UC Irvine

Title
Second choices in a visual span of apprehension task

Permalink

Journal

ISSN

Authors
Yellott, John I
Curnow, Paul F

Publication Date

DOI

Copyright Information

Peer reviewed

Second choices in a visual span of apprehension task

JOHN I. YELLOTT, JR. AND PAUL F. CURNOW

UNIVERSITY OF MINNESOTA

Four Ss were run in a visual span of apprehension experiment to determine whether second choices made following incorrect first responses are at the chance level, as implied by various high threshold models proposed for this situation. The relationships between response biases on first and second choices, and between first choice biases on trials with two or three possible responses, were also examined in terms of Luce's (1959) choice theory. The results were: (a) second choice performance in this task appears to be determined by response bias alone, i.e., second choices were at the chance level; (b) first and second choice response biases were not related according to Luce's choice axiom; and (c) the choice axiom predicted with reasonable accuracy the relationships between first choice response biases corresponding to trials with different numbers of possible response alternatives.

Recent papers by Estes (1965), Estes and Taylor (1964, 1966), and Estes and Wessel (1966) describe a new method of determining visual apprehension span, and investigate several models designed to account for results obtained with this method. In simplest form the procedure in these experiments involves the tachistoscopic display of a set of letters, or other symbols, arranged in matrix form. Among the symbols in a given matrix there is exactly one belonging to a predesignated set of "critical" symbols; the S's task is to report which critical symbol has been presented. The design then is essentially that of a forced choice signal detection experiment. The models that have been studied in connection with these experiments have all been "high threshold" models, in the sense that all assume that on any given trial the critical symbol is either perfectly recognized, in which case the S makes the correct response with probability 1, or else the critical symbol is not detected at all, in which case the S must guess. The designation "high threshold" is used here in the sense described by Swets, Tanner, and Birdsall (1961).

The present study was designed to test a prediction common to all high threshold models: that second choices made following incorrect first responses must be at the chance level. Better than chance performance on second choices has been found in signal detection (Swets et al, 1961), speech recognition (Clarke, 1960), and in the recognition of tachistoscopically presented words (Bricker & Chapanis, 1960). In view of these results it might be expected that better than chance second choice performance would also appear in the Estes-Taylor task, and that consequently the class of high threshold models would be inadequate for that situation.

In the present study, there were three critical symbols, exactly one of which appeared in each stimulus. Let S_1, S_2, and S_3 denote the critical symbols, A_1, A_2, and A_3 the corresponding responses, and let $A_{i,j}$ denote the event of response A_i being given as the jth choice (j = 1 or 2). Then $P(A_{i,1} | S_j)$ denotes the probability that A_i is the first choice, given that the critical symbol is actually S_j, and $P(A_{k,2} | S_j, A_{i,1})$ denotes the probability that A_k is the second choice, given that the critical symbol is S_j and the first choice is A_i. The class of high threshold models of interest here can be represented in a general way by assuming that on every trial there is a (fixed) probability δ that the critical symbol will be detected; if detection occurs the S makes the correct response with probability 1. Otherwise, with probability $1-\delta$, the critical symbol is not detected and the S guesses, selecting response A_i with bias probability p_i. Thus,

$$P(A_{i,1} | S_i) = \delta + (1-\delta)p_i \qquad (1)$$

$$P(A_{i,1} | S_j) = (1-\delta)p_i \qquad (2)$$

for all i, j, $i \neq j$. If the S's first response is wrong, the high threshold model implies that a second choice made between the two remaining alternatives can be based only on pairwise response biases p(i,j), where p(i,j) denotes the (bias) probability of choosing A_i over A_j when only these two responses are available. Thus, in the present experiment a straightforward test of the high threshold assumption can be based on the predictions

$$P(A_{1,2} | S_1 A_{2,1}) = P(A_{1,2} | S_3 A_{2,1}) = p(1,3) \qquad (3)$$

$$P(A_{2,2} | S_2 A_{3,1}) = P(A_{2,2} | S_1 A_{3,1}) = p(2,1) \qquad (4)$$

$$P(A_{3,2} | S_3 A_{1,1}) = P(A_{3,2} | S_2 A_{1,1}) = p(3,2) \qquad (5)$$

There is no necessary relationship between the pairwise biases p(i,j) and the first choice biases p_i, although various choice theories provide for one. In particular, Luce's choice theory (1959) suggests that we ought to find

