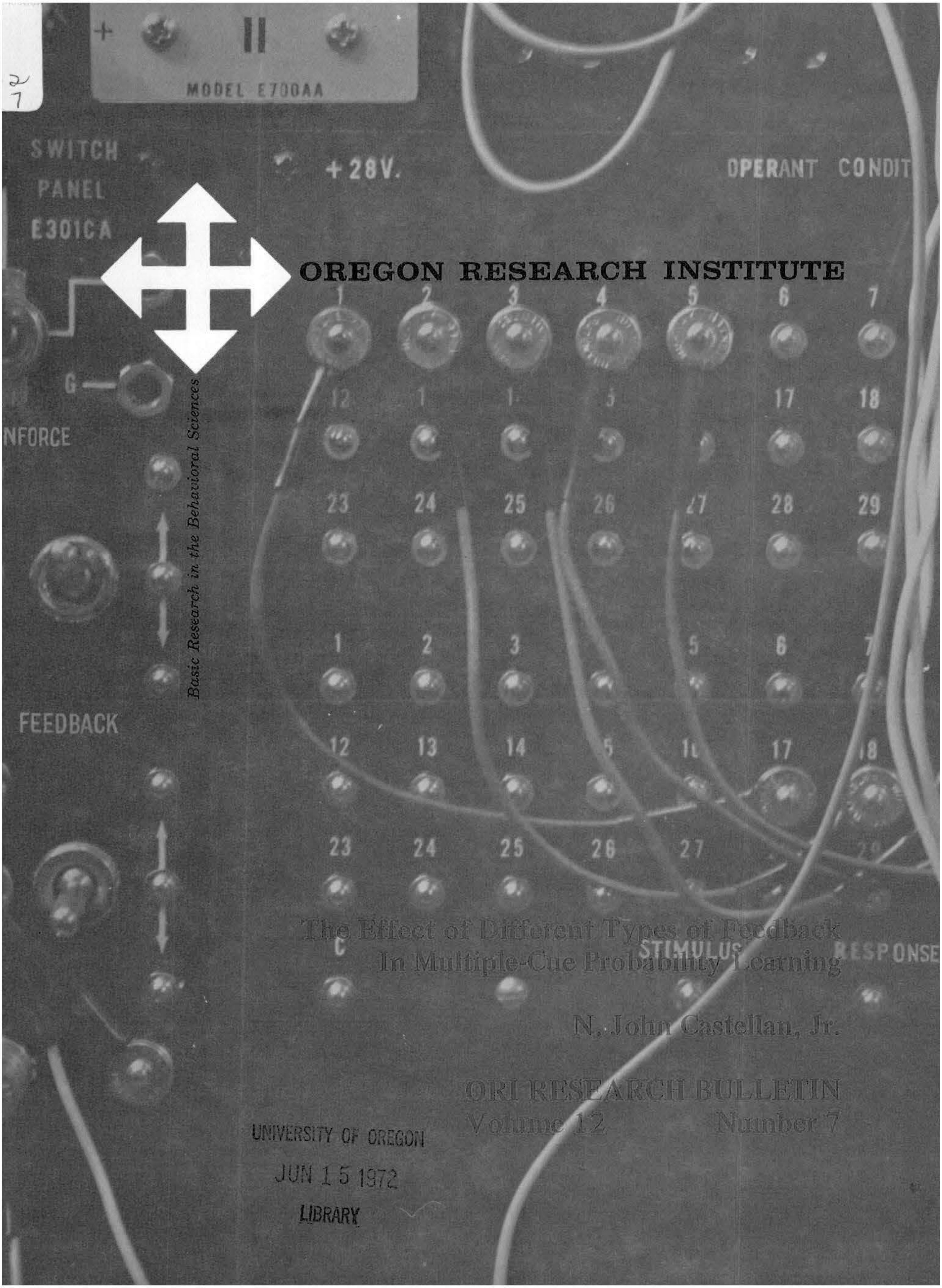

OREGON RESEARCH INSTITUTE

Basic Research in the Behavioral Sciences

The Effect of Different Types of Feedback In Multiple-Cue Probability Learning

N. John Castellan, Jr.

ORI RESEARCH BULLETIN
Volume 12 Number 7

OREGON RESEARCH INSTITUTE

THE EFFECT OF DIFFERENT TYPES OF FEEDBACK IN MULTIPLE-CUE PROBABILITY LEARNING

N. John Castellan, Jr.

Indiana University

Preparation of this report was supported in part by Grant No. MH 12972 from the National Institute of Mental Health, U. S. Public Health Service.

ORI RESEARCH BULLETIN

Vol. 12 No. 7

April 1972

Also published as: Indiana Mathematical Psychology Program

Report No. 72-2

THE EFFECT OF DIFFERENT TYPES OF FEEDBACK IN MULTIPLE-CUE PROBABILITY LEARNING

N. John Castellan, Jr.

Indiana University

Abstract

The purpose of this study was to determine the effect of different sorts of feedback upon performance in a multiple-cue probability learning task with binary cues and binary events. In addition to traditional outcome feedback, subjects were given periodic feedback messages throughout the experiment. Different types of feedback included percentage of correct responses, cue-event validity coefficients, cue-response utilization coefficients, and a combination of both of the last two types. Feedback was based upon either all previous trials (long-term feedback) or the last 10 or 20 trials (short-term feedback) or both. Two different tasks were used. Nine hundred seventy-two subjects were assigned to 32 groups in a 4 (feedback type) by 2 (short-term feedback) by 2 (long-term feedback) by 2 (task) design, and were individually run for 300 non-contingent trials.

In general, the results were that no type of feedback enhanced performance, but all feedback types except percentage correct feedback resulted in a decrement in performance. Long-term and short-term feedback interaction effects were found.

criterion variable, and found that the use of various sorts of feedback enhanced performance and achievement; in particular, feedback regarding the relation between the cues and the criterion was an aid in judgment. Hammond (1971) reported another study in which the achievement of subjects was not enhanced by knowledge of the relation between cues and criterion variables when the cue-criterion relation was essentially non-linear and the relation was described to the subject in terms of a function of a product-moment correlation coefficient. However, the subjects did perform better when more precise information about the tasks was presented to the subject in the form of graphical or verbal descriptions of the relations.

While the effect of different sorts of feedback is far from clear, it appears that the following generalization may be made in the metric case when cues and criterion are continuous: Feedback other than outcome feedback is facilitative to the judgment process if the feedback is appropriate or informative with respect to the true nature of the relation between the cues and criterion or distal variable.

The purpose of the present study was to determine the effects of different kinds of feedback upon the performance of subjects in a non-metric multiple-cue probability learning task. What will be the effect upon performance if the subject is given, in addition to the usual outcome feedback on each trial of the experiment, periodic information about either the structure of the environment and/or information about his own performance on the task?

Non-Metric Multiple-Cue Probability Learning

In the present experiment, there were two binary cue dimensions, and on each trial of the experiment the subject was presented with one value from each dimension. On the basis of experience gained in the experiment, and on the basis of the cues presented on each trial, the subject's task was to predict which of two events would occur on a given trial. In a recent study it was found that subjects do attend to the dimensionality of the cues in making judgments (Castellan, 1972) rather than attending to the cue pattern which is presented on each trial.

In such a non-metric multiple-cue probability learning situation, suppose that there are $\underline{n}$ cue dimensions and that each cue dimension is binary. Let C_i be one value of the $\underline{i}$th cue and let C_i' denote the other value. On each trial of the experiment, C_i is presented with probability $P[C_i]$; C_i' is presented with probability $P[C_i'] = 1 - P[C_i]$. Events are also binary and are labeled E_1 and E_2. On each trial, event E_1 occurs with probability $P[E_1] = \pi_1$. The cue-event contingencies may be described by the two conditional probabilities of one event given the cues. Define $P[E_1|C_1]$ to be the probability of event E_1 given presentation of cue C_1. Let $P[E_1|C_i']$ be similarly defined. Then

$$P[E_1] = \pi_1 = P[E_1|C_i]\ P[C_i] + P[E_1|C_i']\ P[C_i'],\ i = 1, 2, \ldots, n$$

are $\underline{n}$ equations relating the conditional and unconditional probabilities of an event. The ecological validity of a cue is then the product-moment correlation, or in the case of binary cues and events, the phi-coefficient (ϕ) between the event and the cue (Björkman, 1967). The squared validity of the $\underline{i}$th cue dimension is given by the following equation:

$$\phi_i^2 = \frac{\{P[E_1|C_i] - \pi_1\}^2}{\pi_1\ (1-\pi_1)} \quad \frac{P[C_i]}{1-P[C_i]} .$$

If the cue dimension and events are independent, then the validity will be zero.

