Context and learning in the processing of information from brief visual displays

H. A. TAYLOR²
DOUGLASS COLLEGE

Research on visual detection with tachistoscopically presented multiple-element displays has indicated that a nonrandom sample of the elements is processed on any trial. The effect of learning on the sample composition was assessed with a task in which S reported whether matrices of nine letters contained one or two Xs. After a first session involving two-signal displays with signal pairs oriented either horizontally or vertically, two groups of nine Ss saw only one signal-pair orientation, while a comparison group continued to see all display types. The group that switched to all-vertical orientation significantly out-improved and out-performed the comparison group on the vertical two-signal displays, demonstrating that the central processing mechanism may adapt to take advantage of display pattern regularities.

When visual displays composed of many discrete, unconnected elements are presented at tachistoscopic durations, not all elements appear to contribute with equal probability to the transfer of information from display to S. Several sources of the inequality may be tentatively identified: Various structural features of the displays and the precise nature of the task set for the S contribute effects that, though perhaps predictable in theory, are complex and effectively idiosyncratic. For example, Estes and Taylor (1966) found, in the context of a visual detection task that required S to report which of two signal letters was present among a set of noise consonants, that the signal B was more easily detected than the signal F.

A second source of differential perceptability doubtless lies in the well known variation of acuity over the visual field (Woodworth & Schlosberg, 1954). When displays occupy an appreciable visual angle, elements that fall nearest the point of fixation will be relatively favored in their probability of being clearly perceived.

Beyond any possible effects of simple acuity, however, the position of an element within a display has been amply demonstrated to influence the probability that the element will be effectively perceived. In a task that required Ss to report visually marked single letters from 2 by 8 arrays, Averbach and Sperling (1961) found that, despite a central fixation point, letters in the upper row of the array all were reported more accurately than their counterparts in the lower row; and letters in the four outer positions, though farthest removed from the fixation point, were reported with greater accuracy than were their immediate neighbors. In a detection experiment in which the displays were 4 by 4 arrays of letters, Estes and Taylor (1966) found that when two redundant signal elements were present, performance was related to the separation of the two signals within the display. In the same study, these investigators found that successive responses to widely spaced repeated presentations of the same displays departed substantially from independence. Holmgren (1968) found that, for an appreciable proportion of Ss, detections of critical elements within 4 by 4 arrays occurred with significantly greater frequency when the critical element was located in the left half of the display. Taken collectively, these results indicate that when displays contain more elements than can be effectively processed, the elements favored on any trial have not been simply drawn randomly from the display; the

position of an element within the display is related to its probability of inclusion in the sample of elements processed. This finding, while not inconsistent with the suggestion (e.g., Estes & Taylor, 1964) that elements are processed sequentially by a central scanning mechanism, which may follow some quasi-regular path through the displays, does not in itself necessarily imply serial processing. The strongest inference supported by the results cited is that some (presumably central) selective mechanism, possibly probabilistic but definitely nonrandom, determines which particular elements are processed, in a fashion that is relatively stable over trials. For convenience, the selective mechanism will hereafter be referred to as the "scanning process."

The research to be described is addressed to the question of whether the scanning process is fixed by structural factors or is plastic and subject to modification through feedback so as to adapt toward a sampling procedure that is optimal for local display conditions. An appropriate setting is provided by a task in which S, when presented with displays containing noise elements and some number of redundant signal elements, is required to report the number of signals in each display. A systematic relationship between the positions of the signals within the displays may influence the scanning process: Assuming that the average number of elements processed is fixed by S's limitations, he might nevertheless improve his performance by learning to scan in a way that maximizes the probability of encountering all the signals that are present. This point may be illustrated clearly in the context of a simple model that assumes that a "sample" consists of elements correctly classified as signal or noise and that unclassified elements are not available to S; the principles may be generalized to more complex formulations that allow for imperfect classification, or confusion errors. Let

$D_{h,i}$ = a display containing h elements, i of which are signals;
 $S_{j,k}$ = a sample containing j elements, k of which are signals;
 A_m = a response that the display contained m signals.

