The non-monotonicity of the psychometric function in recognition memory[1]

PHILIP TWITCHELL SMITH[2]
UNIVERSITY OF OXFORD

The technique of using the signal detectability theory parameter d' as a measure of trace strength in recognition memory is extended in a decision-making model that predicts that a S's psychometric function [p(Y/s) plotted against stimulus intensity] will have a peak for intermediate values of trace strength when the S has a bias towards a "Yes" response. An experiment is reported where Ss were required to recognize previously presented nonsense syllables under a number of pay-off conditions designed to give different groups of Ss different response biases. The predictions of the model were confirmed. It is suggested that a S's discriminability and response bias parameters may not be as stable as previous studies, that pool large samples of data, have indicated, and the results also provide support for a two-stage model of recognition, where a preliminary estimate of the trace strength of an item is obtained before the location of the response criterion is decided.

Attempts to explain performance in memory tasks in terms of the concept of *trace strength* in memory are not new (e.g., Hull et al, 1940). The basis of such explanations is that a S's response probability for a given item is related to experimental variables (such as the delay between presentation and recall) by the mediation of the theoretical construct of the strength of the item's memory trace in the S's mind; for example, a simple decay theory of memory would assert that trace strength decays with time, resulting in a decreasing probability of correct response, with increasing time interval. With the advent of signal detectability theory the inadequacies of an approach that related trace strength and response probability in such a direct way were quickly realized, and a number of workers have used d', the signal detectability theory measure of discriminability, as a measure of trace strength (Norman & Wicklegren, 1965; Murdock, 1966; Wicklegren, 1966; Wicklegren & Norman, 1966). With this approach, decision processes in perception and in memory are to be viewed in the same way, with the analysis of the detection process in terms of the *discriminability* of the items presented (the parameter d') and of the location of a *response criterion* (the parameter β) applying equally well to items directly presented, as in perception, or indirectly retrieved from a conceptual memory store, as in trace strength analyses of memory.

This paper represents an attempt to develop this parallel between perception and memory with the construction of psychometric functions for items in recognition memory tasks [a psychometric function is a plot of p(Y/s)—the probability of responding "Yes" given an appropriate signal—against the strength of s]. In so doing it will be demonstrated that the decision processes normally assumed for recognition memory are inadequate and need extension. In particular, it will be demonstrated that the psychometric function in recognition memory has a peak at an intermediate value of d' (the measure of trace strength), which necessarily implies that interactions must exist between the discriminability and response bias parameters of the detection model. In a wider context, such interactions mean that the generality of studies attempting to show independent variation of the discrimination and response bias components of a decision is called into question. For example, studies of variation of β with d' fixed (Green, 1960;

Broadbent & Gregory, 1963; Broadbent, 1967; Murdock, 1965), may depend critically upon the value of d', and studies of the relationship between confidence ratings and the a posteriori probability of correctness should not be expected to show the uniformity that has been claimed in other studies (e.g., Pollack & Decker, 1958; Clarke, 1960; and Kintsch, 1968).

In what follows, the standard assumptions of signal detectability theory will be made, namely that in a binary decision task (signal vs noise) an S calculates a likelihood ratio from his observation, and makes a positive response if and only if this likelihood ratio exceeds a criterion point; the distributions of likelihood ratios when signal and when noise are present are assumed to be normal, with equal variance.

Even for traditional decision models, the form of the psychometric function is critically dependent on the type of decision rule the S is using. Two such rules will be discussed here: (A1) the Neyman-Pearson criterion—maximize p(Y/s) subject to the constraint that p(Y/n) < α, and (A2) maximize expected pay-off.

Decision Rule A1 produces a monotonic psychometric function: If p(Y/n) is held constant at α and discriminability is increased, p(Y/s) will increase. Decision Rule A2 produces a U-shaped psychometric function if the pay-off situation favors a "Yes" response and a monotonic psychometric function if the pay-off favors a "No" response. The minimum of the "U" is at a point corresponding to p(Y/n) = ½ (see Fig. 1).

However, the situation is more complicated if certain levels of discriminability make certain levels of response bias more or less difficult to maintain. This situation arises in the author's decision-making model (Smith, 1968), which is concerned with the *cost* incurred in making a psychophysical decision. This model supposes that an S incurs particularly high cost whenever he has to make a decision involving a likelihood ratio particularly close to his criterion. This has the consequence that to a fair approximation in a binary decision task (signal vs noise) the average cost incurred in making a decision is a linear function of the sum of the two probability density functions at the criterion. This has the further consequence that for poor discriminability (d' < 2) it is less costly to be biased than unbiased, but for high discriminability (d' > 2) it is less costly to be unbiased than moderately biased. It is assumed that cost is something that the S wishes to keep as small as possible, within the limits imposed by other requirements of the task.

Decision rules in this situation might then take the form: (B1) maximize p(Y/s), subject to the constraints that p(Y/n) < α and decision cost < K, and (B2) maximize expected pay-off, subject to the constraint that decision cost < K.

There will in general be a range of values of d' where the criterion C defined by Decision Rule A1 or A2 does not coincide with the corresponding criterion C' for Decision Rule B1 or B2, because in the latter case the cost that would be incurred in maintaining a criterion at C is too high. For Rule B1, as Fig. 2 shows, this would produce an uneven psychometric function because of a discontinuity in the location of C'. For Rule B2, as Fig. 3 shows, at levels of very high and very low discriminability C and C' coincide, and when the pay-off situation favors a "Yes" response the
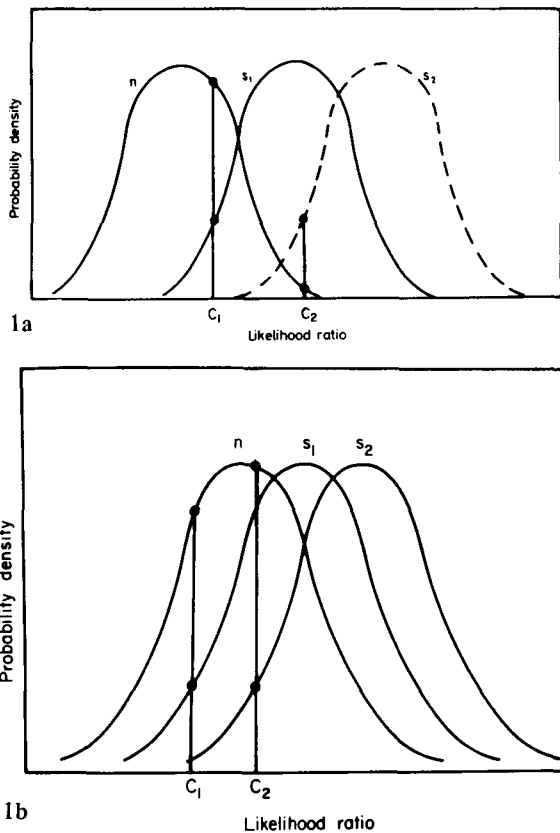
Fig. 1. Location of criterion under Decision Rule A2: n = noise distribution; s_1 = signal distribution; s_2 = signal distribution when d' is increased; C_1 = criterion for s_1 vs n condition; and C_2 = criterion for s_2 vs n condition that produces the same hit rate as C_1 produces for the s_1 vs n condition.

