

Visual search in a forced-choice paradigm*

JOHN E. HOLMGREN

University of Maryland, College Park, Maryland 20742

The processing of visual information was investigated in the context of two visual search tasks. The first was a forced-choice task in which one of two alternative letters appeared in a visual display of from one to five letters. The second task included trials on which neither of the two alternatives was present in the display. Search rates were estimated from the slopes of best linear fits to response latencies plotted as a function of the number of items in the visual display. These rates were found to be much slower than those estimated in yes-no search tasks. This result was interpreted as indicating that the processes underlying visual search in yes-no and forced-choice tasks are not the same.

In recent years, a number of investigators have used a variation of Sternberg's (1966) memory scanning paradigm to study the processing of visual information (e.g., Atkinson, Holmgren, & Juola, 1969; Cavanagh & Chase, 1971; Holmgren, 1974). In this variation of the Sternberg task, a small number of items in a visual display is briefly presented to the S. The S must then determine as rapidly as possible whether or not a particular predesignated critical item is present in the display. The dependent variable of primary interest is response latency, the error rates generally being quite low. This task can be considered a type of yes-no task in that the S must make a positive or negative decision as to the presence of the critical item.

In contrast to the above studies, another line of investigation of visual processing has been based on a forced-choice task introduced by Estes and Taylor (1966). In this task, a S must decide which of two predesignated critical items is present in a visual display; each display contains one of the critical items (but never both). In most of the forced-choice studies, the dependent variable of primary interest is the probability of a correct response; the exposure times of the visual displays are lower and consequently the error rates are generally higher than in the yes-no studies.

These two separate lines of investigation have led to rather different views of the processes involved in this type of visual information processing. On the one hand, response latencies in the yes-no studies have been accounted for in terms of serial or fixed-capacity parallel search processes (e.g., Atkinson et al, 1969). In contrast, the forced-choice studies have led to the view that there is no limit on processing capacity and that performance is determined primarily by the

confusability of the critical items with other items in the visual display (e.g., Eriksen & Lappin, 1967; Estes, 1972; Shiffrin & Gardner, 1972). It may be that the two approaches have led to different conclusions because they are tapping different processing stages. For example, the forced-choice studies may be sensitive to early perceptual stages of processing, while the yes-no studies may be more sensitive to some later decision stage. On the other hand, it is possible that because of the differences in the two tasks, the actual processes employed by Ss differ in the two situations.

The purpose of the studies reported here is to investigate response latencies in a forced-choice search task under conditions as similar as possible to those used in a typical yes-no task—that of Atkinson et al (1969). If distinct differences are observed between the pattern of response latencies in the forced-choice and yes-no tasks, it is reasonable to conclude that in some sense information is not processed in the same way in the two situations. Moreover, it may be possible to get some insight into the source of those differences.

In the Atkinson et al study, Ss were presented on each trial with a single consonant, which served as the critical item for that trial. A display containing a horizontal row of from one to five consonants was then presented for 400 msec. The Ss' task was to determine as quickly as possible whether or not the critical letter was present in the display. If the critical letter was in the display, Ss made a positive response by hitting a response key; otherwise, a negative response was made by hitting a second key. The data of primary interest were the latencies of correct responses as a function of the number of letters in the visual display (the error rates were negligible). Mean latencies for both positive and negative responses were found to increase roughly linearly with display size. This linearity was interpreted by Atkinson et al as an indication of a serial or fixed-capacity parallel search process. Moreover, when best-fitting lines were fitted to positive and negative latencies as a function of display size, the slopes were found to be nearly equal.

*This work is based upon a portion of a doctoral dissertation submitted to Stanford University. The research was carried out under a National Institute of Mental Health Terminal Graduate Fellowship and was supported in part by National Aeronautics and Space Administration Grant NGR-05-020-244.