$$p(i,j) = \frac{p_i}{p_i + p_j} \qquad (6)$$

However, in pilot work it was found that although the high threshold model gave a good account of both first and second choice data, the estimated response biases did not satisfy Equation (6), or any other obvious relationship. The present study then was designed to serve two purposes. The first was to provide a stronger test of the prediction of chance second choice performance. The second was to see whether the Luce theory could predict the relationship between p_i, p_j, and $p(i,j)$ if the choice set were reduced from three responses to two by the E, rather than by the S's making an incorrect first response. For this purpose two types of trials were randomly interspersed in each experimental session. On some trials the S made a first choice from the responses A_1, A_2, or A_3, and then, if this choice were incorrect, a second choice from the two remaining alternatives. On other trials the E eliminated one of the two possible incorrect responses before the S's first choice, so that the first (and only) response on these trials was a choice between two responses. The high threshold model for these two response trials can be represented by

$$P(A_i \mid S_i, \{A_i, A_j\}) = \delta + (1-\delta)\, p'(i,j) \tag{7}$$

$$P(A_i \mid S_j, \{A_i, A_j\}) = (1-\delta)\, p'(i,j) \tag{8}$$

where A_i, A_j denotes the pair of available responses, and $p'(i,j)$ a pairwise response bias probability similar to $p(i,j)$, although not necessarily numerically equal to the $p(i,j)$ of Equations (3), (4), and (5). Interest in the two response trials centered on the question of whether $p'(i,j)$ might not be predictable from p_i and p_j according to Luce's choice axiom, i.e., whether

$$p'(i,j) = \frac{p_i}{p_i + p_j} \tag{9}$$

METHOD

Stimuli

The perceptual task here was essentially that described in Estes and Taylor (1964). The stimuli were 48 5 x 7 in. white cards, each containing a 4 by 4 array of 16 black decal letters (Paratype 36 pt., futura medium). Each stimulus consisted of 15 irrelevant consonants and one letter from the set of critical symbols $\{S, T, U\}$. Each letter occupied a .5 x .5 in. cell of a 2 x 2 in. square field. Forty-eight stimuli were generated by letting each of the three critical letters appear once in each of the 16 possible cells. The three stimuli corresponding to a given position of the critical letter each had the same configuration of irrelevant letters in their remaining cells.

Procedure

A trial consisted of presenting one stimulus at a viewing distance of 48 in. via a Scientific Prototype Model GB tachistoscope. The pre-field on each trial was a white field (8.5 ft-L) containing a small fixation

CRITICAL LETTER	AVAILABLE RESPONSES	RELATIVE FREQUENCY
S —	S,T,U	.223
3R TRIALS T —	S,T,U	.223
U —	S,T,U	.223
S <	S,T	.055
	S,U	.055
2R TRIALS T <	S,T	.055
	T,U	.055
U <	S,U	.055
	T,U	.055

Fig. 1. Relative Frequencies of Trial Types, Sessions 2-10.

point. On a signal from the E the S initiated the stimulus display by means of a hand switch. The luminance of the white background of the letter matrix was 10.5 ft-L. The stimulus display was followed by a blank white field (8.5 ft-L) on which the fixation point reappeared after 3 sec. Each session consisted of 200 such trials and lasted about 1 h. Each S was run through ten 200-trial sessions. In Sessions 2 through 10 there were two types of trials. On three response (3R) trials, all three of the responses S, T, and U were available as possible first responses. If the S's first choice on a 3R trial was incorrect, he was so informed and required to make a second choice from the two remaining alternatives. If the first choice was correct no second response was required. On two response (2R) trials, the S was told, after the stimulus but before his response, that a certain one of the responses was not correct on that trial, and that he should therefore select his response from the two potentially correct responses. The S then made his choice from these two responses. On both 2R and 3R trials the correct response was always indicated to the S after his final choice.