In Fig. 1a, C_1 and C_2 produce false positive rates $[p(Y/n)]$ of less than ½, and thus the likelihood ratio at C_2 in the s_2 vs n condition is larger than it is at C_1 in the s_1 vs n condition; therefore C_2 must be moved to the left to reduce the likelihood ratio in the s_2 vs n condition to its former optimal value; this necessarily increases the hit rate $[p(Y/s)]$.

In Fig. 1b, C_1 and C_2 produce false positive rates of more than ½, and thus the likelihood ratio at C_2 in the s_2 vs n condition is smaller than it is at C_1 in the s_1 vs n condition. Therefore C_2 must be moved to the right to increase the likelihood ratio in the s_2 vs n condition to its former optimal value; this necessarily reduces the hit rate.

If the pay-off situation favors a "Yes" response, then for some sufficiently small value of d' the criterion should be placed to produce a false positive rate of more than 50%. Thus a U-shaped psychometric function will arise, $p(Y/s)$ decreasing from 1 as d' increases from zero until the point where $p(Y/n) = $ ½ is passed, thence onward $p(Y/s)$ increasing with d'.

If the pay-off situation favors a "No" response, the false positive rate should never exceed 50%, and thus $p(Y/s)$ will always be monotonic increasing with d'.

resultant discontinuity in the location of C' at a level of moderate discriminability produces a sharp fall in $p(Y/s)$ at this level. The discontinuity is greater and occurs at a larger value of d' if the overall bias is very high, since the S in this case will be more ready to move his criterion C' to a position of extreme bias and less ready to transfer it to the more moderately biased ideal criterion C.

Rule B2 therefore is the only decision rule that has been considered that predicts $p(Y/s)$ will have a maximum for intermediate values of stimulus strength. It is not suggested that the decision rules considered represent the only or even

the most likely strategies available to the S; the point being made is that with the model of cost proposed, if performance is constrained by a ceiling to the cost that can be tolerated, a psychometric function with an intermediate peak is likely to arise, whereas it is difficult to see how this could come about without such a model. The following experiment, in demonstrating the existence of a peak, will at least discredit the simple Decision Rules A1 and A2.

METHOD

Subjects were required to carry out a series of recognition memory tasks. In each task, the S was presented with a list of 18 three-letter nonsense syllables, which he was asked to read out in time with a metronome which was beating at the rate of one beat per second. After reading out the list, the S was given an interference task for exactly 2 min (crossing out particular letters in a text). At the end of this time, he was given a second list of 20 three-letter nonsense syllables in which, he was told, there was a 50% probability that any particular syllable had occurred in the first list, and he was asked to work serially through the list, writing "Yes" or "No" beside each syllable to indicate whether he did or did not think that that particular syllable had occurred in the first list. (Stimulus list and test list were both random samples of the same 36-syllable population, and this population was changed from trial to trial). When the S had completed the test list, the E worked through this list with the S, awarding points for correct and incorrect responses on the basis of one of the following pay-off matrices.

			S's Response		Number of Ss assigned to this condition
			Yes	No	
I. Unbiased condition	Correct Answer	Yes	2	−2	5
		No	−2	2	
II. "Yes" bias condition	Correct Answer	Yes	3	−3	6
		No	−1	1	
III. "No" bias condition	Correct Answer	Yes	1	−1	4
		No	−3	3	

Any particular S always had the same pay-off matrix, and knew in advance, before the task began, which pay-off matrix this was. The S was told his task was to maximize the number of points he was awarded, and it was pointed out that this need not be the same as maximizing the number of correct answers. In one session, each S carried out seven tasks, with only a brief pause between each; the S was told that material learned in one task was not relevant to the other tasks, i.e., he could safely forget the items in the lists once the task was completed. All the Ss attempted the same series of lists in the same order. At the end of the session, points awarded from the pay-off matrix were converted into money at the rate of 100 points per shilling. This meant that Ss took away between threepence and two shillings per session in addition to the standard payment of four shillings per hour. Each session lasted about 45 min. Each S came for four sessions, each session being separated by a few days (not more than a week).

The critical features of this experimental design are as follows.

