negative slopes to positive slopes. The 2:1 hypothesis was tested as in the previous experiments. For each subject, least squares slope estimates were computed separately for negatives and positives, for display sizes 2, 4, and 8 only. Next, the ratio of the negative to the positive slope was computed separately for each subject. For 13 of 24 subjects, this ratio was less than 1:1 (this includes 5 subjects whose slopes for negatives were less than zero, with slopes for positives greater than zero). A sign test was employed to evaluate the null hypothesis that the true median of these ratios is greater than or equal to 2.0. Eight of the 24 medians were greater than or equal to 2.0 (this included the data of the 2 subjects with positive slopes less than zero). The difference was not quite significant ($p < .06$). Because of the negative slopes for some subjects, the difference statistic used in the previous experiments could not be employed. The fact that the 2:1 ratio could not be statistically rejected here probably reflects the instability of the means due to the small number of trials per subject in Part A.

Part B. Figure 5 presents subjects' mean correct RTs in the 10 different conditions, and Table 5 presents the error rates in these conditions. RTs less than 200 msec or greater than 3,000 msec were discarded. First, the

Table 5
Percent Errors -Experiment 3, Part B

Display Size	Target Present	
	No	Yes
2	2.8	2.7
4	2.5	2.8
8	1.9	3.1
16	2.1	5.5
24	2.2	10.4

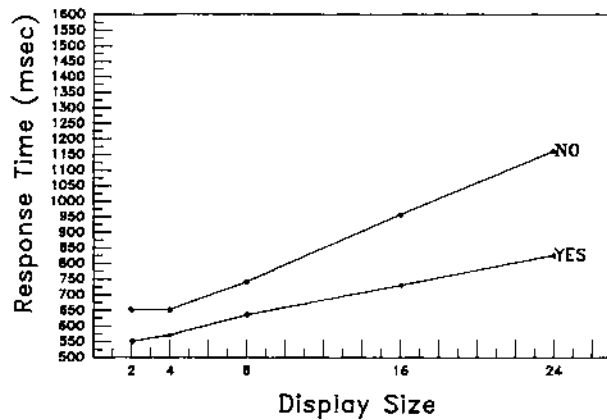


Figure 5. Mean response times in Experiment 3, Part B.

complete data set (i.e., from display size 2 to display size 24) was analyzed. The slopes computed over the means (across subjects) were 24.3 for negatives and 12.6 for positives, with the linear trend accounting for 99.1 % and 99.8% of the variance, respectively. Next, display sizes 2, 4, and 8 were analyzed separately. As in the previous experiments, for each subject, least squares slope estimates were computed separately for negatives and positives. The mean estimate (averaged across subjects) was 15.0 for negatives and 13.7 for positives (medians 11.9 and 13.1, respectively). Then the ratio of the negative to the positive slope was computed separately for each subject. The median was 0.97. For 12 of the 24 subjects, this ratio was less than 1:1. A sign test was used to evaluate the null hypothesis that the true median of these ratios is greater than or equal to 2.0. Four of the 24 medians were greater than or equal to 2.0, rejecting the hypothesis that the true median is greater than or equal to 2.0 ($p < .001$). Because 1 subject had a negative slope less than zero, the difference statistic used in previous experiments could not be applied.

Discussion

Part A was designed as a fairly close replication of Egeth et al.'s (1984) Experiment 2. The only major difference between the studies was in the number of display size points (2, 4, 8, 16, and 24 in the present study, vs. 5, 15, and 25 in Egeth et al.'s study). The present results replicate those of Egeth et al. very well. The overall slopes of 33.2 and 18.5 found in the present study are comparable to the slopes of 32 and 21 found by Egeth et al. A clear practice effect was apparent: for the 720

trials in Part B of the experiment, the slopes averaged 24.3 and 12.5, values very similar to those found in Experiment 2.

The present experiment has several implications. Among them is the concern caused by the fact that the overall slopes reported above are simply lower than those previously reported in the literature. The results of Part A in the present experiment are quantitatively and qualitatively very similar to those of Egeth et al. (1984), after which Part A was closely modeled; thus, the fact that the slopes observed here are uniformly shallower than any of Treisman and Gelade's does not appear to be due to idiosyncracies in the present procedures. The reason for the differences cannot be determined from available information.

The second notable result of Experiment 3 is that the main results of Experiments 2A and 2B were closely replicated, both with the very low levels of practice observed in Part A and over the remainder of the session (Part B). As in the previous experiments, the slopes from 2 to 8 are essentially the same for negatives and positives, whereas the negative slopes are much steeper as display size increases from 8 to 24.