Figure 1 shows the various types of trials in Sessions 2-10 and their relative frequencies. Overall a random 67 percent of the trials in these sessions were 3R trials. Session 1 consisted entirely of 3R trials. In this session, the duration of the stimulus display was varied over trials to determine a value at which the S would be correct about half the time on 3R trials (i.e., a δ of .25). The duration setting obtained in this fashion was then used in the remaining sessions.

Subjects

Four untrained Ss were obtained from an introductory psychology course. Each S was paid a base

rate of one dollar per session, plus one cent for each correct first response on 3R trials, and one-half cent each for all other correct responses. The duration settings as determined in Session 1 were 3.3, 7.0, 4.9, and 4.6 msec for Ss 1, 2, 3, and 4, respectively.

RESULTS

The S_i, A_{ij} notation will be used in reporting results. The correspondence is $S_1 = S$, $S_2 = T$, $S_3 = U$. Data from 2R and 3R trials were analyzed separately; the 3R trials will be considered first. Table 1 shows the first choice response probabilities for each S. The "observed" entries are the relative frequencies of each of the three responses as a function of the critical letter actually presented. The proportions shown are based on all nine experimental sessions; each row corresponds to an N of around 400. The high threshold model of Equations (1) and (2) was evaluated by estimating the parameters δ, p_1, p_2, and p_3 for each S and then using these estimates to predict the first choice probabilities. These predicted values are shown in parentheses in Table 1. Parameters were estimated jointly by the method of least squares. To determine goodness of fit a chi square statistic was computed using the observed and predicted absolute frequencies corresponding to Table 1. Under a minimum chi square estimation procedure this statistic would have df = 3 (Cramer, 1946); the values here were 5.8, 4.9, 1.2, and 7.2 for Ss 1, 2, 3, and 4. None of these is significant at the .05 level (under the rather conservative assumption of df = 3) so the high threshold model appears to given an adequate account of the first choice data.[2] The parameter estimates for each S are shown in Table 2.

To evaluate the high threshold prediction of chance performance on second choices, the three pairs of second choice frequencies corresponding to Equations (3), (4), and (5) were subjected to chi square tests for equality of proportions. Each of the three equations corresponds to a chi square statistic with one

Table 1.

Observed and Predicted First Choice Probabilities on 3R Trials

Subject	Critical Letter	First Choice A_1		A_2		A_3	
		Obs.	Pred.	Obs.	Pred.	Obs.	Pred.
1	S_1	.46	(.47)	.26	(.28)	.27	(.24)
	S_2	.28	(.26)	.51	(.50)	.21	(.24)
	S_3	.24	(.26)	.29	(.28)	.46	(.46)
2	S_1	.27	(.29)	.40	(.39)	.33	(.32)
	S_2	.14	(.14)	.56	(.54)	.30	(.32)
	S_3	.16	(.14)	.36	(.38)	.48	(.48)
3	S_1	.45	(.46)	.32	(.32)	.23	(.22)
	S_2	.22	(.22)	.56	(.57)	.22	(.22)
	S_3	.23	(.22)	.30	(.32)	.47	(.47)
4	S_1	.43	(.46)	.33	(.29)	.24	(.25)
	S_2	.16	(.14)	.62	(.61)	.23	(.25)
	S_3	.15	(.14)	.25	(.29)	.60	(.57)

Table 2. Parameter Estimates from 3R Trials

Subject	$\hat{\delta}$	$\hat{p}_1$	$\hat{p}_2$	$\hat{p}_3$
1	.22	.33	.36	.31
2	.16	.16	.46	.38
3	.24	.29	.42	.29
4	.32	.21	.43	.36
Average	.24	.25	.41	.35

degree of freedom, and the sum of the three statistics for a given S yields a chi square statistic with df = 3 under the high threshold hypothesis of chance second choice performance. The total X^2 values were 3.98, 3.63, 7.08, and 6.75 for Ss 1, 2, 3, and 4 respectively. None of these is significant at the .05 level. The overall probabilities of a correct second choice for Ss 1, 2, 3, and 4 were .53, .53, .54, and .51, as compared with a high threshold prediction of .50. (This prediction does not depend on the values of the pairwise biases p(i,j).) Finally, of the 24 second choice probabilities examined (six for each S), 14 corresponded to a probability of being correct greater than .50, 10 to a probability less than .50. Overall, it appears that the second choices here were determined simply by chance guessing based on response biases, i.e., the second choices did not convey any significant information about the stimulus.