(1) This is a task where a self-paced S produces a series of responses from memory; since it is a memory task, there is some premium on the S's responding reasonably rapidly (particularly if he believes in a trace-decay theory of memory). This then is a condition where the S is likely to consider the cost and speed of making a particular decision in relation to the accuracy of this decision; thus, the preconditions of the
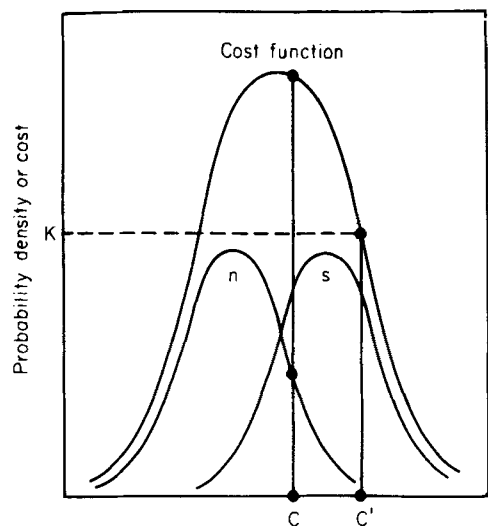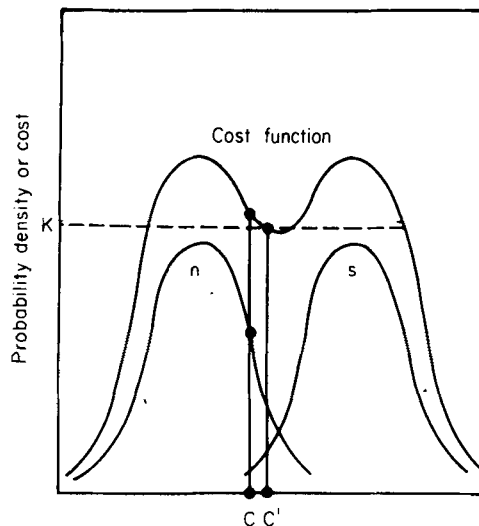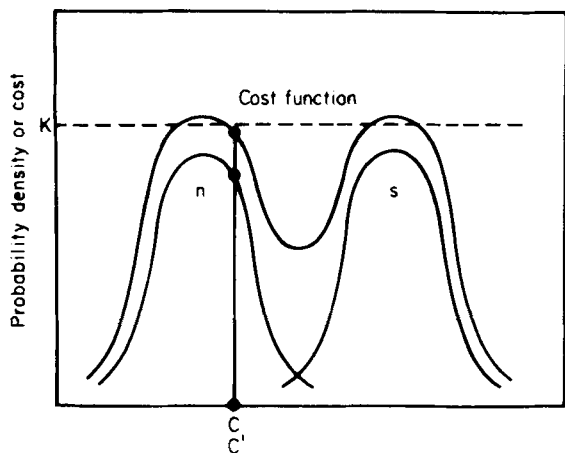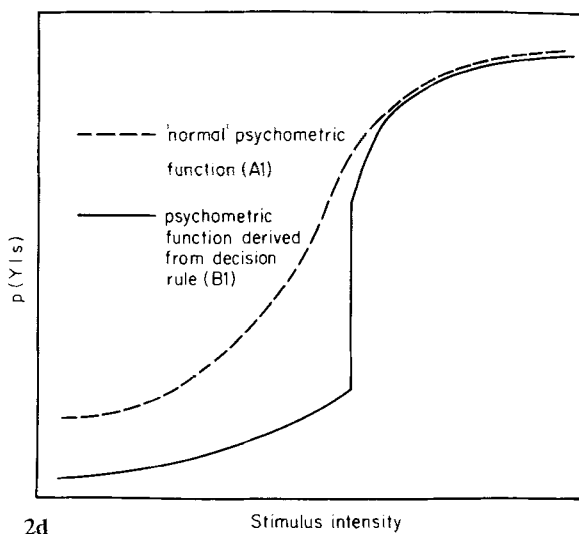
Fig. 2. Location of Criterion C' under Decision Rule B1 compared with location of Criterion C under Decision Rule A1. A discontinuity in the location of C' occurs between Figs. 2a and 2b. Fig. 2d shows the corresponding psychometric functions.

model proposed above—that there shall be an interaction between cost and other features of performance—are likely to be met.

(2) By pacing the S as he reads out the stimulus list, it is hoped that gross individual differences in *learning* strategies are avoided. The S just does not have time to utilize peculiar mnemonic strategies such as "I'll only learn the first half of the list thoroughly" or "I'll see how many syllables begin with p"; he is briefly exposed to each stimulus, and forced, by his reading of it, to give it at least a minimum of attention. Attempts to explain the results of this experiment in terms of learning rather than response strategies are made correspondingly difficult. This does not preclude, for example, the existence of position effects in the test list that interact with experimental conditions; this will be dealt with in the results section.

(3) The use of different pay-off matrices is intended to produce different response biases in different groups of Ss; the arguments in the previous section suggest that level of response bias will critically effect the appearance of the psychometric function.

(4) The stereotyped method of constructing each list (the details of which were not explained to the S) means that roughly the same level of difficulty and the same type of confusions will occur on every list, so with repeated tasks the S will soon learn the type of task to expect. Also, with the amount of feedback given at the end of each task (the S is shown what the correct answers are), the S's response strategy ought also to achieve some sort of stability. It is hoped then that the performance obtained does not contain many random or superstitious responses from Ss, but represents the intelligent guesses of well-informed Ss attempting to produce optimal behavior.

RESULTS

Pooled Data

General trends in the data can be seen by pooling data from Ss who received the same pay-off matrix and classifying performance in terms of (1) order within test list—first, second, third, and fourth quarters of each test list being considered separately; (2) trial number within a session (1 to 7); and (3) session number (1 to 4).

