Bulletin of the Psychonomic Society
1976, Vol. 8 (1), 1-3

Feedback effects in a metric multiple-cue probability learning task

R. JAMES HOLZWORTH
Ohio Northern University, Ada, Ohio 45810

MICHAEL E. DOHERTY
Bowling Green State University, Bowling Green, Ohio 43403

Three forms of augmented feedback were compared with simple outcome feedback in a computer-run, metric, multiple-cue probability learning task. These were: (a) lens model feedback, (b) provision of the predicted criterion value, and (c) provision of categorical information about the criterion. There was significant improvement in performance across six blocks of 25 trials, but no differences among the feedback treatments. In a second study, two additional groups were run without outcome feedback. These groups received either lens model feedback or predicted criterion feedback. The major result was a sharp increase in consistency, which mediated an increase in achievement, when outcome feedback was not provided.

A number of recent studies comparing different types of feedback in metric multiple-cue probability learning tasks have found that subjects given only outcome feedback learn less rapidly than subjects given cognitive feedback (Deane, Hammond, & Summers, 1972; Hammond, 1971; Hammond & Boyle, 1971; Todd & Hammond, 1965).

Outcome feedback refers to the operation of providing subjects with the value of the criterion on each trial immediately after the subject's prediction of the criterion. Cognitive feedback refers to the operation of providing subjects with some lens model components (such as those described below) after each block of trials, block size being generally from 10 to 25 trials. Both forms of feedback, of course, may be provided in a given study. Experiment I investigated different methods of augmenting traditional outcome feedback in a multiple-cue probability learning situation.

EXPERIMENT I

Method

Subjects. Forty introductory psychology students served as subjects to fulfill a course requirement.

Apparatus. Cue values and feedback were presented to individual subjects on a ComData Series 33 terminal Teletype operating within IBM Call/360-OS. The subjects read the cue values from the Teletype, then typed their predictive judgments on the Teletype keyboard. Keypunching errors were excluded from the data. These were defined as responses outside the range of the criterion, and these were fewer than 1%. A shield permitted subjects to see only the information from three immediately preceding trials.

Design. A split plot factorial (four feedback conditions by six blocks of 25 trials) design was employed. Ten subjects were randomly assigned to each of four feedback conditions. Each subject was presented the same metric multiple-cue probability learning task with three cues and a criterion. One hundred and fifty-three judgments were required of each subject in a single 2-h session. The first three trials were practice and were not analyzed.

Task characteristics. The relationship of each cue to the criterion was positive and linear. The mean of each cue was 50, that of the criterion 200. The standard deviation of each cue was 5, that of the criterion 15. All cue and criterion values were whole numbers. The ecological validities of the cues were .551, .384, and .213. Cues were uncorrelated with each other; the highest intercorrelation was −.046. The multiple correlation between the cues and criterion in each block was .71 ± .01.

Feedback conditions. In the first experiment, all subjects were given outcome feedback on each trial and also r_a^2, the squared correlation of their predictions with the criterion values, after each trial block. Four experimental treatments were defined by type of augmented feedback. Group O was given no additional information. Group O+L was provided the following lens model feedback: (a) relative weights (Hoffman, 1960) computed on the environment and on their own predictions; (b) R_s^2, a measure of the extent to which their predictions were linear composites of the cues; and (c) G^2, the square of the correlation between the least squares predictions of their responses and the least squares predictions of the criterion values. These indices are fully described elsewhere (e.g., Dudycha & Naylor, 1966; Hammond & Summers, 1972). Group O+C received a categorical value, a quartile, of the criterion along with the numerical outcome feedback on every trial. Group O+P received the optimally predicted criterion value (computed using the multiple regression equation based on all 150 trials) along with outcome feedback on every trial.

All subjects were instructed on the procedure for interacting with the computer, and were informed that there would be six blocks of 25 trials each. On a sheet developed separately for each feedback group, subjects were further informed that the average of each cue would be 50, and that of the criterion 200. All terms used in the particular feedback conditions were defined as simply as possible for the subjects.

Results and Discussion

The dependent variables in this study are r_a, G, and R_s. These indices were computed on each of six blocks of trials for each subject. Prior to analysis, each index was transformed on these repeated measures across six blocks of trials (Finn, 1972).

Table 1
Mean Values of r_a, G, and R_s by Trial Blocks for All Groups in Experiments I and II

	Group	Trial Block 1	2	3	4	5	6
r_a	O	.22	.38	.39	.42	.39	.54
	O + C	.34	.44	.49	.50	.51	.57
	O + L	.27	.37	.45	.42	.46	.55
	O + P	.45	.56	.58	.46	.51	.58
	L	.57	.50	.56	.58	.60	.64
	P	.45	.57	.58	.55	.63	.58
G	O	.72	.92	.94	.94	.91	.97
	O + C	.77	.93	.96	.96	.91	.96
	O + L	.85	.92	.90	.94	.92	.95
	O + P	.90	.96	.97	.95	.91	.96
	L	.92	.86	.90	.93	.98	.97
	P	.90	.92	.95	.95	.94	.94
R_s	O	.63	.68	.73	.75	.73	.78
	O + C	.64	.76	.75	.81	.85	.85
	O + L	.60	.77	.82	.85	.84	.83
	O + P	.73	.78	.86	.85	.84	.87
	L	.86	.93	.93	.95	.92	.94
	P	.81	.88	.92	.94	.94	.91

Mean values of r_a, G, and R_s obtained in Experiments I and II are presented in Table 1.