EXPERIMENT 4

In each of the previous experiments, for display sizes up to about 8, slopes for negative and positive displays have been fairly comparable. This raises the possibility that the search of these displays may be serial and exhaustive, as suggested by Houck and Hoffman (1986), who obtained very similar results with conjunctions of color and orientation of grating patches. If subjects search exhaustively, then positive responses should not be any faster when more than one target is present than when there is only one. This point was first noted by Holmgren, Juola, and Atkinson (1974), who observed an advantage with redundant targets in an ordinary letter search task. In contrast, a self-terminating search, whether serial or parallel, will be sped up by the presence of redundant targets (see Townsend & Ashby, 1983, for a thorough discussion). The present experiment used display sizes of 2, 4, and 8 (as in Experiment 1), but on half of the positive trials with display sizes 4 and 8, two, rather than one, targets were presented.

Method

Subjects. Eighteen undergraduates at the University of California, San Diego participated as subjects in the experiment in return for payment.

Apparatus and Stimuli. The same equipment was used as in the previous experiments. The stimulus presentation followed that of Experiment 1: displays consisted of 2, 4, or 8 elements, half red and half green, arrayed around an imaginary approximate circle. Stimulus placement was as in Experiment 1.

Design. The experiment was divided into 12 blocks, each of which consisted of 60 trials. There were equal numbers of trials (10 per block) in each of the six combinations of positive versus negative trial \times display size (2, 4, or 8). For display sizes 4 and 8, there were two types of positive trials: those with a single target present

and those with two targets present. Single-target and double-target positive trials occurred equally often; thus, for display sizes 4 and 8, there were 5 trials per block of each type. The constraint that half the items in each display should be red and half green was maintained on the double-target displays. (It was in order to accomplish this that double-target displays were not presented with display-size 2.)

Procedure. The procedure followed that of Experiment 1 exactly, except that the instructions to the subjects mentioned that some displays would contain two targets and that the task was simply to detect the presence of one or more targets in a given display.

Results

Figure 6 presents subjects' mean correct RTs in the eight different conditions, and Table 6 presents the error rates for the same conditions. RTs less than 200 msec or greater than 1,400 msec were discarded.

The main point of interest in the experiment is the comparison of the RTs for single-target and double-target positive trials. A planned comparison was performed to determine the effect of target redundancy on the positive-trial RTs at the two display sizes (4 and 8) for which double-target positive trials were presented: a two-way ANOVA with factors of display size (4 vs. 8) and number of targets (1 vs. 2). The effect of target redundancy was significant [$F(1,17) = 137.5, p < .001$], as was the effect of display size [$F(1,17) = 111.5, p < .001$]. The interaction was not quite significant [$F(1,17) = 3.96, p < .07$].

Slopes were computed over the mean RTs (averaged over subjects). These slopes were 11.7, 14.4, and 10.5 msec/item for negatives, single-target positives, and double-target positives, respectively.

Discussion

The results of the present experiment replicate the failure of the 2:1 generalization over the lower range of display sizes. As in Experiment 1, the ratio of the negative to the positive slopes is not 2:1—it is actually slightly less than 1:1. The main point of interest in the present experiment, however, is the fact that responses on double-target positive trials were significantly faster than responses on single-target positive trials under conditions that produce the roughly 1:1 slope ratios. This rejects a serial-exhaustive-search account of these near-parallel slopes: if the search continued until all items had been searched, two targets would not lead to a faster response than one target. The results extend the finding of Holmgren et al. (1974), who used an ordinary letter search, to the search for a conjunction of color and form.

Townsend and Ashby (1983) pointed out a complication for this form of argument that is worth noting, although it does not undermine the present conclusions. They observed that a serial exhaustive search could produce faster positive responses for redundant-target displays if the targets and nontargets were searched at different rates. If targets were searched faster than nontargets, the search would be completed faster when the more targets were present in the display. However, the present