Discriminability. Tables 1-3 show median d' for various combinations of these conditions. Although trends are suggested, the only significant effect is the variation of d' with session (Friedman two-way analysis of variance, $p < 0.05$).
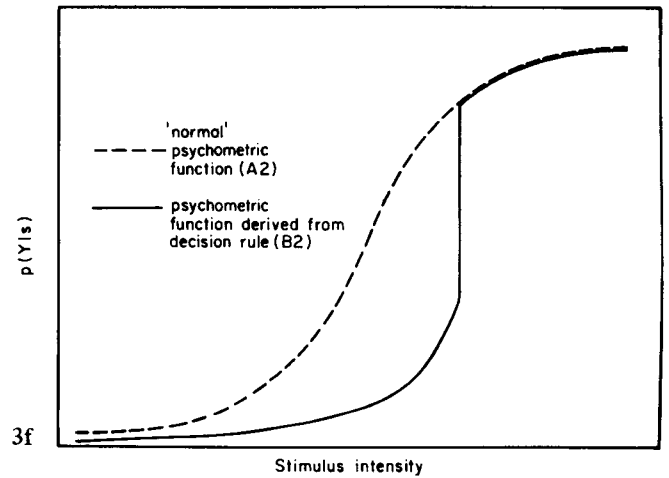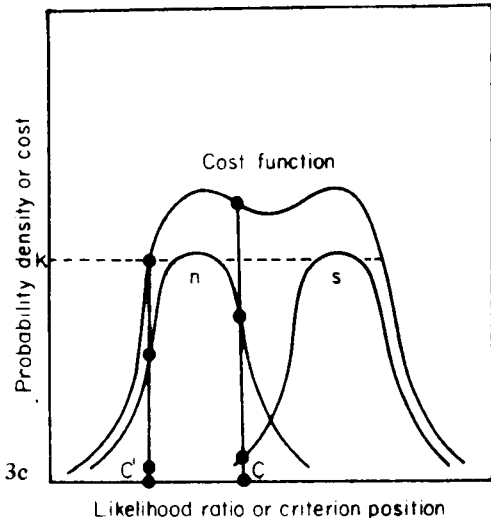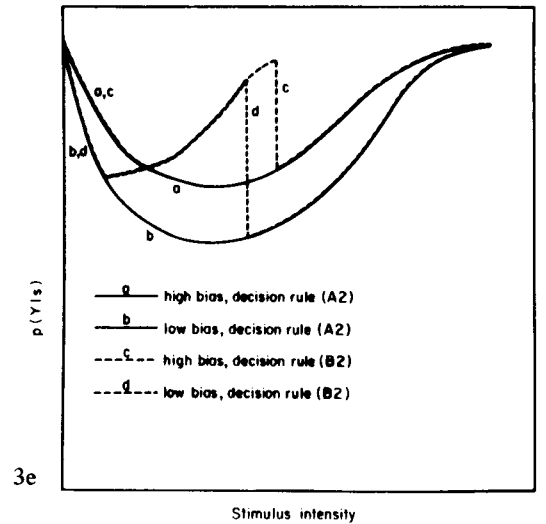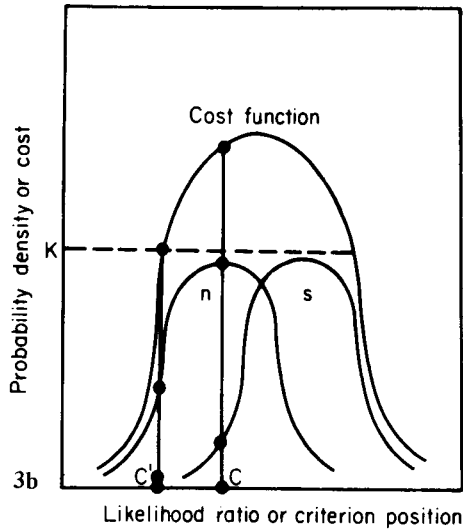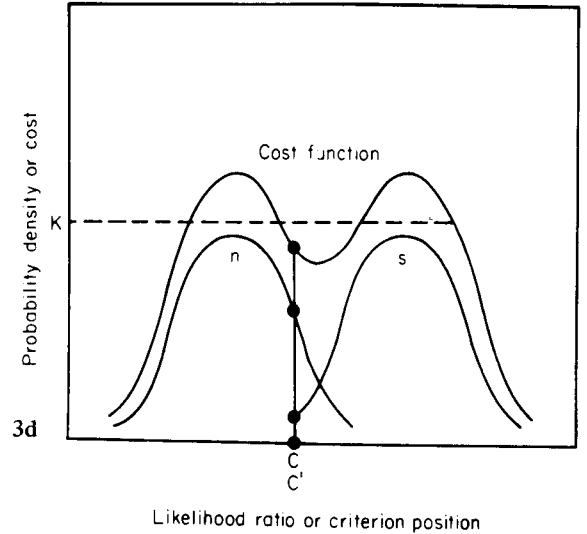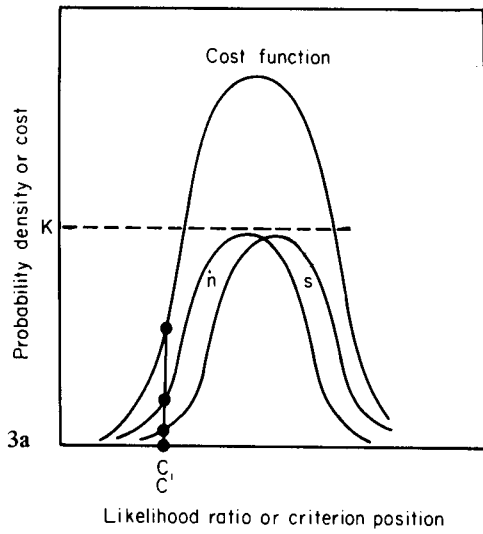
Fig. 3. Location of Criterion C′ under Decision Rule B2 compared with location of Criterion C under Decision Rule A2. C and C′ coincide for low and high discriminability (Figs. 3a and 3d, respectively), and there is a discontinuity in the location of C′ between Figs. 3c and 3d. Figure 3e shows the corresponding psychometric functions. Figure 3f sketches the psychometric functions for a bias away from (rather than towards) a "Yes" response.

Table 1
Median d' as a Function of Test List Position for the Different Experimental Groups

Position in Test List	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter
Unbiased group	0.96	0.96	1.12	0.84
"Yes" bias group	1.06	1.11	1.13	0.64
"No" gias group	1.23	0.88	1.26	0.95

Table 2
Median d' as a Function of Session for the Different Experimental Groups

Session	1	2	3	4
Unbiased group	0.96	0.78	0.97	0.96
"Yes" bias group	1.00	0.96	1.12	1.10
"No" bias group.	0.95	0.85	1.35	1.35

Table 3
Median d' as a Function of Trial Number for the Different Experimental Groups

Trial	1	2	3	4	5	6	7
Unbiased group	0.78	0.91	1.12	1.25	0.92	0.80	0.96
"Yes" bias group	0.51	0.97	1.25	0.81	1.12	1.18	1.12
"No" bias group	0.68	0.97	1.35	1.35	1.35	1.05	1.00

There are no significant differences for d' with different pay-off matrices, and no suggestion of an interaction between pay-off matrix and d' over any of the other experimental variables. This last point is important when individual data are discussed later.

Response bias. Table 4 shows the distribution of the number of occasions that pooled data showed a bias towards "Yes" response, "No" response, or is unbiased. There is a bias towards a "Yes" response over and above the pay-off matrix condition (Wilcoxon matched pair on "Yes" matrix vs "No" matrix, $p < 0.01$, and on the unbiased matrix $p < 0.05$). There is a session effect: there are many more "wrong" biases in the first session than in subsequent sessions (Kruskal-Wallis one-way analysis of variance, $p < 0.01$ for "Yes" matrix, just fails to reach significance for "No" matrix, $0.05 < p < 0.10$; a "wrong" bias is a non-Yes bias for the "Yes" matrix and a non-No bias for the "No" matrix). And there is an order effect: bias towards "Yes" response is greater at the end rather than at the beginning of the test list (Wilcoxon matched pair, $p < 0.01$ for "Yes" matrix, $p < 0.01$ for "No" matrix) but this effect does not appear for the unbiased matrix condition. There is no detectable variation of response bias with trial number for any of the three conditions.

