Types and frequency of errors in visual search*

MARIANNE W. KRISTOFFERSON

Psychiatry Department, McMaster University, Hamilton, Ontario, Canada

The experimental method allowed false-positive as well as false-negative errors to be measured during visual search. Data were collected from eight Ss on four target-set size conditions on each of 25 days in order to determine whether the finding that search rate is independent of target-set size after practice might be attributable to a systematic variation in total error rate with set size. Even after practice, these Ss continued to search significantly faster for a single target than for target sets which were composed of five or more characters. Further, the incidence of total errors was much lower than in those studies in which search rate had been found to be independent of set size (i.e., parallel processing), even though these latter studies considered only the occurrence of false-negative errors. Additional analysis strongly suggests that: (1) attainment of parallel processing and a highly accurate performance are incompatible, and (2) verbal instructions alone are not sufficient to control the accuracy levels.

Neisser, Novick, and Lazar (1963) have shown that after practice and when Ss perform with a low degree of accuracy, rate of visual scanning when searching for any one of 10 targets is as fast as when searching for any one of 5 targets or for only 1 target.

By comparing mean number of failures to find the target per day for each target-set size condition, they were able to reject the possibility that Ss were equalizing their speed in the different target-set size conditions by varying their error rates. This error analysis was based only on the frequency of occurrence of false-negative errors (failure to find the target). It is, of course, also possible to make false-positive errors in performing the search task; but the task as originally designed did not allow for an accurate and objective tabulation of errors of this type. In the Neisser et al (1963) experiment, the only mention of false-positive errors was "...Errors of commission also occurred, but rarely [p. 957]." It is not clear how the E was made aware of errors of this type, but, at best, the S, after turning a switch to indicate that he had visually located the target, may have subsequently named the target and/or indicated verbally its position on the list. At worst, the authors concluded that these errors occurred infrequently because only rarely did a S spontaneously offer the information that, after indicating that he had located the target, he had rechecked and discovered that he had responded to a nontarget character. Looking at only the best of these two alternatives, the designation of the

target was made after the scanning time was recorded, and there was therefore no control for the possibility that the S might inadvertently or even deliberately indicate that he saw the target when he did not and then, after scanning the list further for a few seconds, locate the target and identify it to the E. This appears to be a serious flaw in the design of the experiment, since it leaves open the possibility that Ss were equalizing their speed in the four conditions by varying their error rates.

The present experiment was designed to replicate the Neisser et al (1963) study and to allow for measurement of both false-positive and false-negative types of error. If the main results of the original study were replicated, it would then be possible to determine whether equalization of speed over the experimental conditions could be attributable to a systematic variation in error rate with target-set size or whether this possibility could be ruled out.

METHOD

Stimulus Materials

Stimulus materials were prepared as described by Neisser et al (1963), except for some minor details of the printout format.

Lists of 50 alphanumeric items composed of six characters each were internally generated by an IBM 6400 computer and automatically printed out in single columns, six rows to the inch, on pages of unlined output paper, with eight 50-item lists on each page. The printout was repeated to provide identical sets of stimulus materials for each S participating in the experiment.

Items in the lists were constructed by random selection from a set of 20 context characters: BCGIJNOQRSTV

WXY13567, with the restriction that no character be repeated within a single item. Within each list, a single item was randomly selected to contain the target character and one of the characters making up that item was randomly chosen to be replaced by an appropriate target character. The same target position was never repeated within the lists making up the stimulus material for a given day and a given condition. The four target-set conditions were K, AFKU9, HMPZ4, and AFKU9HMPZ4 ("TEN"). On all trials of Condition K, the target was the letter "K." On trials in Condition AFKU9, the target was equally likely to be any one of those five characters. Similarly, in Condition HMPZ4, the target was equally likely to be any one of those five characters. In Condition TEN, the target was equally likely to be any one of the 10 characters making up that set.

Apparatus and Procedure

A box with a sloping top and a rectangular opening was used to display the stimulus materials. The opening, which was slightly wider and longer than a single 50-item stimulus list, was covered with a metal shield which could be released to expose the entire list by pressing a button positioned on a ledge at the base of the face of the apparatus. To initiate a trial, the S pressed the button with his nonpreferred hand, which caused the shield to drop, and simultaneously activated a Standard Electric timer. Also on this shelf was a holder for a removable metal stylus which housed an ordinary felt-tipped marking pen. When the stylus was lifted from the holder, the timer was stopped and simultaneously a second timer was started. The S was instructed to position the fingers of his preferred hand on the pen prior to initiating a trial, to lift the pen as soon as he saw the target, and to mark the target immediately with the pen. When the pen made contact with the stimulus list, it terminated the second timer and left a mark on the list.