To evaluate the relationship between p_i, p_j, and p(i,j) suggested by Luce's theory, the p_i estimates from the first choice data (Table 2) were used, via Equation (6), to predict p(i,j) values, and these in turn were used to predict second choice absolute frequencies. A comparison of these predicted values and the observed data indicated a generally unsatisfactory fit: The X^2 goodness of fit statistics (df = 6) were 10.14, 61.95, 44.46, and 33.70 for Ss 1, 2, 3, and 4, respectively. All but the first of these are significant at the .05 level. The observed pairwise probabilities are given in Table 3, together with the values predicted by Equation (6) (shown in parentheses). Each "observed" entry is the overall proportion of times A_i was chosen over A_j on all second choices in which the pair $\{A_i, A_j\}$ was available. These proportions are based on Ns ranging from 123 to 314. It will be observed that there is no particularly striking correspondence between observed and predicted values here, except perhaps in the case of S 1. It appears that Luce's choice axiom does not, in general, correctly predict the relationship between first and second choice response biases.

Data from the 2R trials were first examined to determine whether the sensitivity parameter δ had the same value on these trials as on the 3R trials. The estimated δ values were .25, .20, .21, and .32 for Ss 1 through 4, respectively. (These estimates were based on the fact that the sum of the six error probabilities (Equation (8)) generated by the 2R trials is equal to $3(1-\delta)$.) A comparison with the estimates in Table 2 will show that the δ estimates from 2R

trials are quite close to the corresponding estimates from the 3R trials.

The viability of the high threshold model for the 2R trials was tested first by estimating pairwise biases for these trials (using Equations (7) and (8)) and then using these biases, together with the δ estimates from 3R trials, to predict the various response probabilities on 2R trials. These predictions are shown in Table 4 (the "b" rows) together with the observed probabilities. For each S the table gives the observed and predicted probabilities of a correct response on each of the six types of 2R trials. The chi square values (df = 3) corresponding to these predictions are 0.86, 2.03, 8.28, and 0.96 for Ss 1, 2, 3, and 4. The value for S 3 is significant (p < .05), the rest are not.

To test the choice axiom prediction of Equation (9), p_i estimates from the 3R trials (Table 2) were used to predict $p'(i,j)$ values, and these predicted biases, together with δ estimates from the 3R trials, were used (via Equations (7) and (8)) to predict the 2R probabilities. These predictions are shown in the "a" rows of Table 4. These predictions are generally quite satisfactory. The chi square goodness of fit values (df = 6) were 25.26, 4.08, 12.29, and 6.87 for Ss 1, 2, 3, and 4. The first of these is significant (p < .05), the rest are not. It is worth noting that these predictions for the 2R data are based entirely on parameters estimated from the 3R trials, and Equations (6), (7), and (8); there is no overlap between the data used in estimating the parameters and the data being predicted. The overall goodness of fit across all four Ss is reflected in the average pairwise biases: The observed averages for p(1,2), p(1,3), and p(3,2) were .33, .43, and .46, while the averages of the corresponding predicted values were .37, .42, and .45.

DISCUSSION

This experiment shows, first of all, that under the conditions of the Estes-Taylor span of apprehension task second choices made following incorrect first responses convey little or no information about the stimulus; second choice behavior here can apparently be accounted for simply in terms of non-discriminative response biases. This result is consistent with the class of high threshold or "all or none" models that have been proposed for this situation (Estes & Taylor, 1964, 1966). Although none of the models suggested so far has been entirely

Table 3. Pairwise Bias Probabilities on Second Choices

Subject	$\hat{p}(1,2)$		$\hat{p}(1,3)$		$\hat{p}(3,2)$	
1	.45	(.48)[a]	.52	(.52)	.54	(.46)
2	.45	(.26)	.39	(.30)	.45	(.45)
3	.53	(.41)	.65	(.50)	.35	(.41)
4	.46	(.33)	.44	(.36)	.36	(.46)

a Entries in parentheses are the values predicted by substituting the p_i estimates of Table 2 into Equation 6.