Raw scores. Mean total scores for the three conditions were: unbiased matrix, 391; "Yes" bias matrix, 509; and "No" bias matrix, 420. "Yes" bias score was significantly greater than "No" bias score (Mann-Whitney U, $p < 0.05$). The S who always responded "Yes" in the "Yes" bias condition or "No" in the "No" bias condition would have an expected total score

Table 4
Distribution of Response Biases for the Different Experimental Groups

	"Yes" bias	Responses show Unbiased	"No" bias
Experimental group receiving			
1 Unbiased matrix	59	7	43
2 "Yes" bias matrix	95	1	13
3 "No" bias matrix	35	14	60

of 560. No S in the bias matrix conditions achieved as high a score as this, indicating a marked departure from optical behavior, if optimality is defined as getting as many points as possible (because the S could do better without even seeing the stimulus list).

Individual Data

The conclusions of the above section are that the group of Ss receiving the unbiased matrix produce a fairly stable performance with only minor variations in median d' within lists and over trials and sessions, and a relatively unbiased criterion that also shows no changes within lists and over trials and sessions. It is this group that will be used as a standard to which the performance of the other groups may be related.

The technique of analysis will be to use the proportion of "Yes" and "No" responses made by members of the unbiased group to a particular item as a sort of confidence rating for that item. Note that obtaining confidence ratings directly from Ss in this situation is inappropriate, because of the interference effects on the Ss' mnemonic performance that are likely to result. The distribution of responses for the unbiased group is shown in Table 5.

It is not appropriate to regard this distribution as the product of chance variations based on a constant d' for all items. The distributions for "Yes" and for "No" correct responses both differ sharply from binomial (for "Yes" correct response $\chi^2 = 21.8$, $p < 0.001$; for "No" correct response

Table 5
The Distribution of k (the Number of "Yes" Responses made to an Individual Item by the Five Members of the Unbiased Group)

Number (k) of "Yes" responses made to an individual item by the five members of the unbiased group	Number of occasions k "Yes" responses were made	
	Correct Response "Yes"	Correct Response "No"
5	59	8
4	85	25
3	74	43
2	39	70
1	18	75
0	1	63

$\chi^2 = 58.3$, $p << 0.001$). The departure from binomial distribution is in the direction of too many values at the extremes of the distribution. This suggests that it is more natural to regard the proportions of "Yes" and "No" responses made by Ss to individual items as indices of the discriminability of those items: five "Yes" responses to an item when the correct response is "Yes" suggests that this is an unusually easy item (high d'); zero "Yes" responses for correct response "Yes" suggests an unusually difficult item (low d'); the reverse is true if the correct response is "No." If it is further assumed that a high correlation exists between the discriminability of an item for any S in any experimental condition and the discriminability of that same item for any other S (an assumption that the results so far give no evidence for contradicting), then the distribution of the proportion of "Yes" and "No" responses for the unbiased group can be used as a distribution of discriminabilities to investigate the effect of discriminability on the performance of the other groups. [The decision rules, A2 and B2, discussed in the introduction both predict monotonicity of p(Y/s) with d' for an *unbiased* S.]

Specifically, for individuals in the bias groups, the functions p(Y/s,k), p(Y/n,k) will be determined, that is, the probabilities of saying "Yes" correctly and "Yes" incorrectly to items that k members of the unbiased group ($0 \le k \le 5$) had said "Yes" to. p(Y/s,k) and p(Y/n,k) were estimated for each S for the

Table 6
Mean Values of p(Y/s,k) and p(Y/n,k) for the "Yes" Bias and for the "No" Bias Groups

	"Yes" group	"No" group
p(Y/s,5)	0.849	0.718
p(Y/s,4)	0.814	0.686
p(Y/s,3)	0.813	0.617
p(Y/s,2)	0.737	0.490
p(Y/n,3)	0.636	0.404
p(Y/n,2)	0.581	0.330
p(Y/n,1)	0.411	0.203
p(Y/n,0)	0.266	0.132

Table 8
Distribution of Spearman Rank Order Correlations (r_S) between p(Y/s,k) and k, and between p(Y/n,k) and k, for the "No" Bias Group

	r_S [p(Y/s,k),k] < 0	r_S [p(Y/s,k),k] ⩾ 0
r_S [p(Y/n,k),k] ⩾ 0.8	6	12
r_S [p(Y/n,k),k] < 0.8	0	15

first half and the second half of all lists for a given session. No interesting variations occurred in these estimates for a given S, and in what follows they are regarded as eight independent samples (2 list positions x 4 sessions) of a S's performance. A summary of these estimates is provided in Table 6. (The fact that no significant variation of these estimates occurs across sessions or within test lists again supports the argument that no interaction exists between item discriminability and experimental condition.)

Although pooled data show uniformity, individual data suggest that in some cases p(Y/s,k) and p(Y/n,k) are not monotonic with k. To demonstrate this, Spearman rank order correlations are calculated between p(Y/s,k) and k for k = 5,4,3,2, and between p(Y/n,k) and k for k = 3,2,1,0. (Other values of k produce so few observations as to preclude any claim to statistical reliability for them.) Now even if the relations between p(Y/s,k) and k, and between p(Y/n,k) and k were monotonic, sampling errors would produce on occasion imperfect correlations between these functions. It is possible to demonstrate that the imperfect correlations obtained in this experiment are not due to sampling errors but reflect unusual variations of the Ss' criterion placement, because there is a negative correlation between (1) the correlation between p(Y/s,k) and k, and (2) the correlation between p(Y/n,k) and k. Such a situation would not arise simply by sampling errors, but would arise, for example, if the S moved his criterion at a given level of discriminability in such a way as to *increase* p(Y/s) as discriminability *decreased*. This would have the effect of sharply increasing p(Y/n), and so in any sample of the S's responses, criterion placement by the S that makes the probability of p(Y/s,k) exceeding p(Y/s,k + 1) more likely, at the same time reduces the probability of p(Y/n,k') exceeding p(Y/n,k' + 1), where k' is a number corresponding for n to roughly the same level of discriminability as k corresponds to for s. Decrease in correlation between p(Y/s,k) and k should therefore be marked by increase in correlation between p(Y/n,k) and k. That the data for the "Yes" bias group have this property is most conveniently demonstrated by Table 7, which shows that extremely poor correlations between p(Y/s,k) and k (r_S < 0) are associated with high correlations between p(Y/n,k) and k (r_S ⩾ 0.8). This interaction is significant at the 2% level (using t test to test the effect over each individual S). Such a trend is apparent but fails to reach significance for the "No" bias group (Table 8). The two groups combined show a significant interaction beyond the 1% level (t test).