Table 1
Sample composition, likelihood ratios, and probabilities of correct responses as a function of the size of a random sample.

	$P(S_{j,0} D_{9,1})$	$P(S_{j,1} D_{9,1})$	$P(S_{j,0} D_{9,2})$	$P(S_{j,1} D_{9,2})$	$P(S_{j,2} D_{9,2})$	$LR_{j,0}$	$LR_{j,1}$	$P_j(C D_{9,1})$	$P_j(C D_{9,2})$	$P_j(C)$
0	1.000	.000	1.000	.000	.000	1.000	---	.500	.500	.500
1	.889	.111	.777	.223	.000	.875	2.000	.889	.223	.556
2	.778	.222	.583	.389	.028	.750	1.750	.778	.417	.598
3	.667	.333	.417	.500	.083	.625	1.500	.667	.583	.625
4	.556	.444	.278	.556	.167	.500	1.250	.557	.723	.639
5	.444	.556	.167	.556	.278	.375	1.000	.722	.556	.639
6	.333	.667	.083	.500	.417	.250	.750	1.000	.417	.708
7	.222	.778	.028	.389	.583	.125	.500	1.000	.583	.792
8	.111	.889	.000	.222	.778	.000	.250	1.000	.778	.889
9	.000	1.000	.000	.000	1.000	---	---	1.000	1.000	1.000

Note: $P_j(C)$ should be read as the probability of a correct response when the sample contains j elements.

In particular, consider an experiment in which nine-element displays ($h = 9$) contain with equal frequency either one or two signals ($i = 1, 2$). It can be noted that the sample is unambiguous whenever all elements are sampled ($j = h$) or whenever the sample contains two signals ($k = 2$). All other samples are ambiguous and may arise from displays containing either one or two signals. Let $P(S_{j,k} | D_{h,i})$ denote the probability that a particular sample is drawn from a particular display type. Under the assumption that elements are drawn randomly without replacement, it can be shown that

$$P(S_{j,0} | D_{h,1}) = 1 - \frac{j}{h}, \text{ and } P(S_{j,0} | D_{h,2}) = \begin{cases} \frac{(h-2)!(h-j)!}{h!(h-j-2)!} & \text{for } j \leq h-2 \\ 0 & \text{for } j > h-2 \end{cases};$$

$$P(S_{j,1} | D_{h,1}) = \frac{j}{h}, \text{ and } P(S_{j,1} | D_{h,2}) = \begin{cases} \frac{2j(h-2)!(h-j)!}{h!(h-j-1)!} & \text{for } j \leq h-1 \\ 0 & \text{for } j > h-1 \end{cases};$$

$$P(S_{j,2} | D_{h,1}) = 0, \text{ and } P(S_{j,2} | D_{h,2}) = \begin{cases} 0 & \text{for } j < 2 \\ \frac{j!(h-2)!}{h!(j-2)!} & \text{for } j \geq 2 \end{cases}$$

One way in which S may use the information contained in an ambiguous observation is specified within the framework of statistical decision theory. Any sample has an associated likelihood ratio:

$$LR_{j,k} = P(S_{j,k} | D_{h,2}) / P(S_{j,k} | D_{h,1}).$$

To generate a decision, it is assumed that on each trial $LR_{j,k}$ is compared to a criterion value, LR_c , with the outcome that S chooses response A_m , where

$$m = \begin{cases} 1 & \text{if } LR_{j,k} < LR_c \\ 1 & \text{with probability } \frac{1}{2} \text{ if } LR_{j,k} = LR_c \\ 2 & \text{if } LR_{j,k} > LR_c \end{cases}$$

