THE IMPACT OF EVALUATIVE FEEDBACK ON LEARNING AND PERFORMANCE IN A MULTIPLE-CUE DECISION PROBLEM

A THESIS

Presented to

The Faculty of the Division of Graduate

Studies and Research

By

Donald Edward Rose, Jr.

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Operations Research

Georgia Institute of Technology

May, 1974

THE IMPACT OF EVALUATIVE FEEDBACK ON LEARNING AND PERFORMANCE

IN A MULTIPLE-CUE DECISION PROBLEM

Approved:

Terrence Connolly, Chairman

Douglas Montgomery

T. L. Sadosky

Date approved by Chairman: 3 May 76

ACKNOWLEDGMENTS

I would like to thank Dr. Terrence Connolly, whose time and efforts made this study possible. I would also like to thank my reading committee, Dr.'s Douglas Montgomery and T. L. Sadosky, whose comments and corrections were very helpful.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF TABLES

LIST OF ILLUSTRATIONS

SUMMARY

It is not uncommon for decision makers to base their decisions on multiple information systems. Doctors evaluate a patient's multiple symptoms, stock brokers evaluate several conditions of the market, and college administrators must evaluate several facets of a potential student. In each case, the central process is one of reducing a number of error-prone informational inputs into a single estimate. Brunswik's lens model has proven to be a valuable tool in conceptualizing this type of judgment process and it is within the framework of this model that the present study was conceived.

In laboratory studies of judgment processes, it is usual to study the subject's learning when provided with feedback of the "correct" answers over a series of trials. As a variant to this, some researchers have provided the subject with information about the problem structure, or his information usage. However no reported study has examined the effect of what might be called "evaluative feedback", where the subject is told only how well he is doing (good or bad), rather than receiving a numerical feedback. This research undertook a laboratory study of the effects of evaluative feedback on learning of, and performance in, a multiple-cue inference task. The findings of the study have contributed both to our understanding of the processes involved in learning of such complex tasks, and to the generation of insights of value for management practice.

CHAPTER I

INTRODUCTION

Purpose

The purpose of this chapter is to provide the reader with an introduction to (Egon Brunswik's) lens model. The original formulation and its application to human judgment has been modified and enhanced by several other individuals and provides a workable framework for studying man's judgmental processes. This chapter will review several empirical studies within this framework leading to an explanation of the research conducted with this paper.

Historical Background

From the beginning of time, man has always been faced with decision making and the problem of how to improve his decision making ability. For practically the entire history of man, decisions have been made almost entirely by intuition. Although man could educate himself concerning a specific task, this education came about primarily through experience for man had yet to develop a method to transfer experience from one man to another. We have yet to completely understand the application of knowledge which in turn can not be approached without first understanding the learning process.

In 1956 Egon Brunswik published his theory of the learning process within what he called the "lens model". It was a most significant theory in that it linked the task with the individual attempting to learn that task (Edwards 1971). Central to Brunswik's work is the observation that the environment in which we operate is uncertain and probabilistic. For example, the phenomenon of size constancy depends on the subject's ability to integrate several error-prone size cues. Estimates of an object's size are based on cues (distance from the observer, angle to the observer, etc.) of differing ecological values (correlation between the object's actual size and the cues). These same cues, however, have different utilization values to the individual perceiving the object. Brunswik's general formulation of the lens model treats these dimensions within a single framework.

A particularly fruitful extension of the lens model has been its application to human judgment and inference phenomena such as diagnosis by clinical neurologists, interpretation of MMPI profiles (Kleinmuntz, 1968) and estimates of student performances (Hammond, Hursch and Todd, 1964). The common features of such situations are that an "expert" or "judge" makes some sort of inference or estimate of an underlying (distal) variable from several imperfect informational inputs (cues). When studying these phenomena consideration is given to the relationships between the distal variable and the set of interrelated cues. Consideration is then given to the relationships between these interrelated cues and the subject's response. The model therefore provides a basis for the study of complex human judgmental tasks. These relationships are displayed in Figure 1.