Table 7
Distribution of Spearman Rank Order Correlations (r_S) between p(Y/s,k) and k, and between p(Y/n,k) and k, for the "Yes" Bias Group

	r_S [p(Y/s,k),k] < 0	r_S [p(Y/s,k),k] ⩾ 0
r_S [p(Y/n,k),k] ⩾ 0.8	12	21
r_S [p(Y/n,k),k] < 0.8	0	15

The second way of demonstrating that p(Y/s) and p(Y/n) are not always monotonic with discriminability is to look at the number of times such a function as p(Y/s,k) exceeds p(Y/s,k + 1), given that in the samples over which this comparison is made, p(Y/s,5) has a fixed value. The reasoning behind this step is that if it is assumed that Ss in a given condition have essentially the same psychometric function (in terms of its general shape), though through differences in the discriminability of the memory trace and the individual S's response bias, different segments of the psychometric function are observed for each distinct sample of the Ss' responses, nevertheless these segments can be superimposed upon one another by anchoring a particular convenient point. That is, it is being assumed that the segments of the psychometric function obtained from different Ss for different sets of lists and parts of lists, but for all of which (say) p(Y/s,5) = 1, are sufficiently similar to make meaningful comparisons with other conditions. Support for this strong assumption comes from the consistent results obtained when this assumption is made. Table 9 shows the results of this procedure, using p(Y/s,5) and p(Y/n,0) as anchor points.

The important results shown by Table 9 are that, for the "Yes" bias group, the proportion of times that p(Y/s,4) exceeds p(Y/s,3) when p(Y/s,5) is large (=1) is significantly different from the proportion of times that p(Y/s,4) exceeds p(Ys,3) when p(Y/s,5) is small (<0.875) [t test on p(Y/s,4) vs p(Y/s,3), p < 0.01]; and for the "No" bias group, the proportion of times p(Y/n,2) exceeds p(Y/n,1) when p(Y/n,0) is small (=0) is significantly different from the corresponding proportion when p(Y/n,0) is of moderate size [0 < p(Y/n,0) ⩽ 0.143] and when p(Y/n,0) is large (>0.143) [t test on p(Y/n,1) vs p(Y/n,2), p < 0.01 and p < 0.05, respectively]. For no other conditions are there significant differences in proportions. These differences cannot be explained by assuming a monotonic psychometric function and supposing that differences in the number of times (say) p(Y/s,4) exceeds p(Y/s,3) is the result of sampling this psychometric function at points of different slope [the flatter the psychometric function, the less likely is it that a sample estimate of p(Y/s,4) will exceed p(Y/s,3)] because, as Table 9 shows, the mean level of p(Y/s,4) and p(Y/s,3) combined remains much the same, independent of the level of p(Y/s,5). So, if the psychometric function they belong to is monotonic, all samples have been taken around the same average point on this psychometric function, which invalidates this alternative explanation.

Note also that these data do not provide support for a U-shaped psychometric function: For the "Yes" bias group for the lower values of p(Y/s,5) [p(Y/s,5) < 1] (which values presumably correspond to the lower levels of the psychometric function) p(Y/s,3) is greater than p(Y/s,2) (t test over Ss, p < 0.02); this suggests that p(Y/s) reaches a peak for an intermediate level of discriminability. Similarly, for the "No" bias group, p(Y/n,2) has a tendency to be less than p(Y/n,1) p(Y/n,0) small (=0) and to be greater than p(Y/n,1) for p(Y/n,0) large (<0). Again this supports the "peak" explanation.

In summary, two sources of evidence suggest that the functions p(Y/s) and p(Y/n) are not monotonic with discriminability:

(1) Large departures of p(Y/s) from monotonicity in a

Table 9
p(Y/s,k) and p(Y/n,k) for Various Ranges of Values of p(Y/s,5) and p(Y/n,0)

"Yes" bias group	p(Y/s,5)=1	0.86≤p(Y/s,5)≤0.9	p(Y/s,5)<0.86
p(Y/s,4)≷p(Y/s,3)	11	7	6
	2	7	10
p(Y/s,3)≷p(Y/s,2)	5	10	11
	9	4	7
Mean values			
p(Y/s,4)	0.868	0.824	0.764
p(Y/s,3)	0.766	0.835	0.726
p(Y/s,2)	0.790	0.694	0.726
	0≤p(Y/n,0)≤0.143	0.143<p(Y/n,0)≤0.333	0.333<p(Y/n,0)
p(Y/n,2)≷p(Y/n,1)	13	12	12
	5	3	3
p(Y/n,3)≷p(Y/n,2)	10	6	11
	8	8	4
Mean values			
p(Y/n,1)	0.336	0.490	0.422
p(Y/n,2)	0.520	0.619	0.615
p(Y/n,3)	0.588	0.606	0.725
"No" bias group	0.8≤p(Y/s,5)≤1	0.625≤p(Y/s,5)<0.8	p(Y/s,5)<0.625
p(Y/s,4)≷p(Y/s,3)	7	8	4
	6	2	5
p(Y/s,3)≷p(Y/s,2)	8	6	8
	4	4	2
Mean values			
p(Y/s,4)	0.732	0.701	0.603
p(Y/s,3)	0.619	0.572	0.664
p(Y/s,2)	0.525	0.471	0.459
	p(Y/n,0)=0	0.111≤p(Y/n,0)≤0.143	0.143<p(Y/n,0)
p(Y/n,2)≷p(Y/n,1)	4	11	8
	8	0	1
p(Y/n,3)≷p(Y/n,2)	9	7	3
	3	3	5
Mean values			
p(Y/n,1)	0.270	0.111	0.224
p(Y/n,2)	0.285	0.371	0.430
p(Y/n,3)	0.422	0.386	0.401

sample of a S's responses are accompanied by a decreased probability of nonmonotonicity appearing in the corresponding sample of p(Y/n). This suggests that such departures from monotonicity are not merely sampling errors.

(2) The proportion of times p(Y/s,k) exceeds p(Y/s,k + 1) [or p(Y/n,k) exceeds p(Y/n,k + 1)] varies as a function of p(Y/s,5) [or p(Y/n,0)], and this fits in better with the concept of p(Y/s) or p(Y/n) as being a function looking like a cubic curve (i.e., some nonmonotonic function) rather than like the cumulative normal curve (i.e., a standard monotonic function).