After each trial, the search time and the time required to move the stylus from its holder to mark the target were each recorded, along with the number of the line on which the target occurred. The shield was then raised by the E, the next stimulus list was positioned in the display opening, and the S was given an auditory ready signal to indicate that he could initiate the next trial.

If the S failed to find the target after scanning through the list (cf. instructions to Ss), no response times were recorded, that trial was tallied as a false-negative error, and an additional list was presented. Similarly,

*This study was supported by a grant from the Alcoholism and Drug Addiction Research Foundation of Ontario, Ontario, Canada.

Fig. 1. Mean time per item scanned as a function of experimental days (1-25) for each of the four experimental conditions.

if a false-positive error was made, that search time was not used and a new list was presented as a replacement. A false-positive error was recorded on a trial if any one of the following occurred: (1) the S lifted the stylus from the holder, indicating that he had located the target, and then failed to mark the target; (2) the S lifted the stylus from the holder and there was a delay of 3 sec or more before he marked the target; or (3) the S marked the target within this time limit, but the mark was placed three or more lines above or below the item containing the target.

Subjects

Eight male McMaster University student volunteers who were naive to the task served as paid Ss in the experiment. The Ss were run individually.

Experimental Design

In each successive daily session (weekends excluded), each S performed on each of the four experimental conditions (K, AFKU9, HMPZ4, and "TEN") until he produced, in addition to 3 practice trials for each condition, 12 errorless trials per condition, for which search times were obtained to allow for the calculation of scanning rate. The order of conditions was randomized over

days. Every S participated for at least 25 days in the main experiment.

Following the main experiment, data were collected daily from each S under the same set of conditions and procedures until there appeared to be no change with practice in a given S's scanning rate for 6 days on the basis of visual inspection of the data. This was done for purposes unrelated to the present study. The actual number of days over which data were collected from each S were: one S, 26 days; one S, 30 days; two Ss, 31 days; one S, 36 days; one S, 38 days; one S, 46 days; and one S, 48 days.

The Ss were instructed in the following manner: "This experiment is concerned with visual search. You will be searching through lists of alphanumerical characters to locate any of the target characters which are appropriate for that trial or condition. For example, you might be told to search the list for any of the following characters: AFKU9—only one of these characters will be present in any one list, and its position will be unpredictable from one list to the next. There is a target on every list, and its position is randomly chosen. We will be measuring the speed with which you can locate the target."

The procedure and the apparatus were then explained to the S, and the instructions continued in this way:

"As soon as the shield drops, start searching through the list. Always search from the top down. Try not to recheck preceding lines. You are allowed to search through the list only one time. If you come to the bottom of the list and have not located the target, tell me and I will set up a new trial. Do not be disturbed if you miss a number of targets. As soon as you do find a target, lift the pen from its holder and mark the target with it. Never lift the pen until you have seen the target. Speed is important—you should scan the lists and mark the targets as quickly as possible. You do not have to be highly accurate in marking the target, but the mark should either be on the line with the target character or on the line immediately above or below it. For this group of trials the character(s) making up the target set is (are)

A card with the current target-set character(s) typed on it was attached to the apparatus at the S's eye level and to the right of the display window so that the target set was always visually available to the S.

RESULTS

Analysis of Data

A single measure of the time per item scanned was calculated from each set of 12 search times and their related target positions by the method described by Neisser (1963). Equations for the least-squares linear function relating response latencies and target positions were calculated for each S, each set-size condition, and each day separately. Isolated points which deviated markedly from the others were deleted by the method described by Neisser (1963). Slope values obtained from these equations, which represent the time required to scan over each nontarget item independently of response factors, were combined and averaged over the eight Ss for each day and each set-size condition separately.

In Fig. 1, mean time per item scanned in seconds averaged over all Ss is shown plotted against days (1-25) for each of the four experimental conditions separately. It is apparent that, with practice, scanning rate is increased for all conditions and that the difference in scanning rate between conditions diminishes.

A total of 10,221 test trials were administered to the group of eight Ss over the first 25 experimental days. On 6.07% of these trials, a false-negative or false-positive error occurred. False-negative errors numbered 589, occurring on 5.76% of the trials; false-positive errors totalled 32, occurring on .31% of the trials.

In order to determine whether rate

of scanning differed significantly across target-set size conditions after prolonged practice, an A by B by S ANOVA was performed on slope values obtained on Days 18-25 of the experiment. The main effect of experimental conditions (target-set size) was highly significant: $F(3,21) = 17.495$, $p = .0001$, showing that after practice there remained a significant difference in rate of scanning among the experimental conditions. Application of Duncan's range test to these means revealed that time per item scanned was significantly less for Condition K than for each of the multiple target-set conditions. Further, there were no significant differences in scanning rates among Conditions AFKU9, HMPZ4, and TEN. The main effect of days was also highly significant: $F(7,49) = 7.311$, $p = .0001$. The time per item scanned as calculated from each of the target-set size conditions continued to decrease with practice, there being an insignificant Days by Conditions interaction: $F(21,147) = 1.421$, $p = .1171$.