Definitions

The following terms are those primarily used when working with the

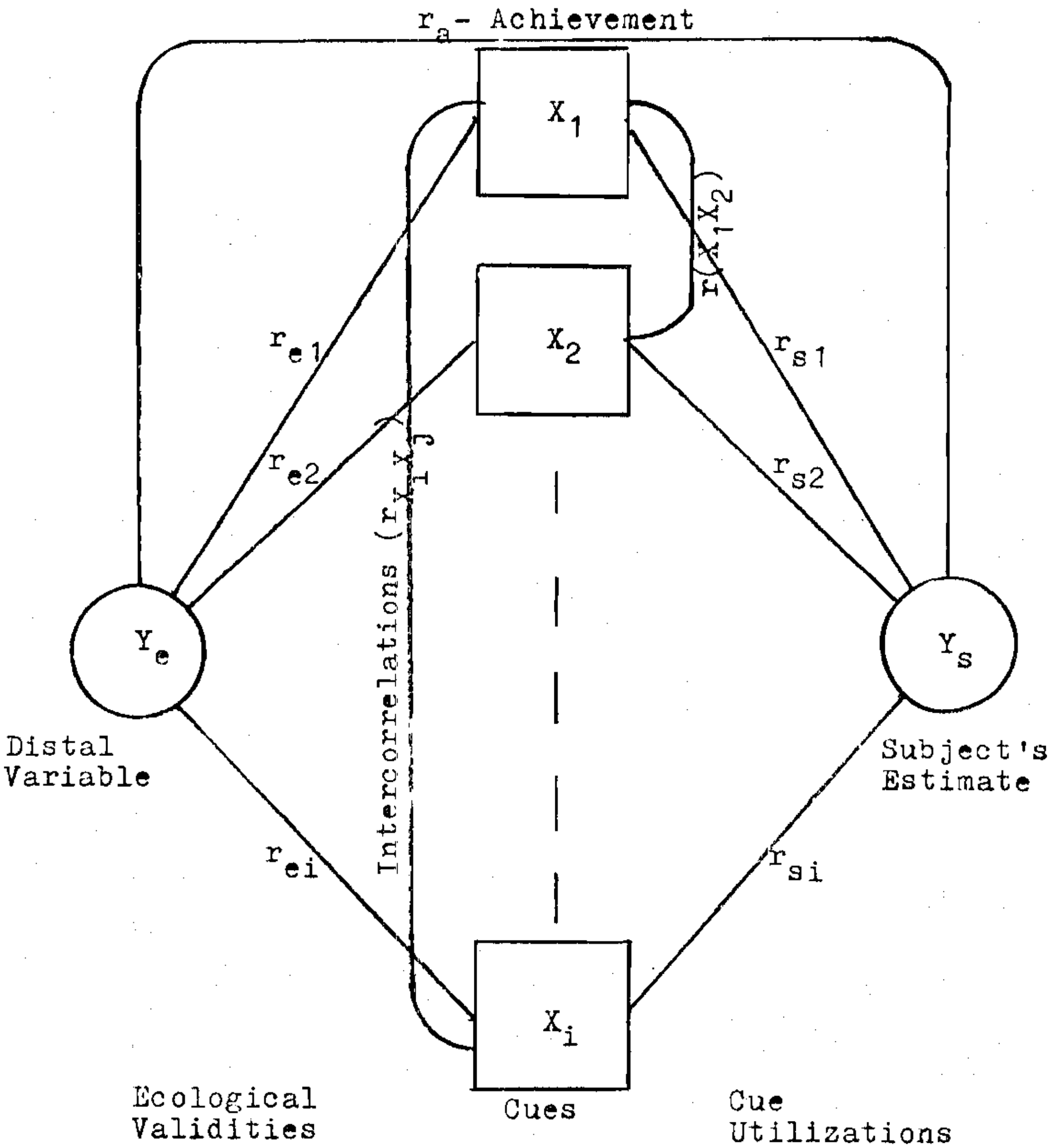

Figure 1. Brunswik's Lens Model

lens model:

Y_e - the established true state of the distal variable. The subject may or may not have access to its actual value.

X_i - represents the ith cue or information source describing the distal variable. These cues must be quantifiable if only at the 0-1 level (yes/no, hi/lo).

Y_s - the subject's response or his estimate of the value of the distal variable, Y_e.

$\hat{Y}_e$ - the best linear prediction of the distal variable from the cues.

$\hat{Y}_s$ - the best linear prediction of the subject's response from the cues.

r_{ei} - the simple correlation between an individual cue, X_i, and the distal variable. This is called the ecological validity of the ith cue.

r_{si} - the simple correlation between an individual cue, X_i, and the subject's estimate. This is called the subject's utilization coefficient for the ith cue.

r_a - achievement coefficient of the decision maker: the correlation between his prediction and the actual distal variable ($r_{Y_s Y_e}$).