DISCUSSION

Finally, some more strong assumptions will be made to derive the form of the psychometric functions. These assumptions are not to be regarded as candidates for the final version of any complete analysis of this experiment, but as first approximations that enable an intelligent guess to be made at the underlying form of the results.

(1) For all Ss in a given condition, the same general psychometric function is valid and individual psychometric functions differ only in their rate of change with stimulus intensity (i.e., any psychometric function in a given condition can be changed into another in the same condition by a *linear* transformation of the stimulus intensity axis).

(2) The stimuli (s,5), (s,4), (s,3), (s,2) are equally spaced on the stimulus intensity axis. (s,k) is the stimulus consisting of a signal to which k members of the unbiased group have responded "Yes." Similarly, (n,0), (n,1), (n,2), (n,3) are equally spaced.

(3) If p(Y/s,5) = 1, the portion of p(Y/s) joining p(Y/s,5) to p(Y/s,4) is a straight line. Similarly, if p(Y/n) = 0, the portion of p(Y/n) joining p(Y/n,0) to p(Y/n,1) is a straight line.

With these assumptions, it is possible to use all the data for which p(Y/s,5) = 1 or p(Y/n,0) = 0 to derive four functions:

(1) p(Y/s) for the "Yes" bias group
(2) 1 − p(Y/n) for the "Yes" bias group
(3) p(Y/s) for the "No" bias group
(4) 1 − p(Y/n) for the "No" bias group

These are shown in Fig. 4.

In line with prediction, Functions 2 and 3 are monotonic and Functions 1 and 4 are nonmonotonic; these latter functions depart from the "normal" functions 2 and 3 for low stimulus intensity. Function 4 departs from the normal psychometric function at a stimulus intensity lower than that for Function 1. This
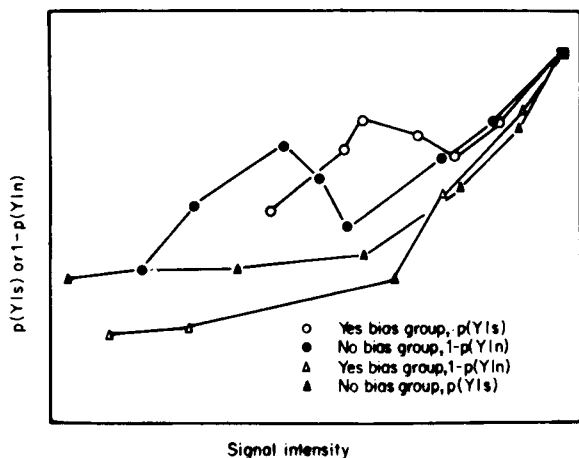
Fig. 4. Psychometric functions derived from the experimental data using the assumptions outlined in the text. The data points in this figure are derived by a moving average technique from the raw data.

is to be expected, since the tendency among the "No" bias group of Ss was to be less biased than the "Yes" bias group of Ss, and the discussion of the cost concept in the introduction to this paper suggests that if the overall bias of a S is relatively small, the point where he changes from an extreme criterion for low discriminability to a moderate criterion for high discriminability will occur at a point of relatively low discriminability. Thus the point of departure from the "normal" psychometric function occurs at a correspondingly lower value of stimulus intensity.

One comment on these results is in order: small samples of individual data in this experiment show wide variations, but some at least of this variance has been shown to be due to unusual psychometric functions and not simply to sampling errors; in contrast pooled data show much greater uniformity. In the past it has been usual in signal detection experiments to obtain very large numbers of responses from Ss to ensure statistical reliability [e.g., in their classic paper, Swets, Tanner and Birdsall (1961) never used less than 60 responses for a single estimate of p(Y/s)], and the data so obtained have indeed shown uniformity and reliability. However, in the present experiment the assumption that the S is operating with discriminability and response bias parameters that change only slowly over time is not supported by samples of small chunks of data, even if larger samples appear more uniform. It is suggested therefore that the E carrying out signal detection experiments must be more conscious of this experimental dilemma, i.e., whether to pool large samples of data and risk obscuring short-term variations in the S's discriminability and response bias parameters, or to collect small samples and risk poor statistical reliability. The results of this paper suggest the latter alternative may be the lesser of the two evils.

This point is well illustrated by the work of Kintsch (1968), who carried out a recognition memory experiment where Ss were required to give confidence ratings with their responses. His graphs of p(Y) against confidence rating are not monotonic (Kintsch, 1968, Fig. 2), though he remarks that "pooling some of the rating categories for each S would remove the noise from the data" [my italics]. In fact, the data probably are not very noisy, as the following analysis shows. The graph of p(Y) against confidence rating can be turned into a psychometric function by assuming the confidence ratings represent different levels of *discriminability* (not different levels of *response bias*, as is generally assumed), and plotting p(Y/s) against confidence rating for the portion of p(Y/s) above 0.5. Overall response bias can be estimated by

calculating p(Y) for those responses the S rated as a "pure guess," counting a S as biased to "Yes," "No," or unbiased, according as p(Y/"pure guess") is greater than, less than, or equal to 0.5. Of Kintsch's four Ss, two are biased to "Yes" and show nonmonotonicity in p(Y/s) against s, one is biased to "No" and shows nonmonotonicity in 1 − p(Y/n) against n, and one is unbiased and has monotonic psychometric functions (see Fig. 5). All these results are predictions of the model presented in this paper. The monotonic psychometric function for an unbiased S is, in terms of our model, a consequence either of the S being able to tolerate high cost and thus not having to move his criterion from an unbiased position, or of the S achieving a net unbiased state by selecting a "Yes" bias criterion or a "No" bias criterion at random, and changing this from trial to trial.

The interpretation of these two experiments in terms of our decision-making model requires an unusual level of knowledge from Ss; for each item, the S must know how discriminable the item is *before* he makes a decision as to whether it is a positive or negative stimulus, in order that he can position his criterion appropriately. This problem is parallel to the difficulty encountered in theories of attention (Broadbent, 1958; Treisman, 1964) where a S must know whether a message is relevant or not before he can decide whether it needs attention; in both cases it seems that the process of identifying the stimulus must be carried out before the process of identifying the stimulus can begin. The solution proposed for the problem in attention is *filter theory*; partial preliminary information can be obtained through a filter before the complete process of identification is brought into action. The same process is assumed to apply in this paper—the S, by preliminary processing, obtains an estimate of the stimulus' discriminability in order to place his criterion appropriately for a more accurate detection process.