Table 6
Percent Errors—Experiment 4

Display Size	Targets Present		
	No	Yes(1)	Yes(2)
2	2.3	2.6	
4	3.1	4.9	1.5

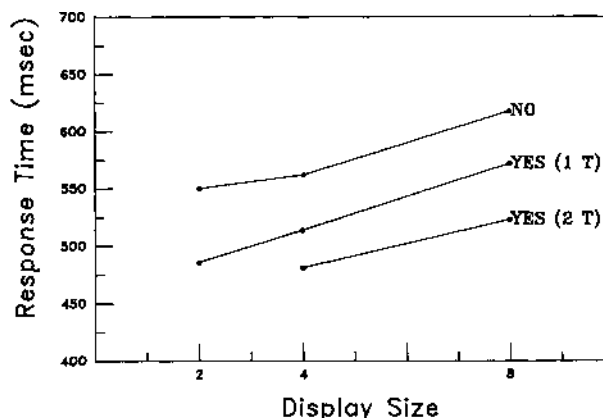


Figure 6. Mean response times in Experiment 4 (1T = one-target positive trials; 2T = two-target positive trials).

data are not at all compatible with this account, given the relative magnitudes of the slopes and the double-target advantage. The maximum decrease in RTs that could be produced by adding a single redundant target to a positive display, according to the differential search rate effect noted by Townsend and Ashby, would be the time to search one nontarget (because the time to search a target cannot be less than zero). Nontarget search time is estimated by the slope of the negative responses. In the present experiment, the slope is 11.7 msec per item, and the advantage for double-target trials is 33 msec and 49 msec for display sizes 4 and 8, respectively. Both of these advantages are significantly larger than the slope: for display size 4, $t(17) = 3.65, p < .002$, and for display size 8, $t(17) = 7.29, p < .001$. Finally, the differential search rate effect would not account for the trend toward an interaction of target redundancy with display size that was noted above: the advantage of redundancy may be greater for display size 8 than for display size 4. It is worth noting, however, that the results might still be consistent with a serial-exhaustive-search model that posits that later stages of processing operate faster when multiple targets have been detected in the search stage. Alternative accounts of what may be responsible for the display size functions will be considered below.

CONCLUSIONS

The data reported here have shown that the empirical generalization offered by Treisman and Gelade (1980)—2:1 slope ratios for negatives and positives in search for conjunctions of features—does not apply to the lower

range of the display-size function. Two later studies (Houck & Hoffman, 1986; Pashler & Badgio, 1985) provided some indication that the 2:1 generalization was not reliable in this range. The apparent conflict between Treisman and Gelade's results and those of the more recent studies raised a number of possibilities, which were enumerated in the introduction. It is now possible to evaluate each of these possibilities in turn.

1. The appearance of slope ratios near 1:1 over the lower display size range does *not* stem from the use of circular displays (as in Houck & Hoffman, 1986, and Pashler & Badgio, 1985); it appears both with circular displays (Experiments 1 and 4 of the present study) and with displays composed of elements randomly distributed in large grids of potential locations (Experiments 2 and 3 of the present study).

2. Plainly, the slope result has nothing to do with the stimulus degradation manipulation included in Pashler and Badgio's (1985) study, inasmuch as degradation manipulations played no role in the present study.

3. The slope result with color/form conjunction search does not depend upon whether the form discrimination concerns orientation of grating patches (as in Houck & Hoffman, 1986) or a simple discrimination between featurally dissimilar letters (as in Treisman & Gelade, 1980, and the present experiments).

4. The slope result does not depend upon whether smaller display sizes are used exclusively (as in Houck & Hoffman, 1986, and Pashler & Badgio, 1985) and the present Experiment 1), or whether the smaller and larger display sizes are included together in mixed trials (as in the present Experiments 2A, 2B, and 3).

Houck and Hoffman (1986) suggested that subjects may exhaustively scan the smaller displays, and search the larger displays in a self-terminating fashion. The present Experiment 4 provided direct evidence against an exhaustive search of the smaller displays. This experiment included only display sizes 2, 4, and 8, and the observed ratio of negative to positive slopes was 0.81. The experiment also included some trials in which there were two targets present, at display sizes 4 and 8. Positive responses to these double-target displays were significantly faster than responses to single-target displays of the same size. Such a finding, as originally pointed out by Holmgren et al. (1974), speaks against an exhaustive model.

Other data from the present study also argue against the hypothesis that subjects search small displays exhaustively and larger displays in a self-terminating fashion, with a constant search rate throughout. Consider the predictions that such a model makes for the slopes for negative trials. Since the negative responses are *always* based on exhaustive search, according to this model, these slopes should not change as display size increases beyond 8. The present data do not support this prediction. In Experiment 2A, the negative slopes are 16.4 for display sizes 2-8 and 24.3 for display sizes 8-24. The corresponding slopes are 19.4 and 29.2 in Experiment 2B, 17.2 and 37.0

in Experiment 3, Part A, and 14.5 and 26.3 in Experiment 3, Part B. Plainly, the negative slopes increase dramatically as display size increases from the 2-8 range to the 8-24 range.