Feedback

Different types of feedback may be appropriate in multiple-cue probability learning tasks. Feedback may be subdivided into three types: outcome feedback, feedback on the subject's performance, and feedback concerning the structure of the ecology in which judgments are being made.

Outcome feedback. As noted earlier, outcome feedback is probably the most widely used type of feedback in multiple-cue probability learning studies. It consists of presenting the subject the criterion or 'correct' value on every trial (e.g., Dudycha & Naylor, 1966; Naylor & Clark, 1968; Naylor & Schenck, 1968; Brehmer and Lindberg, 1971; Castellan, 1972). Classical probability learning studies have relied heavily upon outcome feedback as the informational source about the task (e.g., Castellan, 1969).

Subject performance feedback. While outcome feedback provides a subject with an indication of how accurate his prediction is on each trial, it is implicitly assumed that the subject has memory capacity sufficient to keep an appropriate record of the relevant aspects of his performance which are necessary in order to predict well on subsequent trials. One type of subject performance data is the percentage of correct responses a subject has made across trials. If a subject were given this information in a periodic manner, this form of feedback might affect performance since the subject would not be required to rely only on memory. While the percentage of correct responses is probably not a useful sort of feedback in interval learning tasks where the criterion can take on a large number of values and perfect accuracy is virtually impossible, it would seem to be quite appropriate in non-metric multiple-cue probability learning tasks when the criterion values are relatively few and disti

Another sort of subject performance feedback is the utilization coefficient, the correlation coefficient between the cues and responses (or phi-coefficient in binary cue and binary response environments). The utilization coefficient is an indication of how the subject is weighting the cues in making his judgments. If a cue utilization coefficient is unity, then the subject is relying completely upon that cue dimension in making his judgments across trials; but if the cue utilization coefficient is zero, the cue is, in effect, being ignored by the subject in making his judgments. (However, if the cues are intercorrelated, the cue utilization coefficient value for one cue dimension may reflect dependence on another correlated cue dimension.)

Feedback concerning the task environment. Another sort of feedback is that concerning the detailed structure of the task environment or ecology. For example, if the feedback to the subject includes the cue-event validity coefficients, then the subject has more precise information about the relative usefulness of the cues in making judgments. A few studies have been reported in which such feedback has been presented (e.g., Todd & Hammond, 1965; Summers & Hammond, 1966).

In the case where only cue utilization coefficients are presented, the subject has no way of knowing whether or not he is weighting the cues in an appropriate manner. If the subject is given, in addition to cue utilization coefficients, the ecological validities of the cues, then there exists a basis for comparison which could serve to moderate behavior.

Method

Subjects

The subjects were 972 students from Indiana University who were fulfilling a course requirement in introductory psychology. Subjects were assigned to conditions in sets of four. There were 32 groups and the sample sizes in each group ranged from 21 to 61; the average sample size was 30.

Apparatus

Subjects were run in sets of four. The apparatus for the experiment was one IBM 2260 display station for each subject. The display station consisted of a 10 cm. by 23 cm. cathode ray tube (CRT) for the purpose of displaying cues, feedback, and data, and an alphameric keyboard for entering responses. Each subject's display station was located in a booth designed so that the other subjects and their displays were not visible. All experimental control and monitoring was done by an IBM 1800 Data Acquisition System.

Design

The design of the experiment was a 2 (short-term feedback, designated S) by 2 (long-term feedback, designated L) by 4 (feedback type) by 2 (task) factorial. In all conditions there were two binary cue dimensions and two events. The two tasks or replications were chosen to represent two different probabilistic task environments or ecologies. The tasks will be denoted as Task A and Task B. The two levels of short-term feedback were feedback based on the previous 20 trials or the absence of such feedback. The two levels of long-term feedback consisted of the presence or absence of feedback based upon all previous trials. The text of the various feedback messages is presented in Table 1. Regardless of the presence or absence of short-term or long-term feedback, subjects received outcome feedback on every trial. The four types of feedback messages were the following: Cue-event validity coefficients (CE

Insert Table 1 about here

----

feedback); cue-response utilization coefficients (CR feedback); both cue-event validities and cue-response utilization coefficients (CE + CR feedback); and percentage of correct responses (PCT feedback).

The cue dimensions consisted of shape--square or triangle--and direction--whether the figure was built from vertical or horizontal bars. The events consisted of the symbols " < " and " >," which were named "left arrow" and "right arrow," respectively, in the verbal instructions. Three hundred trials were run in the experiment. The sequences of events and cues were generated so that $P[E_1] = .5$. The validities of the cue dimensions were approximately $\phi_1 = .0$ and $\phi_2 = .6$ in Task A, and $\phi_1 = .4$ and $\phi_2 = .6$ in Task B. Actual validities and conditional probabilities of events—given the cues are presented in Table 2. Sequences of cues and events were generated

----

Insert Table 2 about here

----

by procedures developed by Castellan (1970).

Procedure

After subjects were read instructions on the use of the display stations, they made a few practice responses to gain acquaintance with the use of the keyboard. The experimental instructions were then read to the subjects and they were informed that the cues might be helpful to them in making predictions. The cue dimensions were specifically labeled and each cue value was named. The subjects were told that it was 'virtually impossible' to be correct on every trial, but that they could do better or worse depending upon how much they learned. The subjects were told that their predictions would affect in no way the events which would occur on each trial.

Next, depending upon the experimental group, the subjects were told that additional information would be presented periodically. The subjects in the percentage correct group were told that they would be given the percentage of correct responses which they had made either over all trials, or over the last 20 trials, or both, depending upon conditions. Subjects receiving feedback on the cue-event validities were told that they would be given information concerning how the cues should be weighted in making judgments. Weights were described in a rather general way, and the subjects were told that a weight of zero indicated that the cue was useless in making a judgment and that a weight of one meant that the use of that cue would lead to correct predictions on every trial, and that intermediate values meant intermediate degrees of usefulness. Subjects receiving the cue-response utilization coefficients were given similar instructions but the reference was to the responses rather than the events. Finally, subjects receiving both cue-event validity coefficients and cue-response utilization coefficients were given similar instructions. As in the percentage correct condition, the instructions were modified as to whether the subject was to receive feedback concerning the last 20 trials, all previous trials, or both.

The subject's response consisted of typing either a " < " or " > ," depending upon whether they were predicting the event " < " or " > ," respectively.

At the beginning of a trial the cues were presented on the face of the CRT. There was then a pause of 3 seconds in order for the subject to type a response; when the subject typed in a response, it appeared immediately on the face of the CRT. After the delay for the response, the event was presented for 2 seconds, during which time the cues and response were also visible. At the end of the trial, the CRT was erased and the next trial began. The intertrial interval was about .2 seconds. If a subject failed to respond, the failure was recorded and his response was not analyzed on that trial.

The various feedback messages were presented every 20 trials for subjects in Task A and every 10 trials for subjects in Task B. Each of the feedback messages was presented for 5 seconds.

Results

The data were analyzed by determining two sets of data for each subject. The first set was a measure of subjects' accuracy and consisted of the proportion of correct responses made in each block of 100 trials. Since different response patterns can result in the same proportion of correct responses, the second set of data consisted of the proportions of A_1 responses conditional upon the cues.