The multivariate analysis revealed no significant Groups by Blocks interaction. There were, however, significant increases in r_a, G, and R_s across blocks, multivariate $F(15,22) = 4.11$, $p < .002$. Univariate F ratios for all three indices were also highly significant. Differences between feedback conditions were not demonstrated.

The failure to demonstrate differences between feedback conditions may be due to the possibility that the effects of lens model, predicted criterion, and categorical feedback were degraded by outcome feedback. Although Todd and Hammond (1965) found no significant difference between lens model feedback and lens model feedback plus outcome feedback, Hammond, Summers, and Deane (1973) reported a detrimental effect of outcome feedback in multiple-cue probability learning.

EXPERIMENT II

Method

In order to determine the effects of predicted criterion feedback and lens model feedback without the potentially degrading effects of outcome feedback, two additional groups were run.

Subjects. Eighteen introductory psychology students served as subjects in order to fulfill a research participation requirement.

Apparatus, task characteristics. These were the same as in Experiment I.

Feedback conditions. In Group L, eight subjects were provided with the same feedback as in Group O+L, except that the criterion value was not provided on any trials. Ten subjects in Group P received the same feedback as in Group O+P, save the criterion values. Note that Group P was, for the subjects, a task with perfect environmental predictability. The experimenters computed r_a with the actual criterion values. This was done so that the value of r_a^2, presented at the end of each block, would have the same order of magnitude and the values of r_a in the analysis would be influenced by the same amount of environmental uncertainty as in Experiment I.

Results

Values of r_a, G, and R_s for each block of trials were computed and transformed to Fisher's Z coefficients. A multivariate trend analysis for group differences was then performed (Finn, 1972). There was no significant Groups by Blocks interaction and no significant difference between Groups L and P. However, there were significant increases in r_a, G, and R_s across blocks, multivariate $F(15,2) = 54.85$, $p < .02$. Univariate F ratios for all three indices were also significant. Thus, subjects in the two conditions without outcome feedback did learn the task.

Mean values of r_a were slightly higher for the L and P groups than for all Experiment I groups, but there appeared to be no systematic difference in G, which was already quite high. The effect on R_s was dramatic, however, as is shown in Table 1. The L and P groups were far more consistent than the other four groups. Significance tests were not performed between groups in Experiments I and II, since subjects in Groups L and P were sampled subsequently to the first four groups, and random assignment could not be accomplished.

DISCUSSION

The superior performance of the L and P groups over the O+L and O+P groups supports the contention of Hammond, Summers, and Deane (1973) that outcome feedback has a detrimental effect in multiple-cue probability learning. That effect is to sharply reduce policy consistency. The fact that there was no substantial difference between lens model and predicted criterion feedback may make the latter form of feedback, in which the tendency to "chase error" is removed, an informative treatment in future learning studies.

Finally, the performance of subjects on Block 1 in the L group, where subjects had *no* knowledge of task characteristics (other than means of the cues and criterion) until *after* the first 25 trials, indicates that they came into the laboratory with a response bias favoring positive linear correlations. Brehmer (1974) and Brehmer, Kuylenstierna, and Liljergren (1974) have found that subjects have a hierarchy of hypotheses about functional relations between scaled variables, and that they sample from this hierarchy (positive linear > negative linear > inverted U-shaped > U-shaped functional relationship) when learning a single-cue probability learning task. This suggests that, for unconfounded demonstrations of *learning* in multiple-cue probability tasks, tasks containing only positive linear relationships should not be used.

REFERENCES

Brehmer, B. Hypotheses about relations between scaled variables in the learning of probabilistic inference tasks. *Organizational Behavior and Human Performance*, 1974, **11**, 1-27.

Brehmer, B., Kuylenstierna, J., & Liljergren, J. E. Effects of function form and cue validity on the subjects' hypotheses in probabilistic inference tasks. *Organizational Behavior and Human Performance*, 1974, **11**, 338-354.

Deane, D. H., Hammond, K. R., & Summers, D. A. Acquisition and application of knowledge in complex inference

tasks. *Journal of Experimental Psychology*, 1972, **92**, 20-26.
DUDYCHA, A. L., & NAYLOR, J. C. Characteristics of the human inference process in complex choice behavior situations. *Organizational Behavior and Human Performance*, 1966, **1**, 110-128.
FINN, J. D. *Multivariance: Univariate and multivariate analysis of variance, covariance, and regression. A Fortran IV Program, Version V*. Ann Arbor: National Education Resources, Inc., 1972.
HAMMOND, K. R. Computer graphics as an aid to learning. *Science*, 1971, **172**, 903-908.
HAMMOND, K. R., & BOYLE, P. J. R. Quasi-rationality, quarrels and new conceptions of feedback. *Bulletin of the British Psychological Society*, 1971, **24**, 103-113.
HAMMOND, K. R., & SUMMERS, D. A. Cognitive control. *Psychological Review*, 1972, **79**, 58-67.
HAMMOND, K. R., SUMMERS, D. A., & DEANE, D. H. Negative effects of outcome-feedback in multiple-cue probability learning. *Organizational Behavior and Human performance*, 1973, **9**, 30-34.
HOFFMAN, P. J. The paramorphic representation of clinical judgment. *Psychological Bulletin*, 1960, **47**, 116-131.
TODD, F. J., & HAMMOND, K. R. Differential feedback in two multiple-cue probability learning tasks. *Behavioral Science*, 1965, **10**, 429-435.

(Received for publication November 28, 1975.)