Consider now the positive trials. If responses at display size 16 reflected a serial self-terminating search, then they should be no slower than the positive responses at display size 8. This was clearly not the case in the present study. Furthermore, if the self-terminating search model were accurate, the slope of the positives for display sizes 16-24 should be only half the slope of the positives for display sizes 2-8. In Experiment 2A, however, the slope of the positives was 13.3 msec per item for display sizes 2-8 and 12.7 msec per item for display sizes 16-24. The corresponding slopes were 13.2 and 9.1 in Experiment 2B, 17.7 and 16.9 in Experiment 3, Part A, and 14.5 and 11.9 in Experiment 3, Part B. Plainly, the positive slopes are not halved in the higher display size range.

In summary, the display-size functions in Experiments 2 and 3 and the results of Experiment 4 rule out Houck and Hoffman's (1986) quite reasonable suggestion that performance with lower display sizes might reflect a serial exhaustive search, which changes to a serial self-terminating search at higher display sizes. The present results, then, undermine any form of item-by-item serial model for conjunction search. A fine-grained analysis of the search task does not support the serial interpretation in the way that a relatively coarse examination of slopes and linearity appeared to.

AN ALTERNATIVE ACCOUNT

I will now outline a model that *is* compatible with the data. Suppose that subjects can search up to a modest, fixed number of items at a time; for the present purpose, we will suppose this number to be 8. This search is a capacity-limited parallel self-terminating process that generates comparable, nonzero slopes for negatives and positives. The existence of models with this property has been demonstrated (Townsend & Ashby, 1983). When subjects must search a display of more than 8 items, I propose that they do this in stages, searching clumps of up to 8 items at a time. The parallel self-terminating within-clump searches are embedded in a sequence, that is, a molar serial search process. This molar serial search is also assumed to be self-terminating. Finally, it is assumed that the time to initiate the search of a new clump of items is a nonzero constant. The RT means in Experiment 3, Part B, are based on 1,600 to 1,700 observations per cell, and therefore should provide a data base reliable enough to allow a rough test of the feasibility of this model. The rate of the parallel within-clump search can be directly estimated from the display-size function from 2 to 8. The estimates are 15.0 msec per item for negatives and 13.7 msec per item for positives. When the parameter designating between-clump shift time (*S*) is specified, the four remaining RT means (negative and

positive trials at display sizes 16 and 24) follow-directly. Simply to see if the fit was at all reasonable, I estimated the shift time S based on the positive RTs for display size 16. The relationship is

$$RT(\text{yes}, 16) = RT(\text{yes}, 8) + ** (S + 8[\text{slope-yes}]).$$

The estimate obtained for S was 76.4 msec. This directly predicts mean RTs of 821, 926, and 1,122 msec for positives at display size 24, negatives at display size 16, and negatives at display size 24, respectively. The observed RTs are 822, 938, and 1,140 msec for the corresponding observed means. Thus, this crude model achieves an excellent fit to the data.

Despite the pleasing fit achieved by this simple model, it is not my purpose to make any strong case for the details of the model as stated. Many of these details are surely arbitrary (e.g., the clump size of 8). An understanding of visual search phenomena seems more likely to be advanced by further experimentation, testing the predictions of entire classes of models, rather than by the premature elaboration of any particular model in detail. More important than the fit achieved, then, is the fact that a molar serial self-terminating search, composed of parallel self-terminating searches of clumps of individual items, easily predicts display-size functions of the sort observed in Figures 2-5: relatively parallel negative and positive curves, adopting a 2:1 ratio as display size increases. A possible additional point in favor of this sort of model is that it is not clear (at least to me) that there is any other class of models that can provide any natural account of the totality of the data.

The relationship between the molar search postulated here and saccadic eye movements may not be a simple one. Although no objective evidence will be offered on this point, subjects certainly do make eye movements in the course of searching the larger displays in the tasks reported here. It is quite possible, though, that eye movements tend to occur (and thus presumably contribute a bit to the estimates of shift time) even if the postulated limit on clump size has nothing to do with acuity. This limit may stem from attentional rather than perceptual factors, but subjects may simply adopt the strategy of moving their eyes between clumps (given that an attentional shift is to be executed, an eye movement takes little time and improves acuity, and the impulse to execute it may be difficult to override). This kind of account is suggested by Treisman's (1982) finding that spatial density has little effect on conjunction search RTs. If Treisman's finding is correct, the results might not change much if eye movements could be prevented without dramatically changing accuracy. The technical problem is that brief displays would probably make reasonable accuracy impossible, and it is not clear how to induce subjects to refrain from eye movements if the display remains available. In summary, then, the present conclusions should not be taken to argue that the serial self-terminating pattern evident with large displays in conjunction search

is merely due to eye movements. The available data suggest that it is probably *not* due to eye movements, in any interesting sense.