Subject Accuracy

For each subject the proportion of correct responses--each correct response defined as an E_i - A_i pair--was calculated for each block of 100 trials. These data were then averaged for each condition and are presented in Table 3. Each subject's response proportions were transformed by means

----

Insert Table 3 about here

----

of an arc-sine square-root transformation and a 4 (feedback type) by 2 (long-term feedback) by 2 (short-term feedback) analysis of variance was done for each block of 100 trials. Each Task was analyzed separately.

There were no systematic differences in the proportion of correct responses due to feedback type. Nor were there any short-term or long-term feedback main effects. However, there was a consistent short-term by long-term interaction in both tasks ($p = .054, .024, .122$ for each block of 100 trials in Task A, and $p = .010, .064, .016$ for each block of 100 trials in Task B). This interaction was due to subjects who received either <u>both</u> types of temporal feedback or <u>no</u> temporal feedback to produce a higher proportion of correct responses than subjects who received only one kind of temporal feedback.

<u>Response Proportions Conditional Upon Cues</u>

Since there were two binary cue dimensions, there were four cue patterns observed by the subject: C_1 and C_2, C_1 and C_2', C_1' and C_2, and C_1' and C_2'. For each subject the proportion of A_1 responses was calculated conditional upon each cue pattern for each block of 100 trials. All subsequent analyses are based upon these proportions.

The average A_1 response proportions conditional upon the cue patterns when only outcome feedback was given are presented in Table 4. These response

----

Insert Table 4 about here

----

proportions are also depicted in Figure 1 for Task A and Task B. Examination

----

Insert Figure 1 about here

----

of the figure reveals the differences in the two ecologies upon performance even though the optimal response strategies are the same for both Tasks. The average A_1 responses conditional upon cues were then calculated for the various feedback types, and short-term and long-term feedback conditions; these data are summarized in Tables 5 and 6 for Task A and Task B, respectively.

Insert Table 5 and Table 6 about here

The response proportions for one feedback type condition are summarized in Figure 2. Because it was expected that the results for Tasks A and Tasks B would

---

Insert Figure 2 about here

---

be rather different due to difference in the two ecologies, the results for each shall be presented separately. Prior to all further analyses, the response proportions were subjected to an arc-sine square-root transformation.

The data were first grouped into a 4 (feedback type) by 2 (short-term feedback) by 2 (long-term feedback) non-orthogonal multivariate analysis of variance design with four dependent variables, the dependent variables being the response proportions conditional upon the cues. Each block of 100 trials was analyzed separately by means of a multivariate analysis of variance. In both Tasks A and Tasks B there was a significant feedback type condition effect in the first two blocks of 100 trials ($p = .001$ and .001 in the first two blocks of Task A, and $p = .098$ and .007 for the first two blocks in Task B), while there were no differences in the third block of trials. The exact nature of this effect will be described in detail later. Aside from a feedback type effect, in Task A there was a slight short-term by long-term feedback interaction in the second block of 100 trials ($p = .091$), and a slight feedback type by long-term feedback interaction in the last block of trials ($p = .082$). However, in Task B there was a significant short-term by long-term feedback interaction effect in the first two blocks of trials ($p = .002$ and $.020$, respectively) as well as a significant feedback type by long-term feedback interaction ($p = .001$) and a long-term feedback main effect ($p = .010$) in the first block of 100 trials. All other effects and interactions were not significant (p's $> .12$).

The Effect of Feedback Type

In an effort to specify and understand more precisely the feedback type effect, the feedback type variable, C, was partitioned into the following contrasts: C1, a contrast between cue-event validity coefficient feedback and cue-response utilization coefficient feedback (CE vs. CR); C2, a contrast between cue-response utilization coefficient feedback given singly or in combination (CE, CR vs. CE + CR); C3, a contrast between the percentage correct feedback and the other feedback type conditions (CE, CR, CE + CR vs. PCT).

A multivariate analysis of variance of the partitioned sums of squares of the feedback type variable was performed for Task A and Task B. For Task A the significant feedback type effect noted earlier appears to be due only to the percentage correct (PCT) feedback since the main effects due to the difference between CE feedback and CR feedback present singly (contrast C1) and in combination (contrast C2) was not significant in any block of trials. However, when percentage correct feedback was compared with the other types of feedback (contrast C3) this difference was significant in all blocks of 100 trials ($p = .001$, $.001$, $.044$ in each block of trials, respectively).

This effect is due to the general tendency for the four response proportions to be further from maximizing for each of the four variables in the percentage correct feedback condition compared to the other forms of feedback. However, there was also a slight short-term by long-term by percentage correct feedback type interaction in the first block of trials ($p = .061$) due to subjects in the short-term feedback condition producing response proportions which were far less extreme than those of subjects in the other feedback type groups. This difference accounts for a large part of the main effect. However, this interaction diminished across blocks and was not significant in the second and third blocks of trials. In the third block of trials there was a significant long-term feedback by percentage correct feedback interaction ($p = .012$) due to the subjects who received long-term feedback approaching a more extreme response asymptote than subjects in other conditions.

For Task B the effect of feedback type also appears to be due only to the percentage correct feedback since the contrast effect due to the difference between CE feedback and CR feedback (contrast C1) was significant ($p = .047$) only in the second block of 100 trials, and the difference between CE feedback and CR feedback given singly and in combination (contrast C2) was not significant in any block of trials. The comparison of percentage correct feedback with all other types of feedback (contrast C3) was significant in the first two blocks of trials ($p = .098$, $.015$). However, the effect was smaller than the comparable effect for Task A. Unlike the results in Task A, the four response proportions tended to be closer to maximizing in the percentage correct feedback condition than the response proportions in the other three feedback type conditions. Note, however, the absence of an interaction with short-term and long-term feedback. In the CE feedback and CR feedback, given singly versus given in combination effect (contrast C2), there was an interaction with long-term feedback in the first block of trials ($p = .001$) and an interaction between C2 and short-term feedback in the second block of trials ($p = .057$). These effects appear to be due to subjects doing significantly better with long-term feedback when both forms of correlational feedback are given than when only one of the types of temporal feedback is given in the first block and a lack of increase in response proportions in one condition for the second block. This effect diminishes and disappears by the third block of trials.

Short-Term and Long-Term Feedback with Feedback Type

Each of the four feedback type conditions was then analyzed separately to determine the effect of short-term feedback and long-term feedback on performance. Again each of the four response proportions was used as a dependent variable in a 2 (short-term feedback) by 2 (long-term feedback) non-orthogonal multivariate factorial design. Each block of 100 trials was analyzed separately.

Results for Task A. For the percentage correct feedback condition a long-term feedback effect emerged. Subjects who did not receive long-term feedback produced response proportions which were closer to chance responding than the response proportions of subjects who received long-term feedback. This effect was present in the first block of trials but was not significant until the second and third blocks of trials ($p = .133, .047, .006$, respectively). Subjects who received both short-term and long-term feedback had response proportions slightly more extreme than in other conditions but the effect (on interaction) was not significant.

For subjects who received CE feedback there was a short-term feedback effect that began in the first block of trials but which diminished across blocks ($p = .003, .076, .511$, respectively). In this case, subjects who received short-term feedback tended to produce response proportions closer to maximizing than subjects who did not receive short-term feedback.

For subjects who received CR feedback, there was an interaction between short-term and long-term feedback in the first two blocks of trials which disappeared by the third block of trials ($p = .048, .036, .863$, respectively). The nature of this interaction was that subjects who received only <u>one</u> type of temporal feedback did slightly better than those who received either both short-term and long-term feedback or no feedback (other than outcome feedback).

For those subjects who received CE and CR feedback, no significant differential effect due to the presentation of short-term or long-term feedback emerged, although subjects who received <u>both</u> short-term <u>and</u> long-term feedback produced response proportions closer to maximizing.