Over this same period (Experimental Days 18-25), errors of both types occurred on 123 out of the total of 3,195 trials, or on 3.85% of the trials. Of this total percent errors, 3.47% were false-negative errors and .38% were false-positive errors. The percent of trials upon which errors of either sort occurred ranged for individual Ss from 1.29% to 8.57%. The distribution of combined errors across the four target-set size conditions, given in terms of the mean number of errors obtained by a single S over the 8-day period, was as follows: K = 2.875, AFKU9 = 3.625, HMPZ4 = 5.375, and TEN = 3.500. An A by S ANOVA revealed an insignificant difference between incidence of errors over the four experimental conditions: $F(3,21) = 1.761$, $p = .1845$.

DISCUSSION

The initial question addressed by this study, whether, when both types of possible error are accounted for, equalization of search speed over the experimental conditions might be attributable to a systematic variation in error rate with target-set size, is left unanswered by these data, since, even after prolonged practice, Ss never achieved as rapid scanning rates for 5 or 10 targets as for 1. Under the conditions of this experiment, the incidence of false-positive errors was negligible; over the final 8 days of the main experiment, the frequency of occurrence of the combined errors was found not to differ significantly for the target-set size conditions. These results do add some empirical support

to Neisser et al's (1963) assertion that false-positive errors occur only rarely. However, the answer to the initial question requires measurement of both types of error under conditions which result in achievement of equality of scanning rates over all target-set sizes, which in turn is probably dependent upon performance of the task with a higher overall incidence of errors (Neisser, 1967; Wattenbarger, 1968).

Neisser et al (1963) reported a high level of errors throughout their experiment and an increase in errors over the final 14 of the 27 experimental days after they initiated a competition among the Ss. The median S in their experiment failed to find the target on over 20% of the trials over the last portion of the experiment. Averaging over this same period, even their most accurate S failed to find the target on over 14% of the trials and their least accurate S failed to find the target on nearly 25% of the trials. Note that there was no overlap between error rates attained by individual Ss in the Neisser et al (1963) study and in the study reported here.

Neisser (1967) conceives of the cognitive system used in searching as hierarchically organized. He describes character recognition as a multilevel process in which the outcomes of first-level feature analytic operations (e.g., detection of basic distinctive features such as perhaps angularity, curvature, slopes of lines, etc.) feed into higher-order systems, which in turn perform more complex cognitive functions (e.g., identification of letters as such). On the basis of the finding that rate of search is independent of target-set size (Neisser, 1963; Neisser, Novick, et al, 1963), he concluded that at least at the level of feature detection, several operations may occur simultaneously or in parallel.

Based on both empirical data and a consideration of introspective reports obtained from Ss performing the search task, Neisser (1967) suggested that the conditions of the search experiments encourage the S to rely on fragmentary features of the test stimuli rather than on a full identification of the test stimuli when making comparisons with memory representations of target symbols. He had also postulated that the reason Ss participating in the search experiments made so many errors was that "the preattentive mechanisms are not built for accuracy—that is a matter for focal attention. We can expect a certain crudeness whenever responses occur without attentive construction [1967, p. 100]."

Based on Neisser's findings and

theoretical speculations, one might hypothesize that parallel processing and a highly accurate performance are incompatible. This hypothesis has received some empirical support from a study reported by Wattenbarger (1968) in which he manipulated error level. Using the standard search task procedures and experimental design, Wattenbarger collected data over 15 experimental days from each of two groups: a "speed group" and an "accuracy group." The two groups differed only in the verbal instructions given them. Ss in the speed group were encouraged to scan as quickly as possible, while Ss in the accuracy group were encouraged to scan as quickly as possible without making errors. Over the last eight sessions of the experiment, the error rates were 15% for the speed group and 7% for the accuracy group. One conclusion Wattenbarger (1968) drew from these findings was that "the production of the different error levels in the two groups indicates the effectiveness of verbal intimidation as a means to manipulate the error criterion." Over these same sessions, analysis of variance revealed a highly significant difference between search rates for different set sizes for the accuracy group but not for the speed group. Search rates obtained by the accuracy group were slower than for the speed group, and performance continued to improve with practice for both groups.