For positive latencies, the slope of the best-fitting line was 23.8 msec and the intercept was 444 msec; for negative latencies, the slope was 26.2 msec and the intercept was 474 msec. The simplest interpretation of the above slope equality is that processing is "exhaustive," i.e., on both positive and negative trials, every item in the display is compared with the critical item. This is in contrast to "self-terminating" processing, in which, on positive trials, the response is initiated as soon as a match is detected between the critical item and an item in the display.

EXPERIMENT I

In this study, the apparatus, the visual displays, and the exposure time and intensity of the displays were identical to those used by Atkinson et al (1969). The one difference is that the present task was a two-alternative forced-choice task while that of Atkinson et al was yes-no. It should thus be possible to directly compare results from the two tasks in order to determine if they are comparable with respect to the underlying processes involved.

Method

Subjects. The Ss were five right-handed female undergraduates obtained through the student employment bureau. They all had normal (corrected) vision and none of them wore contact lenses. They each received \$1.75 for each of the nine sessions in this study.

Apparatus. Visual displays were presented to Ss by means of an Iconix automated two-field tachistoscope. The S was seated at a station consisting of a display panel and a response panel. At the bottom of the display panel were mounted two IEE Binaview units, on each of which could be presented any one of a full set of alphanumeric characters. Above these two units was mounted a ground-glass rear-projection screen. Behind the screen was a black metal plate bearing three small colored lights and a large circular aperture, none of which were visible unless illuminated. The response panel consisted of a board on which were mounted three Johnson code transmitting keys. The centers of the keys fell at 3.8-cm intervals along an arc extending from 290 to 340 deg.

Displays were projected onto the ground-glass screen through the large circular aperture, providing an illuminated area 5.2 cm in diam. Stimuli were displayed by means of a random-access slide projector modified for use with a special light source (Sylvania glow modulator, No. R1131C). A second projector, optically identical to the first but holding a single slide, served to illuminate the screen between stimulus exposures.

Stimuli. The stimulus materials for this study were 95 slides, each consisting of a horizontal row of from one to five capital letters plus two dollar signs. The dollar signs were the first and last symbols in the row and were included in an attempt to equate the letters in the display for any possible lateral masking effects. Twenty slides were prepared for each of the five display sizes. Letters appearing in a display were selected at random from the set of 20 consonants with the restrictions that no letter could appear more than once in a display, and at every display size each consonant had to appear exactly once in every possible display location. Due to the limited capacity of the slide magazines, one slide had to be eliminated from each of the five display sizes. The five slides were picked randomly with the restriction that no letter appear more than once in the five eliminated slides. When projected, the displays were 0.5 cm high and varied in length from 1.0 cm, for single-letter displays, to 2.2 cm, for five-letter displays. At a viewing distance of about 60 cm, this provided a visual angle of approximately 2.1 deg for five-letter displays.

Procedure. Each of the 240 trials in a session consisted of the following sequence of events: (1) A pair of consonants (hereafter referred to as the "critical letters") appeared on the two Binaview units for 2 sec. (2) Following the offset of the Binaviews and while holding down the center of the three keys, the S pressed a button held in her left hand in order to turn on a display. (3) Exactly .5 sec after the S pressed the button, a display containing from one to five consonants appeared in the circular aperture for 400 msec.¹ One and only one of the critical letters appeared in the display. Simultaneous with the onset of the display, two latency counters were started. (4) The S was required to respond within 2 sec of the display onset. The S was instructed to hit the left response key if the letter that appeared on the left Binaview (the "left critical" letter) was present in the display, or the right response key if the right critical letter was in the display. The release of the center key stopped one of the latency counters, while hitting either of the other keys stopped the second counter. The center key was held down and responses were made with the right hand. (5) Following the 2-sec response interval, one of three feedback lights came on. A green light indicated a correct response, red indicated an error, and white indicated no response within the allotted time. The feedback light remained on for 2 sec and was followed by a .5-sec blank period, after which the next trial began. The total duration of a trial was 7 sec plus the time from the offset of the Binaviews to the button press by the S.

