A Bayesian prediction of four-look recognition performance from one-look data

MICHAEL E. DOHERTY[1] AND STUART M. KEELEY
BOWLING GREEN UNIVERSITY

The hypothesis that a human O's (S's) performance in a visual recognition task can be modelled by Bayes' theorem was investigated. Two Ss were run for 40 experimental sessions each. Their task was to specify the direction of the gap of tachistoscopically presented Landolt rings (Cs). There were four possible gap directions, and two experimental conditions. In one condition, S responded after each stimulus presentation. In the other, a fixed-observation condition, Ss responded after four consecutive presentations of a C. Exposure durations were such that performance under both conditions was greater than chance, but less than unity. Predictions of four-look performance from one-look data were made. Overall hit rates were predicted closely. The entire pattern of each S's four-look data was also predicted reasonably well. Further tests of the model are currently under way.

Keeley and Doherty (1968) tested three models which describe how an O (S) combines independent pieces of evidence about some state of the world in arriving at some decision about that state. The models are Eriksen's (1966) clearest state model, and an integration model and decision-threshold model as discussed in Green and Swets (1966). Using briefly presented visual inputs, the Es made predictions of repeated observation hit rates (HRs) from single observation HRs. None of the models described Ss' performance adequately, especially in Experiment 1 which employed unidimensional stimuli.

The basic datum which enters into the predictions of all three models is the single observation HR. In other words, the models as we tested them are sensitive to neither differential HRs on the several stimuli nor to nonrandom patterns of errors. In effect, the models assume that perceptual errors are random. This is most clearly seen in the decision threshold model, which explicitly states that on those trials on which S is incorrect, the visual input has been "subthreshold."

If one wishes to predict S's multiple-observation decision behavior, there are compelling reasons for using the entire pattern of single observation responding. Consider an experiment in which S's task is to identify which one of the three graphemes A, O, or U has been briefly presented. The latter two forms appear to be more "confusable" with each other than either of them is with the former. Suppose that a given S tended to "see" Os as Us more often than the reverse. If we now give that S n looks before he must respond, it is very unlikely that his multiple observation performance would be predicted closely by any model using only his overall HR. Furthermore, even a model which takes into account differential HRs on different stimuli would likely be inadequate, unless it also had built into it a means of using the information about the kinds of errors S was expected to make on any given single observation. Putting the same argument into another context, S's probability (P) of a response given n observations is a function of the single observation Ps of all available responses conditional upon the various stimuli. While the argument is clear with the stimuli selected for illustration above, it applies generally to any multiple-observation prediction situations. A model which would describe an S's performance in the sort of situation outlined would likely be appropriate in a situation in which errors are random as well.

The emphasis on S's errors reflects the assumption that S is not responding on the basis of subthreshold (or unavailable) inputs when he makes errors, but is responding frequently on the basis of erroneous information. In other words, he is making "confusion errors." If the latter is the case, Ss would sometime respond erroneously *with certainty,* which they did in Keeley and Doherty (1968). If the stimuli employed were not all equally identifiable, the Ss would have different HRs to the different stimuli. And if the stimuli were differentially confusable with one another, the Ss would have more or less marked departures from randomness in their distributions of errors. While there were insufficient trials in Keeley and Doherty (1968) to support these latter two suppositions, the available data led us to believe that such was indeed the case. Models which take errors into account allow for a suppression in predicted HR which would reflect the behavior one would expect if an S has a high P of making a particular confusion error. Conversely, if our reasoning is correct, models disregarding errors should overpredict. In Keeley and Doherty (1968, Experiment 1) both the Integration and Decision-threshold models overpredicted, the latter for every S.

These considerations led Es to look at a model incorporating considerably more of the single-observation data into the multiple-observation prediction, that model being Bayes' rule. In order to obtain enough data to make estimates of the values of the required conditional Ps, two Ss from Experiment 1 were continued for many more trials in the single and successive observation conditions.

METHOD

Subjects

Two males from Experiment 1 of Keeley and Doherty (1968) served as Ss.

Apparatus

The apparatus and stimuli are described fully in Keeley and Doherty (1968). Briefly, a Scientific Prototype Model GB tachistoscope, with a handswitch permitting S to initiate stimulus presentations, was used. The stimuli were Landolt rings (Cs) with the gaps either right, left, up, or down (R, L, U, or D). A given stimulus would have one C on a corner of an imaginary square centered on a fixation point which was on at all times.

Procedure

Two of the conditions described in the 1968 Experiment 1 were used. In the single observation condition (1C) the Ss were presented with one stimulus and responded with the perceived direction of gap (r, l, u, or d). In the successive condition (1C4), Ss were presented with four successive stimuli with the gaps in the same direction, but with the locus of the C varying randomly. The Ss responded only after the fourth stimulus, qualifying this condition as an instance of what Green and Swets (1966) call the "fixed observation procedure." Exposure durations were such that Ss' HRs exceeded chance but were less than unity in both conditions.