Wattenbarger (1968) interpreted these results to show that "practice alone is not adequate to produce parallel information processing in a visual search task such as this; a lenient accuracy criterion is necessary as well." Another interpretation consistent with these results is that when Ss perform with greater accuracy, the amount of practice required in order to achieve a search rate independent of target-set size is increased. That is, it may be that the rate at which differences between search times obtained from different target-set sizes diminish with practice is slower when a high error criterion is maintained, but that with sufficient practice evidence of parallel processing would be obtained, even when Ss perform with a high level of accuracy.

One way to test between these two alternative interpretations would be to extend the number of practice sessions until search rates obtained from Ss performing the task with a high level of accuracy were stable and then to determine whether search rates over the final experimental sessions remained a function of target-set size.

The results presented for the main experiment reported here are highly consistent with Wattenbarger's findings for his "accuracy" group,

even though, in the present study, 25 sessions instead of 15 were run. A significant difference in rate of scanning was found over the last 8 sessions between a target set of 1 character and target sets made up of 5 or 10 characters; even with this additional practice, scanning rates were still increasing for all target-set sizes as a function of days. Therefore, there remains the possibility that after additional practice, equality of scanning rates might be found.

As mentioned in the experimental design section of this paper, for each individual S in the group, data were actually collected until there appeared, on the basis of visual inspection, to be stability in that S's search rates over a 6-day period. The number of sessions for individual Ss ranged from 26 to 48.

In order to establish more convincingly whether or not achievement of equality of scanning rates across target-set size conditions and a high level of accuracy are incompatible, the data obtained over the final eight sessions for each S were submitted to an A by B by S ANOVA, first to determine whether time per item scanned had, in fact, stabilized and then, if so, to determine whether there remained a significant difference in rate of scanning across experimental conditions.

In Fig. 2, mean time per item scanned in seconds is plotted against the final 8 days of the supplemental experiment separately for each of the four target-set size conditions. Analysis of variance revealed an insignificant main effect of days: $F(7,49) < 1$. Since the Days by Conditions interaction was also not significant [$F(21,147) = 1.38$, $p = .14$], it may be concluded that scanning rates obtained from each set-size condition were stable over this period. The main effect of experimental conditions remained highly significant: $F(3,21) = 9.21$, $p = .0007$. As was the case for the

Fig. 2. Mean time per item scanned as a function of the final 8 days of the supplemental experiment for each of the four experimental conditions.

previous analysis reported, application of the Duncan range test showed that the scanning rate for Condition K was significantly faster than for each of the other conditions, and no significant differences in scanning rate were found among target-set size conditions of more than one character.

Errors of both types occurred on 4.24% of the total trials over the final 8 days: 3.80% were false-negative errors and .44% were false-positive errors. An A by B by S ANOVA revealed insignificant differences in incidence of errors for the different target-set size conditions: $F(3,21) < 1$. It also showed that error levels were stable over all target-set sizes over this period, there being neither a significant main effect of days [$F(7,49) < 1$] nor a significant Days by Conditions interaction ($F < 1$).

These results show quite clearly that even when practice is extended until search and error rates obtained from each target-set size are stable, evidence of parallel processing of memory and stimulus character comparisons is not obtained when visual search performance is accompanied by a high accuracy level. Thus, the conclusion that achievement of parallel processing and highly accurate performance on the search task are incompatible is supported.

On the other hand, the results of the present study do bring into question the validity and/or generality of Wattenbarger's (1968) conclusions regarding the effectiveness of verbal instructions as a successful way of manipulating the accuracy criterion. In the study reported here, verbal instructions stressing speed failed to produce a high error rate; in fact, these Ss performed with a greater degree of accuracy than Wattenbarger's "accuracy" group. Further, in a recent study by Grill (1971), in which she attempted to manipulate error criterion through verbal instructions patterned after those of Wattenbarger in order to specify some of the conditions under which parallel and serial processing occur, it was found that the speed and accuracy instructions produced no differential effects on latency or errors.

It appears that other unspecified variables in the experimental situation may outweigh the influence of verbal instructions in determining the error criterion. Since level of accuracy does have important effects on the nature of processing revealed by this task, it is important that these variables be identified and brought under experimental control.

REFERENCES

- GRILL, D. P. Variables influencing the mode of processing of complex stimuli. *Perception & Psychophysics*, 1971, 10, 51-57.
- NEISSER, U. Decision-time without reaction-time: Experiments in visual scanning. *American Journal of Psychology*, 1963, 76, 376-385.
- NEISSER, U. *Cognitive psychology*. New York: Appleton-Century-Crofts, 1967.
- NEISSER, U., NOVICK, R., & LAZAR, R. Searching for ten targets simultaneously. *Perceptual & Motor Skills*, 1963, 17, 955-961.
- WATTENBARGER, B. L. Speed and accuracy set in visual search performance. Paper presented at the meeting of the Midwestern Psychological Association, Chicago, 1968.