The projector which illuminated the screen between stimulus exposures contained a slide with four dots arranged in a rectangle. When projected, this rectangle measured 2.2 x 1.3 cm. The display always appeared centered within this rectangle. The brightness of the pre- and postexposure field was 6.6 fL, as measured by a Macbeth illuminometer; the display brightness was 7.4 fL.

Of the pair of consonants shown at the beginning of each trial, one was chosen randomly from the set of letters not appearing in the subsequent display; the other was chosen randomly from the set of displayed consonants, subject to the restriction that within each display size, each serial position be tested equally often during the session. The critical letter appearing in the display was presented on the left or right Binaview with equal probability. The slide shown on a trial was selected according to a random schedule, with the restrictions that each display size be presented equally often during each session and that no slide be presented twice in succession. Each S received a new randomization for each session of the experiment.

Each session consisted of 10 warm-up trials followed by 240 data trials. Ss received a short break (about 3 min) midway through each session. Every S ran for nine sessions, either one or two sessions a day, with at least 1 h between successive sessions.

Instructions to the S before the first session were given verbally and consisted essentially of a description of the five trial events described above. In addition, it was emphasized both during the initial instruction period and at the beginning of each subsequent session that Ss should respond as fast as possible while making as few errors as possible. Ss were told that after the first session, which was a practice period intended to familiarize them with the task, they should try to make fewer than 12 errors per session.

Results and Discussion

Before presenting the data, it is necessary to introduce some terminology. Trials on which the left critical letter appears in the subsequent display will be referred to as the "left-critical" trials; the remaining trials will be referred to as "right-critical." The time from the onset of the display to the release of the center key by the S will be referred to as the "release latency." The time from the onset of the display to the pressing of either response key will be called the "terminal latency." Unless otherwise noted, all latencies reported below are terminal latencies.

Figure 1 shows terminal latencies for correct responses as a function of display size averaged across Ss. The data are grouped over Sessions 4-9, the number of observations per point averaging 141. Sessions 1-3 were eliminated, since mean latency declined substantially over these sessions. Also shown here are average error rates at each display size. These error rates were under 5% at every display size. Also shown in Fig. 1 are serial position curves for left-critical and right-critical trials. These curves represent response times for each display size as a function of the location of the critical letter in the display. Note that for right-critical trials, the serial position curves have been aligned on the right rather than the left. That alignment better reflects the tendency of right-critical latencies to increase as serial position of the target moves from right to left in the display.

Using the method of least squares, the best-fitting linear function was found for each of the curves relating display size to response latency for each S. Table 1 gives the mean and standard error of the slopes and intercepts of these lines for each type of trial. The means are of course equivalent to the slopes and intercepts of the lines providing the best fit to the curves in Fig. 1. A trend analysis on the Fig. 1 data shows no significant deviation from linearity. Linear trend accounts for nearly 99% of the variance in these curves. Neither the intercept nor the slope of the left-critical line differs significantly from the corresponding right-critical value ($p > .05$).²

In order to obtain a measure of the general trend in serial position curves for each S, best fitting lines were found for each S for each of the serial position curves associated with Display Sizes 2 through 5. The slopes of these lines were then averaged for both left and right critical trials, resulting in two values for each S. The mean and standard error of these values are given in Table 1. The means differ significantly from zero and from one another ($p < .01$). More will be said about this result below.