There were 40 test observations per experimental session, either 40 1C trials or 10 1C4 trials. Each S ran 40 experimental sessions, 12 1C and 28 1C4. Thus there were 480 1C responses available for purposes of prediction, and 280 1C4 responses for comparison with the predicted frequencies. (Due to an omission, only 478 1C trials were recorded for one S.) Since the Ss were becoming highly practiced, the restriction that an equal number of each of the four gap directions occur five times in each block of 20 trials was no longer observed. The only other procedural difference was that confidence judgments were no longer required of the Ss.

A Bayesian Model

It was assumed that S made an implicit response of r, l, u, or d to each of the four observations on a 1C4 trial, and that the P of an implicit response could be estimated by relative frequencies of the 1C condition. It was further assumed that S would, at the conclusion of the fourth observation of a 1C4 trial, be in a *perceptual state* determined by some combination of the available implicit responses, and that his decision would be determined by that perceptual state. It follows that there is a finite set of possible perceptual states (Bi), consisting of the 4^4 combinations which may result from four presentations of one of four stimuli (Aj).

Briefly, the prediction of the 1C4 results from the 1C data matrix can be considered as involving four steps: (1) The 256 potential perceptual states are listed. (2) The probability of each of these states contingent on each of the four possible stimulus states, P(Bi | Aj), is determined. (3) The most probable state of the world given each particular perceptual state, P(Aj | Bi), is determined by the use of Bayes' theorem. (4) The P(Bi | Aj) values are summed separately for each S for each of the four predictions for each of the four actual stimululus states, which summations yield a matrix of predicted decisions.

Let us consider Steps 2 through 4 in more detail:

In Step 2, there are 1024 separate P(Bi | Aj)s to estimate, 256 for each of the four stimulus states. The calculations employ the relative frequencies from the 1C condition (see Table 1). Each P(Bi | Aj) value is the product of four conditional P values estimated from these relative frequencies. For example, if the B state under consideration is the set of four implicit responses rrlu, the P of occurrence of that perceptual state given the stimulus state of four right gaps is given by:

$$P(rrlu \mid RRRR) = P(r \mid R)^2 \cdot P(l \mid R) \cdot P(u \mid R)$$

and, given four left gaps:

$$P(rrlu \mid LLLL) = P(r \mid L)^2 \cdot P(l \mid L) \cdot P(u \mid L)$$

etc. These calculations are obviously predicated upon the assumption that the four implicit responses are independent. Under this assumption, which renders order irrelevant, there are actually only 35 different values of $P(B_i \mid A_j)$ for each of the four A_j.

In Step 3, for each of the 256 potential perceptual states,

Table 1
Frequency of Each Response to Each Stimulus for the 1C Conditions

	Stimuli							
	Subject 1				Subject 2			
Response	R	L	U	D	R	L	U	D
r	64	14	21	17	65	12	10	12
l	18	63	18	25	25	79	16	21
u	20	25	69	21	18	15	76	24
d	16	17	14	56	13	13	18	63

Table 2
Empirical and Predicted (r′, l′, u′, d′) Frequencies of Each Response to Each Stimulus State for the 1C4 Condition

	Stimulus State							
	Subject 1				Subject 2			
Response	R	L	U	D	R	L	U	D
r	49	8	8	13	52	6	6	1
r′	43	6	7	8	54	4	2	5
l	5	52	8	9	11	52	4	14
l′	6	53	8	10	8	57	6	6
u	5	13	49	3	7	7	52	9
u′	8	10	53	7	4	3	51	8
d	3	5	8	42	2	5	7	45
d′	5	9	5	42	6	6	10	50

the P of each of the four stimulus states was calculated according to Bayes' theorem (e.g., Hayes, 1963, p. 116, Eq. 4.5.2):

$$P(A_j \mid B_i) = \frac{P(B_i \mid A_j)P(A_j)}{P(B_i \mid A_1)P(A_1) + P(B_i \mid A_2)P(A_2) + P(B_i \mid A_3)P(A_3) + P(B_i \mid A_4)P(A_4)} \quad (1)$$

but since the particular experimental conditions are such that the a priori $P(A_j)$ values are all .25, Eq. 1 simplifies to:

$$P(A_j \mid B_i) = \frac{P(B_i \mid A_j)}{P(B_i \mid A_1) + P(B_i \mid A_2) + P(B_i \mid A_3) + P(B_i \mid A_4)} \quad (2)$$

The decision rule which seems most reasonable is that the stimulus with the greatest a posteriori probability would be selected as the response, or decision, in light of the available evidence (i.e., B_i). Therefore, the A_j corresponding to the highest of the four values of $P(A_j \mid B_i)$ was S's predicted response. In those few circumstances in which $P(A_j \mid B_i)$ values were equal, the predictions were divided among the states accordingly.