The predictions of the model sketched here would not change if it turned out that subjects were able to restrict their search to just a subset of the target items, namely those sharing one of the target features. In this case, the true clump size in the case just described would be on the order of 4, rather than 8. (To speak of "restricting the search" is obviously imprecise, because logically the entire display must in some sense be searched to find the correct subset; the point is that, under some conditions, the time-consuming aspects of the processing might be restricted to a subset of the display.) Egeth et al. (1984) reported results they interpreted as favoring this possibility—specifically, that subjects ordinarily detect color/form conjunctions by searching only the items sharing the target color. One finding was that when every display contained only 3 items of the target color, and subjects were instructed to attend only to items of that color, RTs did not increase when distractors that did *not* share this target color were added to the display.

The results of an experiment conducted with the present stimuli and procedures, however, do not support this conclusion for the present experiments. Twenty-four subjects searched displays of 2-8 items similar to those of Experiment 1. The only difference was that display size and number of items sharing the target color was unconfounded: at each display size, there were two possibilities for the number of same-color distractors that might be present. For instance, with display size 4, either 1 or 3 items in the display shared the color of the target. Overall positive and negative slopes again showed a ratio close to 1:1. The results of the manipulation of number of target-color distractors were not totally straightforward, but they can be fairly summarized by saying that it was the overall number of distractors, rather than the number of items sharing the target color, that produced increases in RTs. Under the present conditions, then, it does not appear that subjects are searching only the items of the target color, even though, as Egeth et al. (1984) have shown, such a strategy can be adopted in response to instructions. The results just described are perfectly consistent, though, with the hypothesis that subjects split in adopting the strategy of either searching items with the color of the target or searching items with the form of the target. (Egeth et al. have also shown that either of these strategies can be adopted upon instruction.) And, of course, the results just described are also consistent with the possibility that subjects are simply searching all the distractors in the same way.

Clearly, then, uncertainty remains about the precise operation of conjunction search. However, the present results speak directly against the item-by-item serial processing hypothesis, and strongly suggest that large displays are searched in clumps composed of a number of objects. This conclusion appears to raise serious difficul-

ties for the feature integration theory as it stands. According to that theory, it is the operation of a single mechanism (the "spotlight" of focal attention) that accomplishes both the localization of individual features and the perceptual conjoining of multiple features of a single object. Furthermore, illusory conjunctions are supposed to occur just when this single mechanism cannot be focused on particular objects (Treisman & Schmidt, 1982). It is not clear, given the present results, that a unified explanation of this sort will remain viable. Thus, it remains to be seen whether major modifications of the feature integration theory can accommodate the present results, or whether a radically different alternative must be developed.

REFERENCES

- EGETH, H., VIRZI, R., & GARBART, H. (1984). Searching for conjunctively defined targets. *Journal of Experimental Psychology: Human Perception & Performance*, **10**, 32-39.
- HOLMGREN, J., JUOLA, J., & ATKINSON, R. (1974). Response latency in visual search with redundancy in the visual display. *Perception & Psychophysics*, **16**, 123-138.
- HOUCK, M. R., & HOFFMAN, J. E. (1986). Conjunction of color and form without attention: Evidence from an orientation-contingent color aftereffect. *Journal of Experimental Psychology: Human Perception & Performance*, **12**, 186-199.
- PASHLER, H., & BADGIO, P. (1985). Visual attention and stimulus identification. *Journal of Experimental Psychology: Human Perception & Performance*, **11**, 105-121.
- TOWNSEND, J. T., & ASHBY, G. (1983). *Stochastic modeling of elementary psychological processes*. Cambridge, England: Cambridge University Press.
- TREISMAN, A. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception & Performance*, **8**, 194-214.
- TREISMAN, A. (1985). Preattentive processing in vision. *Computer Vision, Graphics & Image Processing*, **31**, 156-177.
- TREISMAN, A., & GELADE, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, **12**, 97-136.
- TREISMAN, A., & SCHMIDT, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, **14**, 107-141.

(Manuscript received July 7, 1986; revision accepted for publication November 21, 1986.)