All of the above results pertain to terminal latencies. The release data are not presented here for two reasons. First, curves based upon release latencies show the same trends as the terminal curves, except that the slopes and intercepts of the best-fitting lines

Table 1
Summary of Best Fits to Latencies as a Function of Display Size and Serial Position for Experiment I
(Means \pm Standard Errors in Milliseconds)

Best-Fit Statistics	Trial Type	
	Left-Critical	Right-Critical
Slope of Display Size Function	54.5 \pm 9.0	54.2 \pm 6.9
Intercept of Display Size Function	447.4 \pm 22.6	485.6 \pm 30.6
Slope of Serial Position Function	64.5 \pm 7.6	-30.2 \pm 8.8

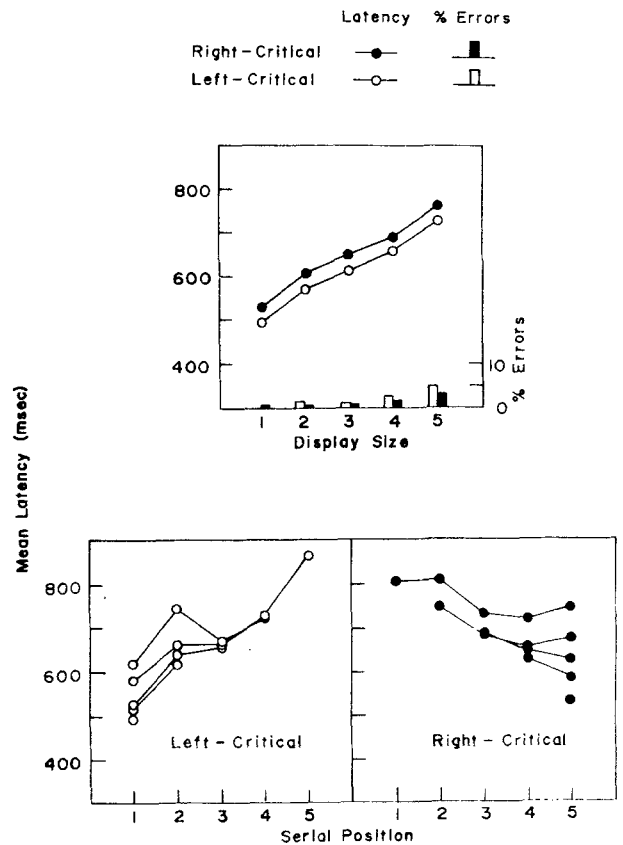


Fig. 1. Overall mean latencies and error rates as a function of display size; serial position curves for left-critical and right-critical trials (Experiment I).

are generally smaller for the former. Second, it is likely that the smaller release slopes are the result of premature releases. That is, on some trials, Ss probably released the center key before determining which of the two critical letters was in the display. If the mean release latency on trials with premature releases showed a smaller increase with increasing display size than did mean latencies on the remaining trials, the slope of the release curves would be spuriously low. Since the low error rates indicate few premature terminal responses, the terminal latencies more accurately reflect the underlying search processes.

It is of interest to compare the results of this study with those of the Atkinson et al experiment since both studies were run under exactly the same conditions, using the same stimuli and apparatus. While the intercepts in the two studies were very similar, the average slope of the forced-choice curves (54.3 msec) is significantly greater than either the positive slope (23.8 msec) or negative slope (26.2 msec) found in the yes-no task ($p < .001$). This indicates that the two tasks were not equivalent with respect to the manner in which Ss processed displays. Ss apparently did not completely ignore one of the critical letters on each trial, a strategy that would have converted the task from forced-choice to yes-no.

In terms of the types of models that have been used to account for response latencies in yes-no search tasks, the simplest way of explaining the difference in slopes between the present study and that of Atkinson et al is to assume that Ss in the forced-choice task carry out an exhaustive search for *both* critical letters, despite the fact that a search for only one would give them sufficient information for a correct response. Assuming that the search rate for a single critical item is about the same in the two studies, this leads to the prediction that the forced choice slopes should be about twice the yes-no slopes—a prediction which is born out by the data.