R_e - the linear predictability of the distal variable: the correlation between the distal variable and the linear model of the distal variable ($r_{Y_e \hat{Y}_e}$).

R_s - the linear predictability of the decision maker: the correlation between the decision maker's predictions and the linear model of the decision maker ($r_{Y_s\hat{Y}_s}$).

G - the correlation between the predicted response from the linear model of the decision maker and those from the linear model of the distal variable. It has been called the "matching index" (Slovic and Tichtenslein, 1971) or a measure of the subject's "knowledge" (Hammond and Summers, 1972).

C - the nonlinear component of the judgmental accuracy: correlation between the residual values of the distal variable and the residual values of the decision maker's predictions after linear components in both have been removed.

In an important work, Hursch, Hammond and Hursch (1964) derived equations facilitating study of human judgmental tasks within the framework of the lens model. Tucker (1964) presented a simpler, more elegant formulation relating the various dimensions of the model in a concise manner. The following was taken from that article.

Assuming all variables standardized Tucker considered the variables $\hat{Y}_e$ and $\hat{Y}_s$ defined as;

$$\hat{Y}_e = \beta_{e1}X_1 + \ldots + \beta_{en}X_n$$
$$\hat{Y}_s = \beta_{s1}X_1 + \ldots + \beta_{sn}X_n$$

where the β's are the multiple linear regression weights for predicting Y_e and Y_s from the multiple cues.

Tucker then defines a residual error for each variable, Z_e and Z_s such that

$$Y_e = \hat{Y}_e + Z_e$$

$$Y_s + \hat{Y}_s + Z_s$$

From multiple linear regression theory

$$\mathrm{Var}(\hat{Y}_e) = R_e^2$$

$$\mathrm{Var}(\hat{Y}_s) = R_s^2$$

and

$$\mathrm{Var}(Z_e) = 1-R_e^2$$

$$\mathrm{Var}(Z_s) = 1-R_s^2$$

while

$$\mathrm{Cov}(xy) = r_{xy}\sqrt{V(x)V(y)}$$

By definition:

$$r_a = \mathrm{Cov}(Y_e Y_s) = \mathrm{Cov}(\hat{Y}_e \hat{Y}_s) + \mathrm{Cov}(Z_e Z_s)$$

and thus it can be shown by substitution where $G = r(\hat{Y}_e \hat{Y}_s)$ and $C = r(Z_e Z_e)$ that

$$r_a = GR_e R_s + C\sqrt{(1-R_e^2)(1-R_s^2)}$$

This is the basic equation used with the lens model when analyzing data.

Notice that this equation reduces to a simple relationship under any of the conditions a) $R_s = 1$, b) $R_e = 1$, or c) $C = 0$. Under these conditions the term

$$C\sqrt{(1-R_e^2)(1-R_s^2)}$$

disappears. Specifically if C = 0 (i.e. there is no systematic non-linear matching), the achievement correlation becomes simply a product of the knowledge (G) the subject has of the task, the predictability (R_e) of the environment and the predictability (R_s) of the subject's response system. That is

$$r_a = GR_eR_s$$

for linear task systems.

Studies Concerning Judgmental Strategies

With the lens model providing the framework there have been a wealth of studies attempting to represent the decision maker's policy in various task environments as a linear model. These tasks include personality characteristic judgments (Hammond, Hursch, and Todd, 1964); performance in college (Dawes, 1971; Einhorn, 1971); attractiveness of common stocks (Slovic, 1969); and physical and mental pathology (Goldberg, 1970; Oskamp, 1962; Wiggins and Hoffman, 1968). At least one thing emerged from all these studies that was of importance - the linear model does a very good job of predicting the judgments of the decision makers (Slovic and Lichtenstein, 1971).

Despite this overwhelming success of the linear model, there still has grown an active interest in what could be called "configural judges" (Goldberg, 1968) or those decision makers who use complex, non-linear strategies. Many early studies using nonlinear models of the judgmental policies, however, showed "no substantial increment of predictable response variance over that contributed by the linear model" (Slovic and Lichtenstein, 1971). But as was indicated by Slovic and Lichtenstein (1971) these findings did not preclude nonlinearity since a lack of nonlinearity between the environment and the response system could produce linearity indications. This fact along with 1) assertions by decision-makers that their processes really are complex; 2) possibilities that previous experimenters had not yet studied a truly configural model; and 3) possibilities that the experimental designs and statistical procedures of previous experiments were not optimally suited for discovering configural effects, kept the search alive for the configural thinker. From a study by Einhorn (1971) in which he concluded that cognitive complexity and mathematical complexity can often counterbalance each other (in other words a strategy with greater mathematical complexity might have a lesser cognitive complexity), it seems appropriate to assume that the configural process will play a greater role in the understanding of learning despite its limited ability to out predict the linear model (Slovic and Lichtenstein, 1971).