Table 4. Correct Response Probabilities on 2R Trials as a Function of Trial Type

Subject	Critical letter:	S1		S2		S3	
	Choice set:	{A1A2}	{A1A3}	{A1A2}	{A2A3}	{A1A3}	{A2A3}
1	Obs.	.53	.55	.72	.53	.71	.70
	Pred.[a]	.59	.62	.63	.64	.60	.58
	Pred.[b]	.51	.52	.70	.52	.69	.69
2	Obs.	.42	.41	.80	.70	.76	.53
	Pred.[a]	.38	.41	.78	.62	.75	.54
	Pred.[b]	.38	.41	.77	.66	.74	.49
3	Obs.	.40	.70	.65	.68	.59	.61
	Pred.[a]	.55	.62	.69	.69	.62	.56
	Pred.[b]	.50	.67	.74	.66	.57	.59
4	Obs.	.53	.65	.74	.74	.71	.57
	Pred.[a]	.54	.56	.78	.69	.76	.63
	Pred.[b]	.56	.63	.76	.75	.69	.58

a Predicted from Luce model and parameters estimated from 3R trials.
b Predicted from 3R δ estimates and pairwise biases estimated from 2R data.

satisfactory (Estes & Wessel, 1966), the present study indicates that one need not abandon the assumption that the critical recognition process here is all or none.

At first glance these results might seem to contradict Bricker and Chapanis' (1960) finding that second, and even later, choices in tachistoscopic word recognition are better than chance. However, Bricker and Chapanis specifically attributed their result to the possibility that a S could recognize some, but not all, of the letters in a word, fail to make a correct first choice, and still use the recognized letters to advantage in subsequent choices. In the present experiment, performance depended entirely on recognizing the single critical letter; any recognition of irrelevant letters in the display could not affect second choice performance. Consequently Bricker and Chapanis' results, as they interpret them, are not incompatible with chance second choice performance in the present experiment.

The fact that Luce's choice axiom (represented here by Equation (6)) did not accurately predict second choice response biases is consistent with results obtained by Clarke (1960) in a speech recognition task. He found second choices to be somewhat better than chance, but was unable to account for the numerical values of second choice probabilities, either in terms of a threshold model, "an extension of the constant ratio rule" (Luce's choice axiom), or any signal detectability model. An adequate model relating first and second choice probabilities is still unavailable. However, it is possible to suggest an explanation of the fact that the choice axiom did, by and large, correctly predict the relationship between first choice biases on 2R and 3R trials, but at the same time failed to account for second choice biases. It is well established that response biases in many psychophysical tasks are not constant within a session, but vary from trial to trial according to

the history of reinforcing events (Atkinson, Bower, & Crothers, 1965). The present study was concerned only with accounting for the average values of the bias probabilities, on the assumption that the bias process is at least stochastically stationary. However, if there are trial to trial fluctuations in response bias, then when a response, say A_1, is given incorrectly as a first choice on some trial n, its momentary bias probability $p_{1,n}$ will tend to be higher than average, and the other bias probabilities ($p_{2,n}$ and $p_{3,n}$ in this case) will tend to be below their respective averages. It is these non-average values that then determine the second choices, perhaps according to the choice axiom:

$$p(2,3) = \frac{p_{2,n}}{p_{2,n} + p_{3,n}}$$

However, even if this relationship obtained we should not, in general, expect to be able to predict the average second choice probabilities correctly if we simply substitute estimates of (the averages of) p_2 and p_3 based on the first choices; these averages will not be estimates of the appropriate conditional probabilities. The 2R trials, on the other hand, are random samples of all trials, and consequently the average bias values over these trials will be equal to the average bias values on 3R first choices. Consequently we would expect 2R and 3R first choice biases to be predictable, one from the other, if the choice axiom holds, as it apparently does to a good approximation in this experiment.