The problem with the above explanation is that it cannot handle the large serial position effects observed in the present study. If processing were exhaustive, then the serial position curves should be flat, since the expected processing time for a given display size would be independent of the location of the critical letter in the display. However, it may be that the serial position effects are due to some other factor, such as stimulus-response compatibility. The fact that the average slope of the left-critical serial position curves (as given in Table 1) has the opposite sign of the corresponding right-critical slope suggests the possibility that when the position of the critical letter in the display conflicts with the position of the appropriate response key, the response is slowed due to the lack of stimulus-response compatibility. For instance, a critical letter at the far right of a five-letter display would result in a highly compatible stimulus-response relation on right-critical trials and a highly incompatible relation on left-critical trials.

EXPERIMENT II

The purpose of this study is to provide information as to whether or not the exhaustive processing hypothesis mentioned above is tenable for the forced-choice task. On a portion of the trials in this study, neither critical letter appears in the display. The addition of these "negative" trials to the "positive" trials (i.e., the left- and right-critical trials) of the previous study changes the forced-choice task into what might be called a "yes-no with recognition" task. That is, the S must not only make a yes-no decision as to the presence or absence of a critical letter, but when one is present must also determine which critical letter it is. However, to avoid cumbersome terminology, this task will be referred to hereafter simply as the "three-choice" task.

If processing in the forced-choice task is exhaustive, then the addition of these negative trials should require no additional processing of the visual display. By searching the display for both critical items, Ss would acquire sufficient information to respond correctly on negative trials. This leads to the prediction that when response latency is plotted as a function of display size, the slope of the functions for right- and left-critical trials in the three-choice task should be

the same as in the forced-choice task. Also, the slope for negative trials should be the same as that for right- and left-critical trials. Note that these predictions apply only to the slopes of the latency-display size functions and not to the intercepts. The addition of a third response to the task may well affect the intercepts, perhaps by increasing the duration of response selection or execution subsequent to processing of the visual display.

Method

Subjects. The Ss for this study were six female undergraduates. As in Experiment I, they all had normal (corrected) vision, none wore contact lenses, and they were each paid \$1.75 for each of the nine sessions.

Apparatus. The apparatus used here was the same as that used in Experiment I, with one modification. A fourth key was added to the response panel. This new key was placed inside the arc formed by the other three keys, with its center lying 3.8 cm from the center of the middle key in the arc.

Procedure. Each trial in this study consisted of the same sequence of events as in Experiment I. However, in this study, each trial had three possible outcomes. One-third of the trials were left-critical, one-third were right-critical, and on the remaining third, neither critical letter appeared in the display. On these negative trials, the correct response was made by hitting the additional key on the response panel. The displays were the same ones used in Experiment I.

Instead of turning on a green light for correct and a red for error, feedback was given by turning on a light in the upper left of the display panel for left-critical, one in the upper right for right-critical, and one directly above the circular aperture for negative trials. All three of these lights were green. As in the previous study, the S received a white light if she did not respond within the 2-sec response interval. Also as in the previous study, each session consisted of 10 warm-up trials followed by 240 data trials. Ss received a short break halfway through the session. Again, it was emphasized at the beginning of each session that Ss should respond as fast as possible while making as few errors as possible.

Results

Latencies of correct responses were averaged across Ss and are shown as a function of display size in Fig. 2, along with serial position curves for positive trials. Corresponding points on the two positive curves were averaged and a test for trend was carried out on the resultant values. The test showed a significant linear trend in the average curve but also significant deviation from linearity ($p < .01$ in both cases). The curve for negative trials also has a significant linear component ($p < .01$), but no significant deviation from linearity.²

The error rates in this study were quite a bit higher than those in the forced-choice task. A summary of the error data is given in Table 2. The most notable finding here is that when Ss make an error on positive trials, they are much more likely to hit the negative key than the other positive key. This is particularly true for the larger display sizes. These "false negatives" suggest that Ss may, on some trials, make a negative response before processing all items in the display. If these false negative responses occurred on trials that would otherwise have had relatively long latencies (providing the S had not responded