As in the theory of signal detection (Swets, Tanner, & Birdsall, 1961), LR_c might be expected to be influenced by such experimental variables as payoffs and the a priori probability of the $D_{h,i}$. Assuming for convenience that $LR_c = 1$, values of relevant variables were calculated for the conditions specified and are presented in Table 1 as a function of the size of a random sample, including the sample composition probabilities and likelihood ratios and the probabilities of a correct response conditional upon display type and averaged over display types. Performance clearly depends upon sample size. For any sample size, however, performance can also depend upon the scanning process. Suppose that the signal in $D_{9,1}$ displays appears with equal frequency in all cells of a 3 by 3 matrix; the sampling probabilities as presented in Table 1 are appropriate as an average over displays no matter how S scans the displays. If $D_{9,2}$ displays are constructed so that the two signals always appear in a common row, although equally often in all rows, horizontal scanning can increase the probability of encountering both signals and thus the probability of a correct response on two-signal

displays. The amount of increase is a function of sample size: For example, with random sampling, $P(S_{3,2} | D_{9,2})$ is .083, and $P_3(C | D_{9,2})$ is .583. If the three elements scanned are always taken from a common row, however, $P(S_{3,2} | D_{9,2})$ becomes the probability that the row scanned contains the signal, which is .333, and $P_3(C | D_{9,2})$ becomes .833, a substantial increase. Similarly, with random sampling, $P_6(C | D_{9,2})$ is .417; with the row as the sampling unit, $P_6(C | D_{9,2})$ becomes .667. The same argument would apply to two-signal displays constructed with the signals placed in common columns rather than common rows; under this condition, the optimal scanning process is vertically oriented. Performance under either of the preceding conditions may be contrasted with that expected when the signal pairs are placed equally often in rows and in columns: For this condition, no single scanning process can attain the efficiency of those optimal for the simpler conditions, so that average performance should be inferior to that attained after practice in the all-rows or all-columns condition.

METHOD

Displays and Apparatus

Displays consisted of nine capital letters arranged in matrices of three-letter rows and columns. They were constructed with dry-transfer black lettering (Letraset Instant Lettering, Sheet No. 729) on white file cards. Individual letters were 4 mm high and from 3 to 5 mm in width; the matrices overall were 17 mm high and varied about a mean of 17 mm in width. From the letter set (A,H,M,T,U,V,W,X,Y) three major classes of displays were generated: (a) Eight SX displays were constructed by random assignment of the letters to matrix positions with the constraint that no letter appear twice in the same position. The X appeared in every position except the central cell of the matrix. (b) Eight DX displays were constructed by replacing each of the letters in the set (except the X) with a second X; the two Xs were positioned in common rows in the matrix. These cards were divided into two subsets, designated DH and FH displays. If cells of the matrix are numbered sequentially by rows starting from the upper left, the positions of Xs in the four DH displays were (1,2), (1,3), (7,9), and (8,9). The four FH displays included the combinations (2,3), (4,5), (5,6), and (7,8). (c) Eight more DX displays were constructed in the same way except that the two Xs appeared in the same column rather than the same row. These cards were divided into a DV subset, including Xs in the positions (2,5), (3,6), (4,7), and (5,8), and an FV subset, with Xs in the positions (1,4), (1,7), (3,9), and (6,9).

The displays were combined into three different 16-item lists, which all included the eight SX displays, but differed in their DX composition. *List M* combined the DH and the DV displays; *List H* combined the DH and the FH displays; and *List V* combined the DV and the FV displays.

The cards were displayed with a Scientific Prototype Model GB tachistoscope. The pre- and postexposure field, which was illuminated at all times except for the warning signal and stimulus presentation, appeared as a 6-cm white circle centered in a black background. At the start of each trial the letter O appeared in the center of the circle for 20 msec, serving as a warning signal and fixation point. The matrix of letters appeared 750 msec later; the duration of exposure is detailed below. At the viewing distance of 1.3 m, the matrices subtended a visual angle of 7 deg 20 min.

Subjects and Procedure

Thirty-five (female) students at Douglass College served as S s for credit in introductory psychology courses. They were not screened for visual defects, but those who used corrective lenses were required to wear them during all sessions.