<u>Results for Task B</u>. For the subjects who received percentage correct feedback there was a significant interaction between the short-term and long-term feedback conditions in the first two blocks of trials ($p = .065, .029, .671$, respectively). There was also a significant long-term feedback effect ($p = .010$) in the first block of trials. The nature of this effect was that subjects who received both short-term and long-term feedback initially had response proportions which were closer to maximizing than when a single form of temporal feedback was given alone. The subjects who received short-term feedback only performed worse than subjects who received only long-term feedback.

For subjects in the CE feedback condition, the four response proportions were essentially the same as in the no-feedback, short-term feedback, and short-term and long-term feedback conditions. The presence of long-term feedback by itself results in less extreme response proportions, but this was significant only in the first block of trials in the form of a short-term by long-term feedback interaction ($p = .048$) and a long-term feedback main effect ($p = .064$).

For subjects who received CR feedback there were no differences as a function of short-term and long-term feedback, although there was a slight tendency for long-term feedback to result in a lower asymptote for two of the response variables.

When subjects received both CE and CR feedback, there were significant effects in every block of trials. The nature of these effects was a <u>decrement</u> in response proportions when short-term feedback was given either by itself or in combination with long-term feedback. This effect appeared to interact with the dependent variables of response type but the effect was not tested.

Discussion

The effect of different types of feedback in multiple-cue probability learning with binary cues and criterion produces little in the way of systematic effects across ecologies and temporal feedback conditions. This is not surprising since it has been found that in interval learning tasks performance is quite sensitive to the ecology (see, e.g., Dudycha & Naylor, 1966).

The Effect of Feedback Type

When the different sorts of feedback are considered it appears that, of the three types of feedback given, percentage correct feedback produces subject response proportions most like those which result when outcome feedback only is given. Performance in general is worse when the various sorts of weights are presented to the subjects. This is particularly true in the early trials of the experiment, and the differences in performance between the various types of feedback decrease across trials.

While the effects appear to depend at least in part upon the ecology itself, the following rather general statements can be made concerning feedback types. In the discussion which follows, each of the feedback types will be discussed with respect to a 'standard' comprising outcome feedback only.

Percentage correct (PCT) feedback. While short-term feedback alone depresses response proportions in the percentage correct feedback condition regardless of the ecology, long-term feedback by itself maintains responding at 'normal' levels while both forms of temporal feedback together tend to elevate response proportions slightly. These effects are sustained in the later trials in Task A but diminish across trials in Task B. Thus the feedback enhances terminal responding significantly only when one cue dimension is irrelevant (as in Task A).

While it might be argued that the effect of periodic presentation of the percentage correct responses might be to encourage the subject to maximize, there was no evidence that such was the case.

Cue-event validity coefficient (CE) feedback. The periodic presentation of cue-event validity coefficients affected acquisition but not terminal responding. The effects were different in the two ecologies, probably due in part to the result that the feedback in Task A regularly informed the subject that he ought to ignore one of the cue dimensions, while in Task B the subject was being periodically reminded that both dimensions were useful.

In Task A, where one cue dimension was irrelevant, short-term feedback enhanced early acquisition (and both short-term feedback and long-term feedback together result in even better acquisition; but the effect--in the form of an interaction--was not significant). In Task B, presentation of either short-term or long-term feedback alone results in an initial decrement in responding; both short-term and long-term feedback together allowed the subject to perform about the same as when only outcome feedback was given. The conclusion to be drawn from data from the two ecologies would seem to be the following: If a cue dimension is irrelevant, remind the subject of the fact but if all of the cue dimensions are relevant, then outcome feedback alone is satisfactory. The latter seems true only for acquisition and not for asymptotic performance since the differences were not sustained across trials.

Cue-response utilization coefficient (CR) feedback. Informing the subject of the manner in which he is in fact weighting the cue dimensions might appear to have little effect upon the subject because the subject has no way of knowing just what the weights should be in the task.

Examination of the data indicates that the subjects ignored CR feedback in Task B, but that there was an effect in Task A early in the experiment. The nature of the effect is not clear but may have been due to subjects 'experimenting' and trying to use both cue dimensions even though one dimension was irrelevant.

Both cue-event validity coefficient and cue-response utilization coefficient (CE + CR) feedback. When both types of weights are presented to the subject, he knows not only what the weights of the cue dimensions should be, but also the manner in which he has been weighting the cue dimensions. That is, the subject has a basis for comparison of the weights produced by his responses. In Task A there were no significant effects although either short-term or long-term feedback by themselves resulted in poorer performance. On the other hand, in Task B there were effects throughout the experimental session. In this case, short-term feedback by itself or in combination with long-term feedback resulted in poorer performance than that when either long-term feedback or just outcome feedback was given.

Short-Term and Long-Term Feedback

There are numerous problems with attempting to compare short-term feedback and long-term feedback. One problem is the differential variability of the two types of feedback--variability which may be observed by the subject. Long-term feedback given at trial _n_ is based upon _n_ trials while short-term feedback at trial _n_ is based on _m_ trials. While _m_ is constant (at least after trial _m_) _n_ is always increasing. Since both forms of feedback are sample values, the short-term feedback statistics will be approximately $\sqrt{n/m}$ times as variable as long-term feedback statistics. For example, if short-term feedback is based on the last 10 trials (_m_ = 10), then at trials 100 and 200 the short-term feedback statistics will be approximately 3.2 and 4.5 times as variable, respectively, as the long-term feedback statistics on those trials. Further, the variability of the short-term feedback statistics remains constant across trials while the variability of the long-term feedback statistics decreases across trials.

If any conclusion concerning the differential effects of short-term feedback and long-term feedback may be drawn, it is that short-term feedback, when it has an effect, acts as a suppressor of responding on the part of subjects. If long-term feedback is given, it may act to diminish the negative effect of short-term feedback. However, the effect of long-term feedback by itself interacts with both feedback type and ecology, and in no case is the effect of long-term feedback much more effective than outcome feedback alone. Both short-term and long-term feedback together tend to enhance performance regardless of feedback type, and while effect is not completely consistent in its effect upon response proportions, it produces a consistent effect when the proportion of correct responses is analyzed.

It may be that the larger variability of the short-term feedback statistics is the major factor contributing to the poor performance when this sort of feedback is used. However, the same reasoning would lead one to conclude that the negligible effect of long-term feedback is due to the fact that it becomes more stable across trials and consequently provides very little new information as the experiment progresses.

Conclusion

The lack of major, systematic, and significant effects in this study may be due to several factors. Of course, it is possible that the different sorts of feedback in the non-metric case simply do not make any real difference in multiple-cue probability learning. Another possibility may be that the subject is overwhelmed with data and is unable to cope with it adequately or appropriately. While this may be a possibility, it is not suggested by the data of this experiment since as more and more feedback was given to the subject there was not necessarily a progressive deterioration in performance. The results may also be due to fairly large individual differences. Björkman (1969) has shown that there are individual differences in apparent strategies in the multiple-cue probability learning situation. A persuasive argument concerning the role of such individual differences in learning with probabilistic schedules has been given by Fiorina (1971). Finally, it should also be recalled that, in the research with metric ecologies, no consistent results have emerged. In the case of binary cues and criteria, further research would seem to be warranted to discover the actual conditions and types of feedback with which performance could be significantly altered or enhanced.

REFERENCES

Björkman, M. Stimulus-event learning and event learning as concurrent processes. Organizational Behavior and Human Performance, 1967, 2, 219-236.

Björkman, M. Individual performances in a single-cue probability learning task. Scandinavian Journal of Psychology, 1969, 10, 113-123.

Brehmer, B. and Lindberg, L. The relation between cue dependency and cue validity in a single-cue probability learning task with scaled cue and criterion variables. Organizational Behavior and Human Performance, 1970, 5, 542-554.

Castellan, N. J., Jr. Effect of change of payoff in probability learning. Journal of Experimental Psychology, 1969, 79, 178-182.

Castellan, N. J., Jr. Determination of joint distributions from marginal distributions in dichotomous systems. Psychometrika, 1970, 35, 439-454.

Castellan, N. J., Jr. Multiple-cue probability learning with irrelevant cues. Organizational Behavior and Human Performance, 1972, in press.