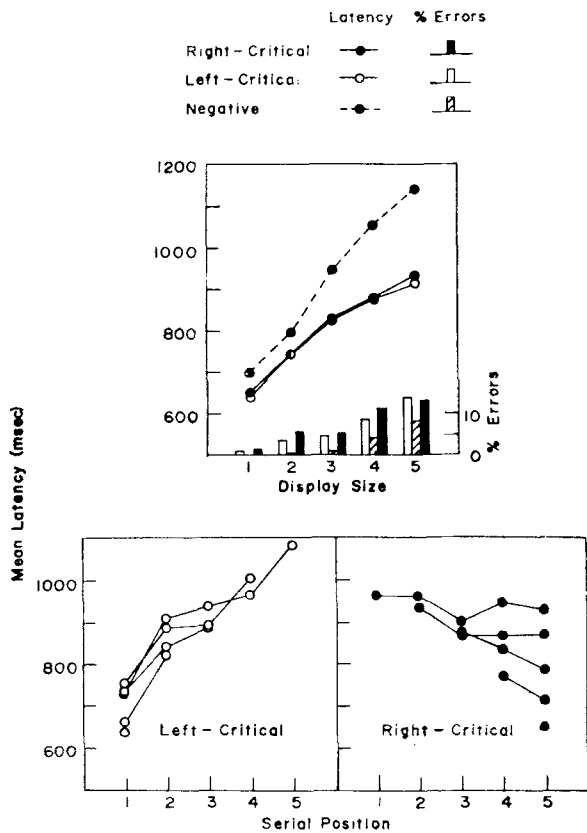


Fig. 2. Overall mean latencies and error rates as a function of display size; serial position curves for left-critical and right-critical trials (Experiment II).

prematurely), the slight negative acceleration in the positive curves could be accounted for.

Despite the departures from linearity in the data, best-fitting linear functions were obtained for all curves in Fig. 2. Table 3 gives the mean slope and intercept and also the standard error of the mean for each type of trial. Because of the high error rates at the two largest display sizes, lines were also fit to only the first three points on each curve. The data for these fits are also given in Table 3. There are no significant differences among the mean intercepts for the three trial types. Also, the two positive slopes do not differ

Table 2
Proportions of Incorrect Responses for Left-Critical (L), Right-Critical (R), and Negative (N) Trials Averaged Across Subjects (Experiment II)

Trial Type	Resp	Display Size				
		1	2	3	4	5
L	R	.002	.005	.011	.016	.029
	N	.007	.031	.034	.067	.102
R	L	.005	.018	.013	.022	.027
	N	.001	.034	.036	.086	.098
N	L	.002	.004	.004	.024	.038
	R	.000	.002	.007	.016	.043

significantly. However, the mean negative slope is significantly greater than the average of the two mean positive slopes ($p < .05$). All of these results apply both to fits based on all display sizes and to those based on only the first three.

As in Experiment I, a measure was obtained of the general trend in the two sets of serial position curves for each S. The measures were based upon Display Sizes 2-5. Measures were also obtained using only Display Sizes 2 and 3. The mean and standard error of the mean for the two sets of measures are given in Table 3. For both measures, the left-critical and right-critical serial position trends differ significantly from one another and from zero ($p < .05$ in all cases).

A number of comparisons were made between Experiments I and II. Using the three-choice slope estimates based upon all display sizes, the difference between the average three-choice positive slope and forced-choice positive slope is not significant ($p > .05$), while there is a significant difference ($p < .01$) between the forced-choice slope and the three-choice negative slope. When the three-choice estimates are based upon only the first three display sizes, both the positive and negative slopes differ significantly from the forced-choice slope ($p < .05$ and $p < .01$, respectively). The difference between positive intercepts in the two studies is significant at the .05 level regardless of which estimates are used. Finally, average trends in corresponding sets of serial position curves show no significant differences across the two studies; again, this holds for both sets of estimates.