During the first experimental session, S was instructed in the general nature of the task and shown several of the cards; she was

then asked not to move from the tachistoscope eyepiece except during a rest period. The 16 displays of List M were presented in random order, at an exposure duration of 200 msec, as a block of trials; lists were reshuffled during a 30-sec pause after each block. Each trial was initiated with a verbal warning from E and terminated with verbal feedback to S after her report of the number of Xs she thought the display contained. An average interpresentation interval of 12 sec was maintained within blocks.

The second block of displays was presented at 100 msec, and the third block was presented at a 50-msec duration. The session was completed with seven blocks run at the terminal duration of 30 msec. On all days a 5-min rest interval was interpolated between the fifth and sixth block of trials.

Eight Ss were discarded after the first session because their performance (46 or more errors on the last 112 trials) did not differ significantly from chance by a chi-square test. The remaining 27 Ss were divided equally into Group H, Group M, and Group V, equated for total errors during the first session by mean and standard deviation.

On Day 2, Ss in Group H were told that two Xs when present would always appear in a common row; Group V was told that the two Xs would appear in a common column; and Group M was told that the Ss would be either in the same row or the same column. Ten blocks of trials were run at the 30-msec duration: Group H saw List H; Group M saw List M; and Group V saw List V. On Day 3 the procedure of Day 2 was repeated.

RESULTS

The analysis that follows excludes data from the first three blocks of trials on all sessions to eliminate any warmup effects. The blocks remaining provide a total of 112 trials for each S from each session; of these, 56 involved SX displays, and the remainder included 28 presentations of each of the two types of DX display.

The dependent variable of primary interest, performance on the DH and DV displays as a function of practice and context, must be assessed with a measure that includes performance on the SX displays as well and thus reduces contamination by problems of response bias or criterion. Weighted error scores were derived by adding to the number of errors on the four DH or DV cards one-half of the number of errors on the eight SX cards, effectively giving equal weight to SX and DX displays. These scores have the advantage of permitting direct comparisons between Group H and Group M and between Group V and Group M, as they are based on the same displays and the same number of trials. Table 2 presents these weighted error scores by group, display type, and experimental session.

The mean (matched) total errors on Day 1 for Groups H, M, and V were 35.4, 35.9, and 36.8. Their respective SDs were 7.5, 7.9, and 8.0. No difference approaches significance, nor is there any significant difference between the groups on weighted error scores for either the DH or the DV displays considered separately. Within-group comparisons do not reveal consistent differences in

Fig. 1. Weighted errors by blocks of trials.

performance with DH and DV displays; for Group H and Group V the differences are in different directions and nonsignificant. However, Group M made significantly more errors on DV than on DH displays ($t = 2.44$, $df = 8$, $p < .05$, two-tail). When the three groups are combined, the mean difference between display types, 0.91, does not approach significance.

Table 2 reveals consistent decreases in weighted errors over days for all groups and display types. Group H shows a significant improvement from Day 1 to Day 2 ($t = 2.52$, $df = 8$, $p < .025$, one-tail) and a nonsignificant improvement from Day 2 to Day 3; the difference between performance on Day 1 and Day 3 is significant ($t = 3.47$, $df = 8$, $p < .01$, one-tail). Group V displays a similar pattern: The improvement from Day 1 to Day 2 is significant ($t = 7.31$, $df = 8$, $p < .01$, one-tail), as is the net gain from Day 1 to Day 3 ($t = 5.85$, $df = 8$, $p < .01$, one-tail), while the improvement from Day 2 to Day 3 is nonsignificant. The performance of Group M on the DH displays shows a nonsignificant improvement from Day 1 to Day 2, a significant improvement from Day 2 to Day 3 ($t = 3.02$, $df = 8$, $p < .01$, one-tail), and a nonsignificant net gain from Day 1 to Day 3. On the DV displays, Group M shows a significant improvement from Day 1 to Day 2 ($t = 1.87$, $df = 8$, $p < .05$, one-tail), a nonsignificant improvement from Day 2 to Day 3, and a significant gain from Day 1 to Day 3 ($t = 2.15$, $df = 8$, $p < .05$, one-tail).