Dudycha, L. W. and Naylor, J. C. Characteristics of the human inference process in complex choice behavior situations. Organizational Behavior and Human Performance, 1966, 1, 110-128.

Fiorina, M. P. A note on probability matching and rational choice. Behavioral Science, 1971, 16, 158-166.

Hammond, K. R. Computer graphics as an aid to learning. Science, 1971, 172, 903-908.

Naylor, J. C. and Schenck, E. A. The influence of cue redundancy upon the human inference process for tasks of varying degrees of predictability. Organizational Behavior and Human Performance, 1968, 3, 47-61.

Naylor, J. C. and Clark, R. D. Intuitive inference strategies in interval learning tasks as a function of validity magnitude and sign. Organizational Behavior and Human Performance, 1968, 3, 378-399.

Summers, D. A. and Hammond, K. R. Inference behavior in multiple-cue tasks involving both linear and non-linear relations. Journal of Experimental Psychology, 1966, 71, 751-757.

Todd, F. J. and Hammond, K. R. Differential feedback in two multiple-cue probability learning tasks. Behavioral Science, 1965, 10, 429-435.

Footnotes

[1]This research was supported by National Institutes of Health Grants No. MH-16817 and No. PHS S05 FR7031-04. Portions of the research reported here were presented at the Fourth Annual Conference on Human Judgment held at the University of Colorado, March 5-7, 1971.

[2]Final preparation of this paper was done during the author's tenure at the Oregon Research Institute, Eugene, Oregon. The author would like to thank Robyn Dawes, Lewis Goldberg, Sarah Lichtenstein, and Paul Slovic for many helpful comments concerning this research.

Table 1

Different Types of Feedback Messages

Used in the Present Study

Message Number[1]	Display
1 S	OVER THE LAST MM TRIALS THE CUE WEIGHTS HAVE BEEN SHAPE .XXX DIRECTION .YYY
1 L	OVER ALL TRIALS THE CUE WEIGHTS HAVE BEEN SHAPE .XXX DIRECTION .YYY
2 S	OVER THE LAST MM TRIALS YOU HAVE BEEN WEIGHTING THE CUES AS FOLLOWS: SHAPE .XXX DIRECTION .YYY
2 L	OVER ALL TRIALS YOU HAVE BEEN WEIGHTING THE CUES AS FOLLOWS: SHAPE .XXX DIRECTION .YYY
3 S	OVER THE LAST MM TRIALS YOU HAVE MADE XX% CORRECT RESPONSES
3 L	OVER ALL TRIALS YOU HAVE MADE XX% CORRECT RESPONSES

(Feedback is given every NN trials.)

[1]S and L denote short-term and long-term messages, respectively.

Table 2

Table of Ecological Validities and Event

Given Cue Conditional Probabilities

	Task A	Task B
$P[E_1]$	.50	.50
$P[C_1]$	.48	.50
$P[E_1 \mid C_1]$	.500	.680
$P[E_1 \mid C_1']$	.500	.320
Validity cue 1	.00	.36
$P[C_2]$	.48	.46
$P[E_1 \mid C_2]$	.833	.826
$P[E_1 \mid C_2']$	.192	.222
Validity cue 2	.64	.60

TABLE 3

Proportion of Correct Responses for Groups by Blocks of 100 Trials (Decimal Point Omitted)

		Task A			Task B		
Blocks of 100 Trials		1	2	3	1	2	3
Feedback Type							
CE	SL	641	730	747	629	704	701
	S	606	722	751	618	696	695
	L	591	702	734	604	673	674
CR	SL	609	720	728	632	707	710
	S	623	735	756	642	709	703
	L	609	696	731	635	698	699
CE + CR	SL	606	728	750	661	702	696
	S	592	675	713	632	680	656
	L	610	699	704	634	694	705
PCT	SL	644	726	769	642	708	722
	S	579	661	713	605	682	697
	L	632	715	751	637	687	698
No Feedback		609	711	736	642	702	711

TABLE 4

Proportion of A_1 Responses Conditional Upon Cues for Subjects Receiving Outcome Feedback Only For each Block of 100 Trials (Decimal Point Omitted)

	Task A			Task B		
	Blocks of 100 Trials			Blocks of 100 Trials		
	1	2	3	1	2	3
$P[A_1 \mid C_1' \& C_2']$	295	163	133	256	173	117
$P[A_1 \mid C_1 \& C_2']$	326	172	128	430	256	273
$P[A_1 \mid C_1' \& C_2]$	675	844	886	645	806	751
$P[A_1 \mid C_1 \& C_2]$	716	887	892	848	885	944

TABLE 5

Proportion of A_1 Responses Conditional upon
Cue Patterns for each Feedback Condition
By Blocks of 100 Trials for Task A
(Decimal Point Omitted)

Feedback Type	CE			CR			CE + CR			PCT			Cue Pattern
Block of 100 Trials	1	2	3	1	2	3	1	2	3	1	2	3	
Feedback Type													
SL	241	130	083	272	141	141	278	159	144	283	167	090	C'_1 & C'_2
	307	173	102	348	198	141	319	132	097	231	146	086	C_1 & C'_2
	736	863	890	698	865	868	674	882	911	653	883	940	C'_1 & C_2
	795	955	921	732	901	858	733	934	936	806	892	944	C_1 & C_2
S	298	161	123	248	152	102	315	228	170	336	245	160	C'_1 & C'_2
	346	169	123	287	148	100	386	249	179	343	252	177	C_1 & C'_2
	669	875	904	699	893	915	656	846	840	590	759	831	C'_1 & C_2
	754	922	924	798	947	937	736	896	869	675	784	837	C_1 & C_2
L	314	215	178	317	206	132	310	214	190	309	171	122	C'_1 & C'_2
	372	191	113	352	222	145	336	226	178	270	148	107	C_1 & C'_2
	680	689	895	675	831	892	707	859	852	652	856	900	C'_1 & C_2
	698	889	868	745	932	908	716	900	873	784	872	907	C_1 & C_2

TABLE 6

Proportion of A_1 Responses Conditional upon
Cue Patterns for each Feedback Condition
By Blocks of 100 Trials for Task B.
(Decimal Point Omitted)

Feedback Type	CE			CR			CE + CR			PCT			Cue Pattern
Block of 100 Trials	1	2	3	1	2	3	1	2	3	1	2	3	
Feedback Type													
SL	258	161	141	228	141	113	196	139	100	198	133	110	C_1' & C_2'
	450	267	316	456	303	317	486	300	361	439	243	293	C_1 & C_2'
	639	777	749	614	773	809	666	694	680	597	826	802	C_1' & C_2
	832	904	942	821	907	950	891	913	932	803	892	941	C_1 & C_2
S	278	166	141	275	167	123	277	182	184	328	184	137	C_1' & C_2'
	482	301	320	470	312	277	451	266	349	486	319	319	C_1 & C_2'
	679	771	748	666	715	761	648	652	644	699	778	743	C_1' & C_2
	825	876	921	864	899	952	768	840	857	847	867	933	C_1 & C_2
L	322	239	186	248	171	127	257	158	119	254	175	120	C_1' & C_2'
	503	325	361	482	362	359	438	266	316	447	296	256	C_1 & C_2'
	605	729	711	586	720	682	638	773	782	688	780	786	C_1' & C_2
	790	838	901	838	908	940	813	865	926	804	880	907	C_1 & C_2