If the argument given above is applicable to the present experiment, it should be possible to demonstrate that response biases did in fact fluctuate over trials. For this purpose the sequential probabilities $P(A_{1,1} | S_1 A_{1,1})$ and $P(A_{1,1} | S_1 A_{x,1})$ were computed for each S. The first quantity is the conditional probability of an A_1 first response (on 3R trials) given that the stimulus is an S_1, and that the first response on the previous trial was A_1. (The previous trial may be either a 2R or 3R trial.) The second quantity is the conditional probability of an A_1 first response given an S_1 stimulus and a non-A_1 (i.e., A_2 or A_3) first response on the previous trial. If the bias probabilities were fluctuating gradually over trials, then we should expect to find that $P(A_{1,1} | S_1 A_{1,1})$ is greater than $P(A_{1,1} | S_1 A_{x,1})$, since the occurrence of A_1 on Trial n-1 would tend to imply an above average value of p_1 on Trial n. In fact, a substantial effect of this sort was found for Ss 3 and 4: The respective values of $P(A_{1,1} | S_1 A_{1,1})$ and $P(A_{1,1} | S_1 A_{x,1})$ were .56 vs .44 for S 3, and .53 vs .42 for S 4. For Ss 1 and 2, on the other hand, the two probabilities were virtually the same: .44 vs .46 for S 1, and .27 vs .26 for S 2. However, when the sequential probabilities $P(A_{i,n} | S_{j,n} S_{k,n-1})$ were examined, S 2 was found to show a very pronounced negative recency effect: When the critical letter on Trial n (a 3R trial) was the same one that had been

presented on Trial n-1, the overall probability of a correct response on Trial n was .38, as compared with a correct response probability of .46 when the stimulus on Trial n differed from that on Trial n-1. A detailed analysis of the sequential probabilities confirmed the significance of this negative recency effect. For S 1 no significant sequential effects due to stimuli could be found.

In summary then, substantial sequential effects reflecting fluctuations in response bias could be demonstrated for Ss 2, 3, and 4, but not for S 1. This pattern of results is consistent with the pattern of successes and failures found in attempting to predict the second choice probabilities. According to our interpretation the substitution of first choice p_i, p_j estimates into Equation (6) will result in successful prediction of second choice probabilities only if the trial by trial variance in the response biases is negligible. This condition was apparently satisfied only in the case of S 1, and S 1 was in fact the only observer for whom the observed second choice probabilities did not differ significantly from their predicted values.

References

Atkinson, R. C., Bower, G. H., & Crothers, E. J. *An introduction to mathematical learning theory*. New York: John Wiley & Sons, 1965.

Bricker, P. D., & Chapanis, A. Do incorrectly perceived tachistoscopic stimuli convey some information? *Psychol. Rev.*, 1953, 60, 181-188.

Clarke, F. R. Confidence ratings, second choice responses, and confusion matrices in intelligibility tests. *J. Acoust. Soc. Amer.*, 1960, 32, 35-46.

Cramér, H. *Mathematical methods of statistics*. Princeton: Princeton University Press, 1946.

Estes, W. K. A technique for assessing variability of perceptual span. *Proc. Nat. Acad. Sci.*, 1965, 4, No. 2, 403-407.

Estes, W. K., & Taylor, H. A. A detection method and probabilistic models for assessing information processing from brief visual displays. *Proc. Nat. Acad. Sci.*, 1964, 52, No. 2, 446-454.

Estes, W. K., & Taylor, H. A. Visual detection in relation to display size and redundancy of critical elements. *Percept. & Psychophys.*, 1966, 1, 9-16.

Estes, W. K., & Wessel, D. L. Reaction time in relation to display size and correctness of response in forced choice visual signal detection. *Percept. & Psychophys.*, 1966, 1, 369-373.

Luce, R. D. *Individual choice behavior*. New York: John Wiley & Sons, 1959.

Swets, J. A., Tanner, W. P., & Birdsall, T. G. Decision processes in perception. *Psychol. Rev.*, 1961, 68, 301-340.

Notes

2. To see whether the actual minimum chi square parameter estimates could produce any substantial improvement in the goodness of fit, the parameters δ, p_1, p_2, and p_3 were reestimated to an accuracy of $\pm .01$ by a chi square minimization procedure. All of the resulting estimates were either identical to the corresponding least squares estimates, or differed by exactly .01. Over all four subjects the total of the minimum chi squares was 19.01 (df = 12) as compared with the value of 19.1 generated by the least squares estimates.