As mentioned, there were no significant between-group differences on Day 1. It is predicted that on Day 2 and Day 3 the performance of Group M should be inferior to that of Group H and Group V; there is no basis for predicting any particular difference between the latter groups. Differences between Group H and Group V were not significant on any day; neither were differences between Group H and Group M. However, weighted error scores of Group M involving the DV displays are significantly larger than those of Group V both on Day 2 ($t = 2.41$, $df = 16$, $p < .025$, one-tail) and on Day 3 ($t = 2.01$, $df = 16$, $p < .05$, one-tail). The SX component of these weighted error scores must not be overlooked. On Day 2, Group M made significantly more errors on SX displays than did Group V ($t = 2.21$, $df = 16$, $p < .05$, two-tail); the respective proportions of correct responses were .75 and .84. This difference disappeared on Day 3, when the comparable proportions were .80 and .79; the difference in weighted errors on Day 3 is clearly based on responses to the DV displays alone.

The most critical comparison involves between-group differences in improvement over days: The prediction is that Group M

Table 2
Proportions of correct responses and means (and standard deviations) of weighted errors by display type.

Group	Day 1			Day 2			Day 3		
	SX	DH	DV	SX	DH	DV	SX	DH	DV
Group H	.722	.599	.651	.821	.667	—	.829	.663	—
	—	19.0	17.6	—	14.6	—	—	14.2	—
	—	(3.96)	(3.76)	—	(5.05)	—	—	(5.19)	—
Group M	.617	.802	.683	.748	.738	.655	.800	.758	.631
	—	16.3	19.6	—	14.4	16.7	—	12.4	16.0
	—	(4.19)	(4.75)	—	(4.43)	(3.76)	—	(3.83)	(6.03)
Group V	.633	.722	.698	.837	—	.726	.792	—	.810
	—	18.2	18.8	—	—	12.2	—	—	11.1
	—	(4.46)	(3.73)	—	—	(3.99)	—	—	(4.53)

Table 3
Per cent reduction in weighted errors from Day 1 to Day 3.

Group	DH displays	DV displays
Group H	25.3%	---
Group M	23.9%	18.4%
Group V	---	40.9%

Note: Per cent reduction = $100 \times \frac{\text{Day 1 score} - \text{Day 3 score}}{\text{Day 1 score}}$

should show less improvement on DH displays than Group H and less improvement on DV displays than Group V. The differences in improvement between Group H and Group M are for the most part in the expected direction, but they are not significant. The difference between Group V and Group M in change in weighted errors from Day 1 to Day 2 is significant ($t = 2.03$, $df = 16$, $p < .05$, one-tail); the difference in change from Day 2 to Day 3 is nonsignificant; and the difference in net change from Day 1 to Day 3 is significant ($t = 1.85$, $df = 16$, $p < .05$, one-tail). Table 3 summarizes the net reductions in weighted errors from Day 1 to Day 3 as percentages of Day 1 scores.

The development of the significant difference between Group V and Group M is described in insufficient detail by data that are pooled by session. A more sensitive test is provided by Fig. 1, which details performance by block over the entire course of the study and compares weighted error scores for the pairs of groups that share display types. Consistent with the analysis by sessions, Fig. 1 reveals no particular difference between Group H and Group M; both groups show improvement with practice, but there is no obvious difference either in trend or in absolute level of performance. The comparison of Group V with Group M reveals striking agreement between the predicted and observed course of change in performance. Neglecting the three initial blocks of training trials, the two groups are clearly comparable through the early blocks of Day 2, at which point they diverge as expected. After the fourth block on Day 2, the performance of Group V remains superior to that of Group M throughout.