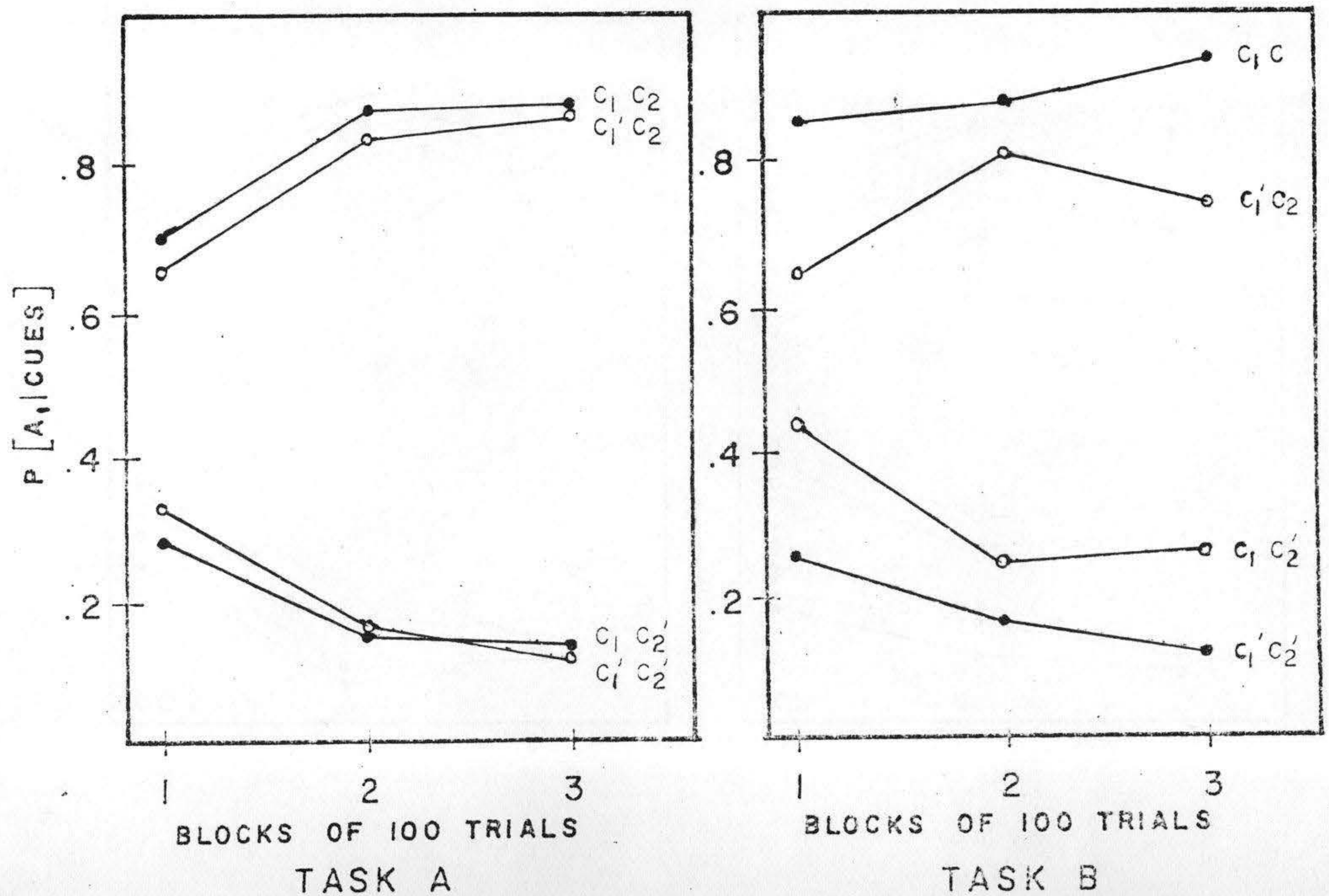

Figure 1: Proportions of A_1 Responses Conditional upon Cues for Subjects Who Received Who Received Outcome Feedback Only.