REFERENCES

BROADBENT, D. E. *Perception and communication.* London: Pergamon Press, 1958.

BROADBENT, D. E. Word-frequency effect and response bias. Psychological Review, 1967, 74, 1-15.

BROADBENT, D. E., & GREGORY, M. Vigilance considered as a statistical decision. British Journal of Psychology, 1963, 54, 309-323.

CLARKE, F. R. Confidence ratings, second-choice responses and confusion matrices in intelligibility tests. Journal of the Acoustical Society of America, 1960, 32, 35-46.

GREEN, D. M. Psychoacoustics and detection theory. Journal of the Acoustical Society of America, 1960, 32, 1189-1203.

HULL, G. L., HOVLAND, C. I., ROSS, R. T., HALL, M., PERKINS, D. T., & FITCH, F. B. *Mathematico-deductive theory of rote learning.* New Haven: Yale University Press, 1940.

KINTSCH, W. An experimental analysis of single stimulus tests and multiple-choice tests of recognition memory. Journal of Experimental Psychology, 1968, 76, 1-6.

MURDOCK, B. B., JR. The criterion problem in short-term memory. Journal of Experimental Psychology, 1966, 72, 317-324.

NORMAN, D. A., & WICKLEGREN, W. A. Short-term recognition memory for single digits and pairs of digits. Journal of Experimental Psychology, 1965, 70, 479-489.

POLLACK, I., & DECKER, L. R. Confidence ratings, message reception, and the receiver operating characteristic. Journal of the Acoustical Society of America, 1958, 30, 286-292.

SWETS, J. A., TANNER, W. P., JR., & BIRDSALL, T. G. Decision processes in perception. Psychological Review, 1961, 68, 301-340.

SMITH, P. T. Cost, discriminability and response bias. British Journal of Mathematical and Statistical Psychology, 1968, 21, 35-60.

TREISMAN, A. M. The effect of irrelevant material on the efficiency of selective listening. American Journal of Psychology, 1964, 77, 533-540.

WICKLEGREN, W. A. Consolidation and retroactive interference in short-term recognition memory for pitch. Journal of Experimental Psychology, 1966, 72, 250-259.

WICKLEGREN, W. A., & NORMAN, D. A. Strength models and serial position in short-term recognition memory. Journal of Mathematical Psychology, 1966, 3, 316-347.
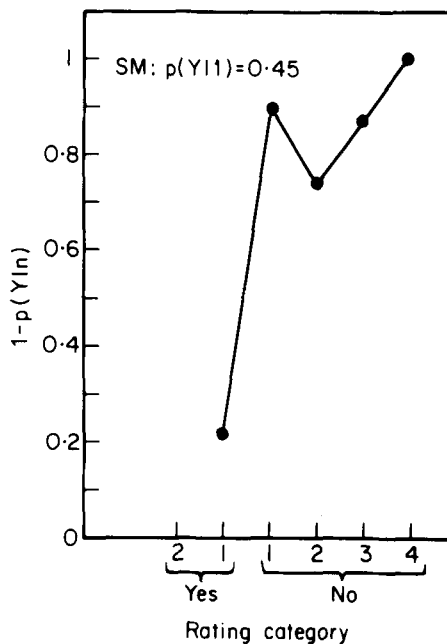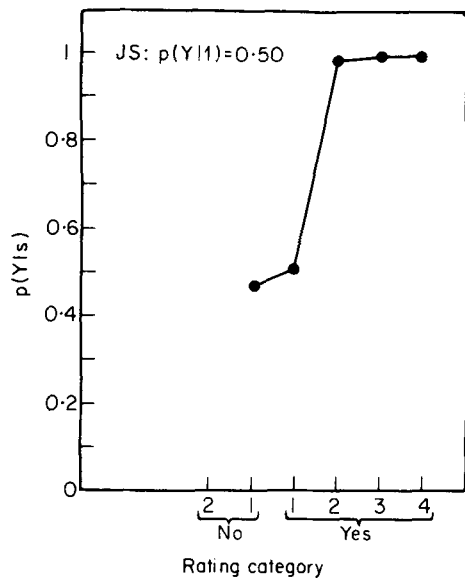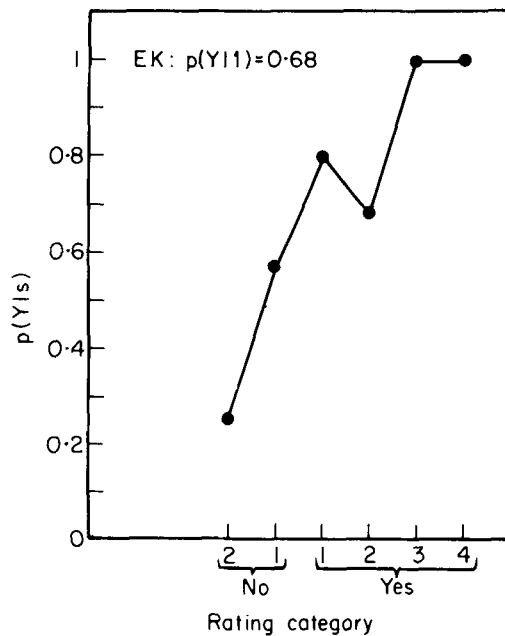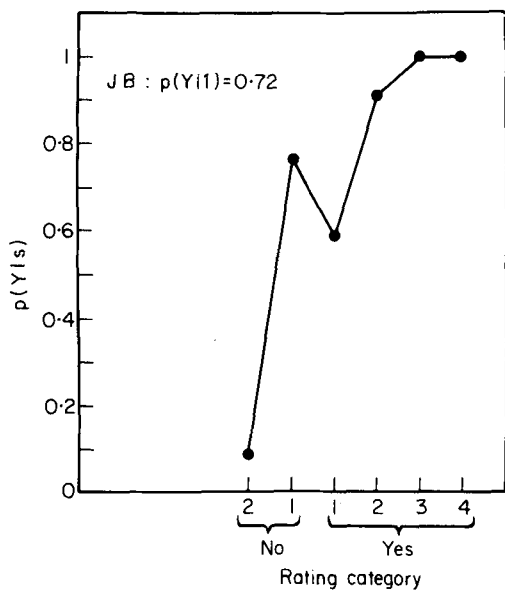
Fig. 5. Psychometric functions for the Ss taking part in Kintsch's (1968) experiment. The curves are adapted from Fig. 2 of Kintsch's study.