DISCUSSION

Since relatively short lists of unique items were presented repeatedly, correct responses may have become linked to distinctive fragments of individual displays in paired-associates fashion, in which case improvement over days need reflect no fundamental change in the scanning process. If the display types are assumed to be comparable in distinctiveness, associative learning should favor the overall performance of Group M, at least on Day 2, because this group sees familiar displays on all days while one-fourth of the displays are novel on Day 2 for Group H and Group V. One check on such learning is provided by the first two blocks of trials on Day 2: Group H did not make appreciably more errors on the novel FH displays than on the DH displays, nor did Group V make more errors on FV than DV displays. At any rate, the restriction of comparisons to those displays that were common to the groups serves to eliminate the possibility of any systematic differential effect between groups; any associative learning common to groups would tend to mask possible between-group differences in processing by reducing the proportion of displays on which extensive scanning for Xs takes place.

Another possibility that must be considered is that, even though they were instructed always to fixate on the warning letter O that appeared in the center of the display circle, Ss in different groups may conceivably have learned to fixate at different points that were maximally efficient for their particular combination of DX displays. In this case the same scanning process might vary in effectiveness as a function of its starting point or locus within the displays, and, rather than having modified their central scanning mechanism, Ss may simply have

learned a peripheral observing response. This interpretation requires both that optimal performance demand different fixation points for List V and List M and that those different fixations systematically affect performance on the DV displays. In all three lists the DX composition made Xs appear twice in all cells of the matrix except for two in which they appeared only once: for List M and List H, the two anomolous cells were the outer positions of the center row, and for List V they were the outer positions of the center column. A prediction of differential predictions under these relatively balanced conditions seems difficult to justify, and the differential fixation hypothesis does not appear to warrant further consideration, although differential fixation might very well be expected to develop under less symmetrical conditions.

There is an indication that some of the improvement over days is due to increase in processing efficiency, or sample size: Table 3 shows that the weighted-error improvement of Group M, while favoring the DH displays, is roughly of the same order of magnitude for both display types; the improvement of Group H is similar. Increase in performance that does not differentiate between DH and DV displays demonstrates an increase in the mean number of elements processed. However, the improvement of Group V is nearly twice the magnitude of that displayed by the other groups, and it must be concluded that Group V either has increased sample size more than the other groups or has focused the scanning process to a level of efficiency not achieved by the other groups. There is no reason to suspect that Group V should be able to process more elements than the other groups, and thus the conclusion that remains is that the data demonstrate a case of modification of the scanning process toward an ideal that is dictated by local conditions.

The pattern that best seems to describe the data is that Group H and Group M orient their scanning horizontally, while Group V orients its scanning vertically. The weighted error scores on Day 3 involving the DH displays for the former two groups and the DV displays for the latter are similar in magnitude. It is of some interest to question why Group M should tend to scan horizontally; it does not seem implausible that some slight bias toward horizontal scanning may be brought into the experiment on the basis of reading habits, although this suggestion is not supported by the lack of a Day 1 difference in success rates with DH and DV displays. Further research with the multiple-signal technique may serve to specify more precisely what kind of scanning process S does in fact bring to the experimental situation.

REFERENCES

- AVERBACH, E., & SPERLING, G. Short-term storage of information in vision. In R. Haber (Ed.), *Contemporary theory and research in visual perception*. New York: Holt, Rinehart & Winston, 1968. Pp. 202-214.
- ESTES, W. K., & TAYLOR, H. A. A detection method and probabilistic models for assessing information processing from brief visual displays. *Proceedings of the National Academy of Sciences*, 1964, 52, 446-454.
- 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. Visual detection with imperfect recognition. Technical Report No. 128, March 1968, Institute for Mathematical Studies in the Social Sciences, Stanford University, Stanford, California.
- SWETS, J. A., TANNER, W. P., & BIRDSALL, T. G. Decision processes in perception. *Psychological Review*, 1961, 68, 301-340.
- WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology*. New York: Holt, Rinehart & Winston, 1954.

NOTES

1. This research was supported by the Research Council, Rutgers, The State University of New Jersey.
2. Address: Douglass College, Rutgers, The State University of New Jersey, New Brunswick, New Jersey 08903.

(Accepted for publication April 1, 1969.)