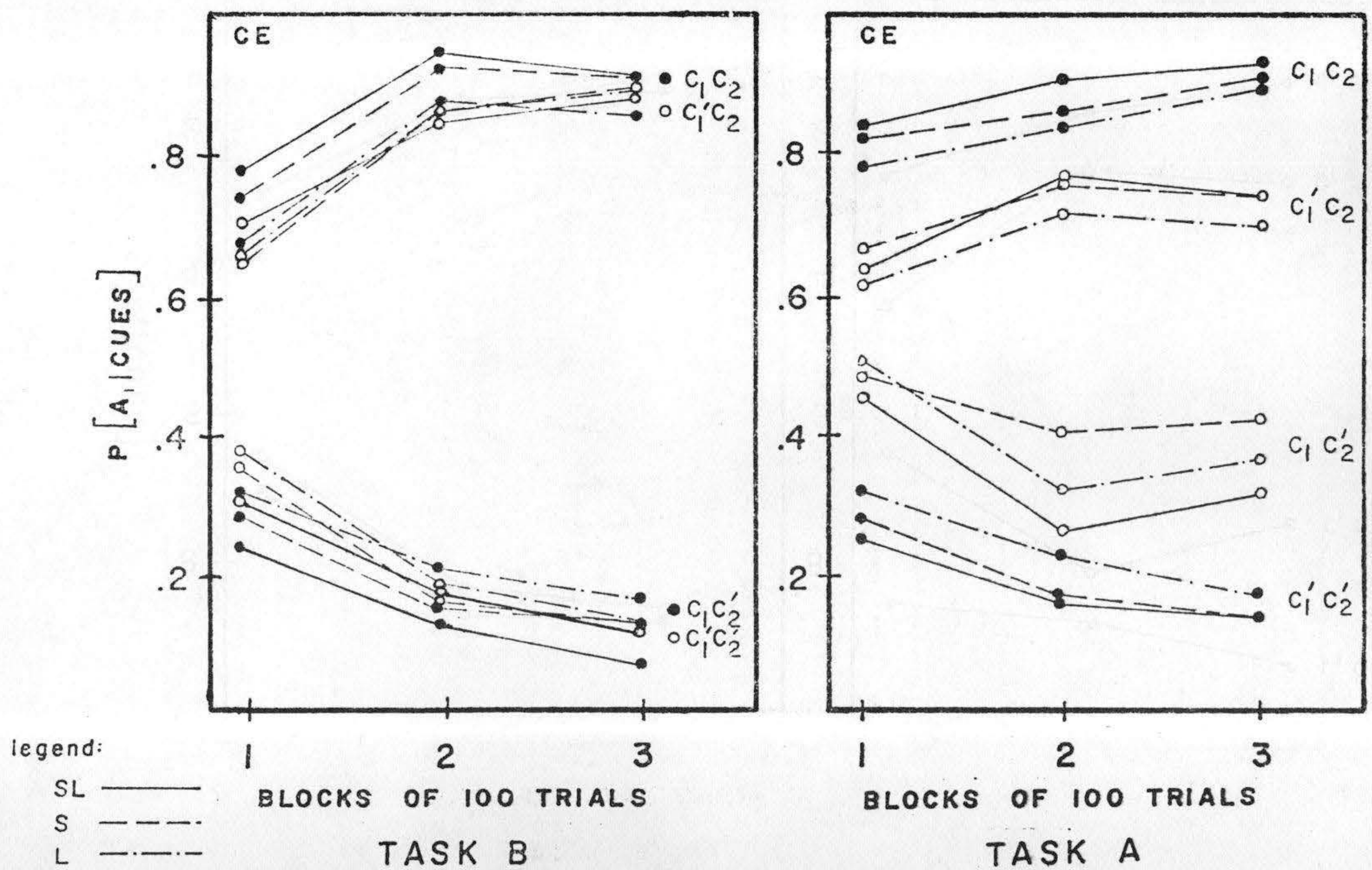

Figure 2: Proportions of A_1 Responses Conditional upon Cues for Subjects

Who Received Cue-Event Validity Coefficient Feedback in the Form

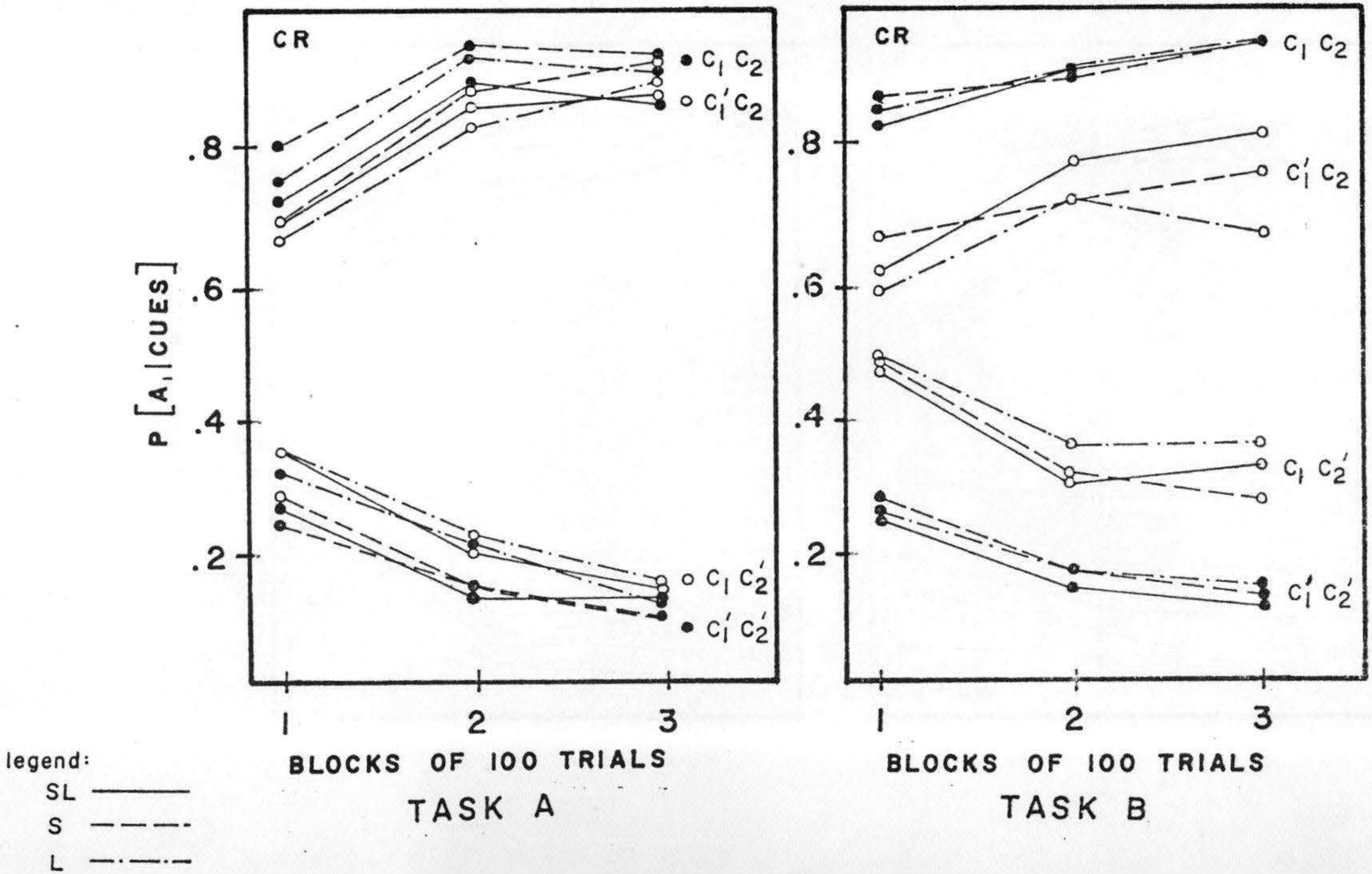

Figure 2b: Cue-Response Utilization Coefficient Feedback

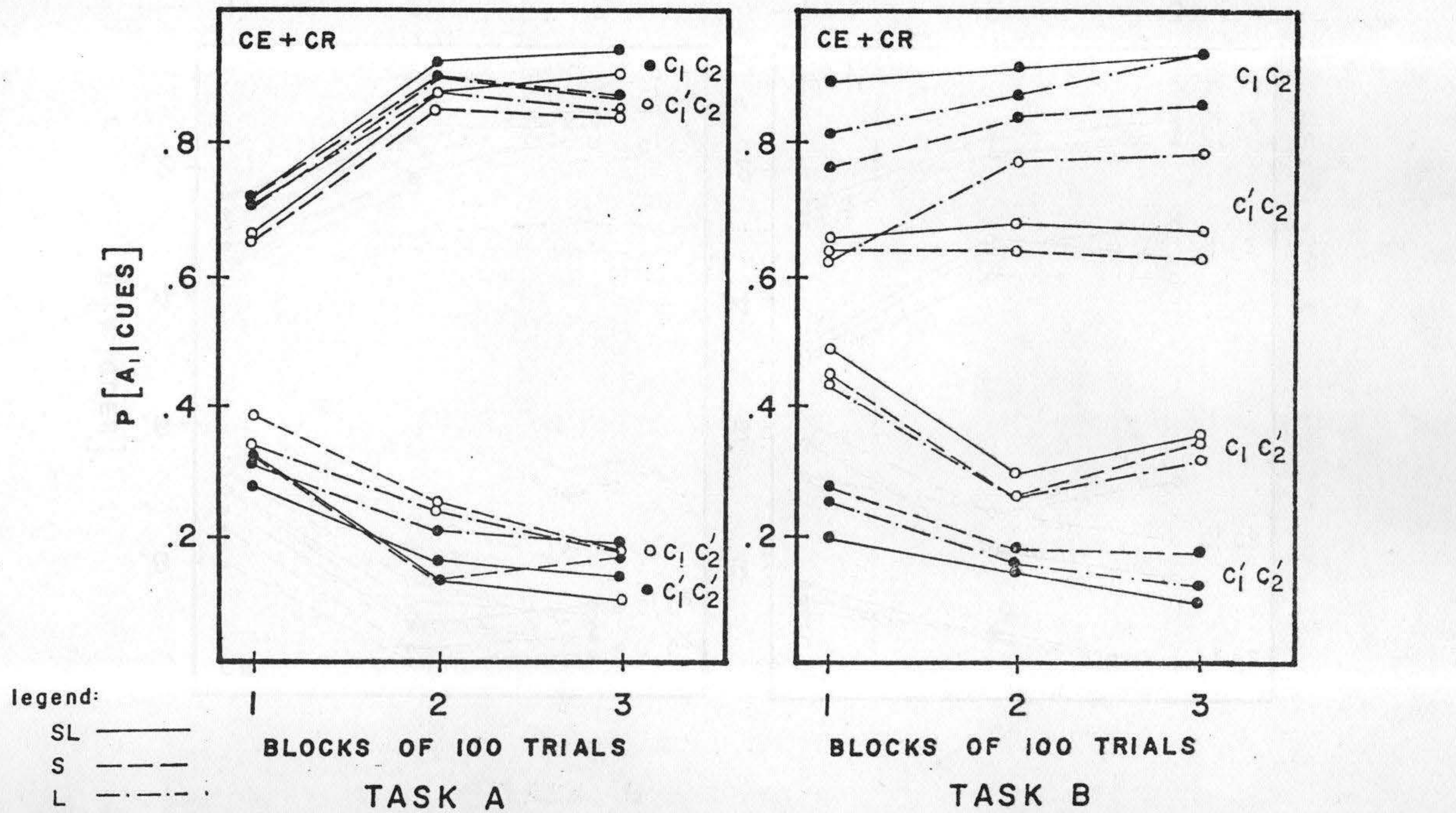

Figure 2c: Both Cue-Event Validity Coefficient and Cue-Response Utilization Coefficient Feedback