In Step 4, the predicted response probabilities for the four-look condition were obtained separately for each S by summing the appropriate P values:

$$P(A_j \mid B_k) = \Sigma P(B_i \mid A_j), \quad (3)$$

where the summation is taken, separately for each j, over those perceptual states for which the decision rule defined above leads to a given outcome. The subscript k indexes the four response outcomes possible for each j. This procedure yields a j by k (in this case 4 by 4) matrix of conditional P values, for which

$$\Sigma P(A_j \mid B_k) = 1.0, k = 1, \cdots 4, \quad (4)$$

for each $j = 1, \cdots 4$. The predicted frequencies of Table 2 were obtained by multiplying the actual frequencies of the four stimulus states by the calculated $P(A_j \mid B_k)$ values.

RESULTS

The overall HRs were predicted remarkably well for S 1 with the actual HR being .686 and the predicted .682, and rather well for S 2, the HRs being .718 and .757, respectively. These predictions compare favorably with those of the

Decision-Threshold model, .882 and .932, which are, as expected, radical overpredictions. The results of primary interest, however, are the predictions of each S's total pattern of results. An inspection of Table 2 reveals that these predictions are reasonably close to the empirical frequencies. The major exception is the model's failure to predict the disproportionate number of left responses given by S 2 to the down stimuli.

DISCUSSION

The closeness of the predictions to the empirical frequencies is remarkable when one considers that the data are based on a procedure which was neither originally designed to test the Bayesian model, nor is an optimal procedure to do so. The very least that can be said for the results is that they are very encouraging. They have prompted Es to commence an experiment which employs the major operation lacking in the present research which makes Bayesian analysis conceptually much more appropriate. That operation is knowledge of results in the 1C condition. While the Es did not provide trial-by-trial feedback in the present experiment, the Ss had considerable inferential knowledge of the veridicality of their perceptions due to the nature of the 1C4 condition, and some direct knowledge due to the procedure of informing Ss of their overall performance at the termination of each block of 10 trials. The accuracy of S's assessments of the $P(B_i \mid A_j)$ values is open to question. Clearly this is a variable which ought to be under E's control if he wishes to perform an optimal test of a model which assumes that Ss are combining these conditional probabilities. Furthermore, a much more refined prediction of S's multiple-observation performance could be made if confidence judgments were recorded and taken into account. They would enter into the prediction by considering different confidence ratings as different responses. Since this would increase the size of the data matrices by a factor equal to the number of levels of confidence permitted, a very large number of trials would be necessary.

The usual assumptions which must be considered before applying Bayes' theorem seem to Es to be fully met. The P values all admit of relative frequency interpretations, and the a priori probabilities are not only known, they are under experimental control. This particular application of the rule is predicated upon another assumption, however, concerning which we have no information. That assumption is that Ss' decision criteria are either constant over time over the course of the experiment, or vary randomly about some constant value. When using stimuli in the visual mode in appropriate HR ranges, at least with the kind of experimental conditions in this study, the attentional demands upon S are so severe that relatively little data can be collected in a session. Change over sessions is thus relatively difficult to detect.

Green and Swets (1966) dichotomize decision models in terms of whether the Ss are postulated to be accumulating and combining *information* (the "integration" models) or are somehow basing an ultimate "decision" on the basis of other decisions (the "decision-threshold" models). It does not seem necessary at this time to the present authors to choose between these categories of decision rules. That question concerns the nature of the coding and storage mechanisms, and S may encode and store information or he may encode and store *a set* of decisions about an input. Either process can be represented equally well by conditional probabilities; S's performance may well be modelled by Bayes' rule, whichever he is doing.

REFERENCES

ERIKSEN, C. W. Independence of successive inputs and uncorrelated error in visual form perception. Journal of Experimental Psychology, 1966, 72, 26-35.

GREEN, D. M., & SWETS, J. A. *Signal detection theory and psychophysics.* New York: Wiley, 1966.

KEELEY, S. M., & DOHERTY, M. E. Simultaneous and successive presentations of single-featured and multi-featured visual forms: Implications for the parallel processing hypothesis. Perception & Psychophysics, 1968, 4, 296-298.

SWETS, J. A., & BIRDSALL, T. G. Deferred decision in human signal detection: A preliminary experiment. Perception & Psychophysics, 1967, 2, 15-28.

NOTE

1. Address: Department of Psychology, Bowling Green State University, Bowling Green, Ohio 43402.

(Accepted for publication January 6, 1969.)