Table 3
Summary of Best Fits to Latencies as a Function of Display Size and Serial Position for Experiment II*

Best-Fit Statistics	Trial Type		
	Left-Critical	Right-Critical	Negative
Slope of Display Size Function	67.8 ± 6.3 (91.4 ± 12.3)	70.2 ± 5.1 (89.5 ± 4.7)	113.1 ± 14.0 (121.8 ± 20.4)
Intercept of Display Size Function	594.5 ± 23.9 (551.7 ± 21.4)	596.0 ± 37.6 (561.4 ± 35.4)	588.0 ± 22.3 (571.0 ± 21.7)
Slope of Serial Position Function	96.6 ± 24.0 (117.3 ± 29.7)	-32.6 ± 8.8 (-50.7 ± 14.3)	--

Note—Means ± standard errors in milliseconds.
*Values in parentheses are based on fits to Display Sizes 1-3 only.

DISCUSSION

Recall that if Ss in Experiment I were employing an exhaustive search process (i.e., searching the whole display for both critical items), then the slope of the latency-display size function for right- and left-critical trials should be unaffected by the addition of negative trials. Moreover, the slope of the latency-display size function for negative trials should be the same as that for right- and left-critical trials. As the results show, these predictions were generally not born out by the data in the present study. Of the various relevant comparisons, the only one that was not significant was the difference of 14.7 msec between the average of the left- and right-critical slopes in Experiments I and II, using the slope estimates based on all five display sizes.

Thus, the data of Experiment II support the view that Ss were not processing exhaustively in Experiment I. The fact that the slopes of the latency-display size functions in the forced-choice task were twice the slopes found by Atkinson et al (1969) in the yes-no task cannot be explained by simply assuming that the forced-choice task involves one exhaustive search for each critical item. The slope differences must be due to more fundamental differences in the underlying processes, although these data are not sufficient to determine exactly what those differences are.

In summary, even when a forced-choice study is run under conditions as similar as possible to those used in a yes-no study, there are substantial differences in the resulting data, indicating that Ss do not process visual information in the same way in the two tasks. Thus, differences in the conclusions about visual processing drawn from forced-choice and yes-no studies may be due to the fact that the two approaches

are studying different processes, not just different stages of the same basic process.

REFERENCES

- ATKINSON, R. C., HOLMGREN, J. E., & JUOLA, J. F. Processing time as influenced by the number of elements in a visual display. *Perception & Psychophysics*, 1969, **9**, 321-326.
- CAVANAGH, J. P., & CHASE, W. G. The equivalence of target and nontarget processing in visual search. *Perception & Psychophysics*, 1971, **9**, 493-495.
- ERIKSEN, C. W., & LAPPIN, J. S. Independence in the perception of simultaneously presented forms at brief durations. *Journal of Experimental Psychology*, 1967, **73**, 468-474.
- ESTES, W. K. Interactions of signal and background variables in visual processing. *Perception & Psychophysics*, 1972, **12**, 278-286.
- ESTES, W. K., & TAYLOR, H. A. Visual detection in relation to display size and redundancy of critical elements. *Perception & Psychophysics*, 1966, **1**, 9-16.
- HOLMGREN, J. E. The effect of a visual indicator on rate of visual search: Evidence for processing control. *Perception & Psychophysics*, 1974, **15**, 544-550.
- SHIFFRIN, R., & GARDNER, G. T. Visual processing capacity and attention control. *Journal of Experimental Psychology*, 1972, **93**, 72-82.
- STERNBERG, S. High-speed scanning in human memory. *Science*, 1966, **153**, 652-654.

NOTES

1. While an exposure time of 400 msec does not guarantee that Ss will be limited to a single fixation on the display, previous work with this apparatus has shown that results obtained with this exposure time show no substantive differences from those obtained with an exposure time of 150 msec (Atkinson, Holmgren, & Juola, 1969).

2. All statistical results reported for this study are based upon analysis of variance or, where appropriate, simple t tests for independent or correlated means. In all analyses, between-S differences are treated as error variance.

(Received for publication March 15, 1974;
revision received May 14, 1974.)