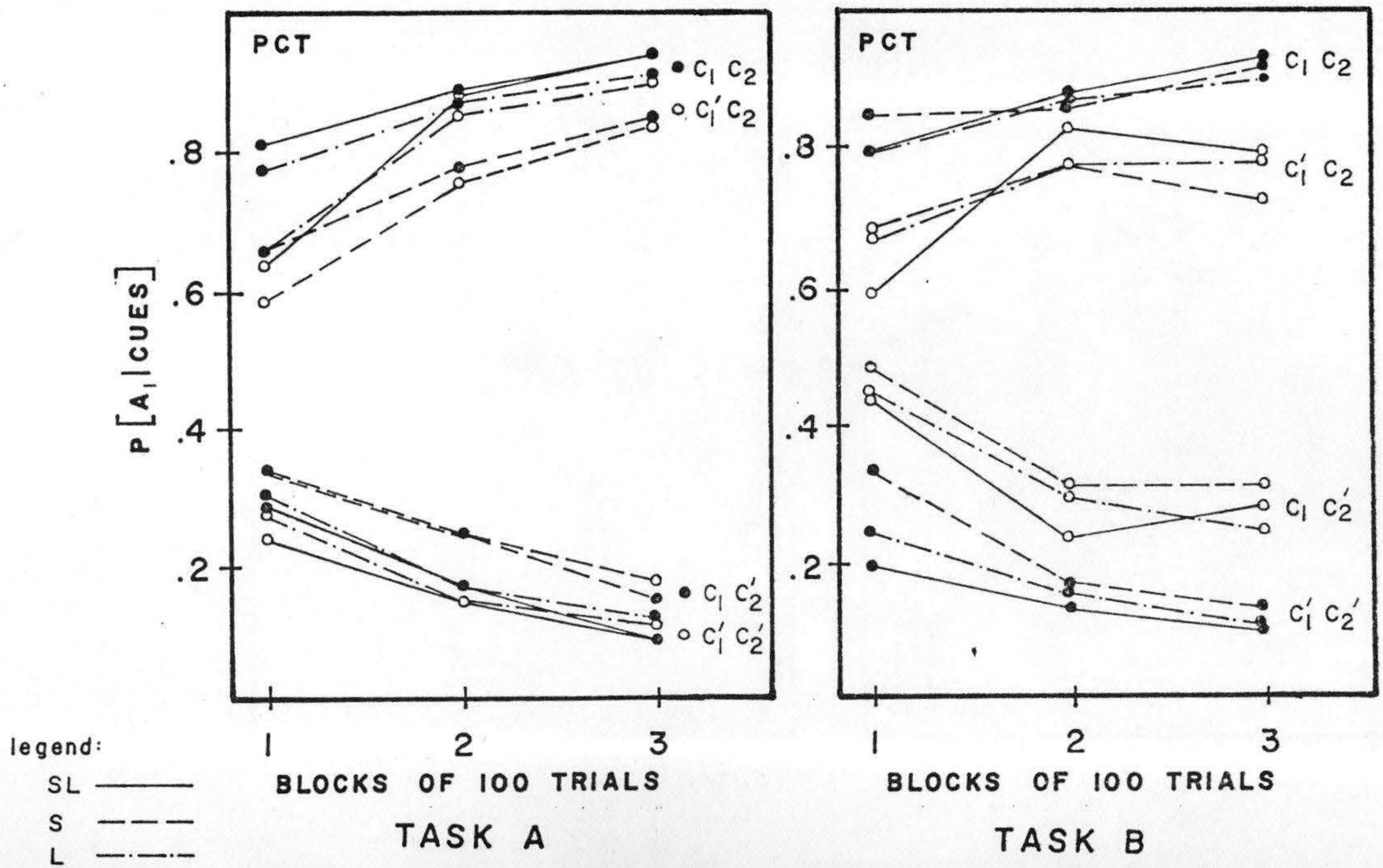

Figure 2d: Percentage Correct Feedback

Appendix A

The appendix contains summaries of the results of the multivariate analyses of variance which were performed. See text for discussion.

Table A-1

Results of Multivariate Analysis of Variance on Feedback Type (C), Short-Term Feedback (S), and Long-Term Feedback (L) by Blocks of 100 Trials (Tabled Values Are Significance Probabilities Derived from the Distribution of Wilks' Lambda Criterion)

	Effect	Blocks of 100 Trials 1	2	3	Degrees of Freedom HYP/ERROR
Task A	CSL	.223	.139	.610	12/1336
	SL	.199	.091	.344	4/505
	CL	.494	.336	.082	12/1336
	CS	.253	.358	.818	12/1336
	L	.228	.120	.457	4/505
	S	.273	.375	.742	4/505
	C	.001	.001	.233	12/1336
Task B	CSL	.381	.334	.804	12/1135
	SL	.002	.020	.119	4/429
	CL	.001	.844	.206	12/1135
	CS	.213	.247	.653	12/1135
	L	.010	.127	.168	4/429
	S	.345	.144	.850	4/429
	C	.098	.007	.335	12/1135

Table A-3

Results of Multivariate Analysis of Variance on Special Contrasts of Feedback Type (C), Short-Term Feedback (S), and Long-Term Feedback (L) by Blocks of 100 for Task B (Tabled Values of Significance Probabilities Derived from the Distribution of Wilks' Lambda Criterion with 4 Degrees of Freedom for Hypotheses and 429 Degrees of Freedom for Error)

Effect[1]	Blocks of Trials 1	2	3
C1 S L	.204	.540	.531
C1 S	.814	.911	.662
C1 L	.385	.420	.215
C1	.122	.047	.537
C2 S L	.488	.127	.326
C2 S	.114	.057	.219
C2 L	.001	.727	.186
C2	.441	.180	.255
C3 S L	.447	.464	.996
C3 S	.154	.265	.794
C3 L	.148	.895	.480
C3	.098	.015	.268

[1]Contrasts used in the above are:

C1 Cue-event validity coefficient feedback vs. cue-response utilization coefficient feedback.

C2 Cue-event validity coefficient plus cue-response utilization coefficient feedback vs. average of cue-event validity coefficient feedback and cue-response utilization coefficient feedback.

C3 Percentage correct feedback vs. average of other types of feedback.

Table A-4

Results of Multivariate Analysis of Variance on Each Feedback Group Separately for Subjects in Task A

(Tabled Values Are Significance Probabilities Derived from the Distribution of Wilks' Lambda Criterion)

Feedback Type	Contrast[1]	Blocks of Trials 1	2	3	Degrees of Freedom HYP/ERROR
Percentage Correct	SL	.151	.513	.319	4/184
	L	.133	.047	.006	4/184
	S	.901	.564	.675	4/184
Cue-Event Validities	SL	.437	.744	.408	4/108
	L	.406	.826	.706	4/108
	S	.003	.076	.511	4/108
Cue-Response Utilization Coefficients	SL	.048	.036	.863	4/107
	L	.769	.236	.678	4/107
	S	.432	.485	.989	4/107
Both Cue-Event and Cue-Response Coefficients	SL	.652	.306	.354	4/97
	L	.623	.329	.694	4/97
	S	.470	.653	.857	4/97

[1]S denotes short-term feedback
L denotes long-term feedback

[1]Contrasts used in the above are:

C1 Cue-event validity coefficient feedback vs. cue-response utilization coefficient feedback.

C2 Cue-event validity coefficient plus cue-response utilization coefficient feedback vs. average of cue-event validity coefficient feedback and cue-response utilization coefficient feedback.

C3 Percentage correct feedback vs. average of other types of feedback.

Table A-5

Results of Multivariate Analysis of Variance on Each Feedback Group Separately for Subjects in Task B (Tabled Values Are Significance Probabilities Derived from the Distributions of Wilks' Lambda Criterion)

Feedback Type	Contrast[1]	Blocks of Trials			Degrees of Freedom HYP/ERROR
		1	2	3	
Percentage Correct	SL	.065	.029	.671	4/93
	L	.010	.494	.321	4/93
	S	.757	.758	.613	4/93
Cue-Event Validities	SL	.048	.248	.323	4/124
	L	.064	.433	.644	4/124
	S	.681	.636	.996	4/124
Cue-Response Utilization Coefficients	SL	.329	.354	.765	4/102
	L	.447	.373	.406	4/102
	S	.461	.943	.536	4/102
Both Cue-Event and Cue-Response Coefficients	SL	.208	.175	.228	4/101
	L	.009	.550	.008	4/101
	S	.013	.001	.305	4/101

[1]S denotes short-term feedback
L denotes long-term feedback