Subjective Policies and Self Insight

Another way to assess the weighting policy other than fitting a regression model to the subject's responses is to have the subject describe the relative weights he used in arriving at a response to the task. The correlation of these subjective weightings and the computed policies gives some indication of the insight the subject has into his actual strategy (see Hoffman, 1960). All studies in this area seem to agree that judges strongly overestimate their use of minor cues and strongly underestimate their use of major cues. Indeed across several studies it has been shown that three cues usually account for 80% of the predictable variance while one cue typically accounted for 40% of the variance (Slovic and Lichtenstein, 1971). A study by Slovic, Fleissner and Bauman (1972) which examined the relationship of self insight to amount of experience, indicated that over time self insight even decreases. It appears that the most experienced judges produce verbal rationales that are less trustworthy than those of the novice judges.

Studies Concerning Variation of Cues and Utilization

There have been quite a few studies that have examined the role of the cues themselves in determining a weighting policy. Slovic (1966) and Hoffman (1968) indicated that when certain cues agreed in their implications, subjects tended to weight them equally. If these cues disagreed, however, other cues were used to resolve the differences. Along these same lines Uhl and Hoffman (1968) hypothesized that an increase in the variability of a salient cue, across a set of stimuli, would lead to greater weighting of this cue by a decision maker. Behind this was the idea that a decision maker tries to make differential predictions and any cues which indicate a differentiation would only reinforce this tendency (Slovic and Lichtenstein, 1971). In testing this hypothesis, Uhl and Hoffman found that by varying a particularly strong cue the hypothesized effect was indicated. When they manipulated a minor cue, however, no effect was found.

Another variable which can be manipulated is the number of cues. Apparently there has been relatively little research done on the effects of varying the number of cues, despite what seems to be obvious interest. One early study by Hoffman and Blanchard (1961) (as reported in Slovic and Lichtenstein, 1971) generated some very interesting results. By increasing the number of cues they found that the subjects produced lower R_s values, decreased accuracy and lower response variance. Unfortunately they used a small number of subjects which somewhat limits their conclusions. Hayes (1964) and Einhorn (1971) reinforced this early study, with Einhorn interpreting the decrease in R_s as an indication of the subject's use of a more complex strategy with more information.

Another interesting finding is reported by Oskamp (1965) who used 32 "experts" and asked them not only to make a judgmental decision but also to give a confidence rating of their decision. When he increased the amount of information available to the subject he found that their confidence increased way out of proportion to the amount of increase in accuracy. The study provided us, therefore, with a small amount of evidence to indicate that by increasing the amount of information available to the decision maker we can increase his confidence while his decision quality is not increased and his decisions become more difficult to predict (Slovic and Lichtenstein, 1971).

One final concept that concerns cue variables is the idea of cue consistency. Hoffman (1968) described cue consistency in the following manner:

> 1) The consistency of a stimulus is the extent to which the distinguishing cues or characteristics of the stimulus match those of the population (of objects or persons) of which the

> stimulus is a member. 2) The consistency of a stimulus is the sum of the consistencies of the seperate cues. 3) The consistency of a stimulus is the degree to which the defining cues or characteristics of the stimulus, when considered in their totality by the judge, are perceived by him to be representative of an actual familiar, known or believable stimulus or object.

It is important, therefore, that the task cues be consistent not only with each other but with any perception the subject has with the problem, a consideration that must be made when designing an experiment. If findings are to be of any value to understanding "real world" learning situations then cue consistency would seem to be an important design criteria. Indeed not only should a consideration be made for cue consistency but as Connolly (1973) suggests,

> The non-conscious process of perception may be ... viewed as an (intuitive) "making sense" of a multiple stimulus situation, where the "knowledge" used is acquired unconsciously over a period of time from everyday experience, ...

In fact Connolly points out that factors such as cue interrelationships, types of feedback, verbal context, time pressures and possibly earlier experiences should be considered together, "and must be so designed as to form an internally consistent package".

Studies on Learning to Use Information

There have been quite a few studies in the area of learning information processing. One easy way to partition these studies would be according to availability of cues, distinguishing those studies with one cue available to the subject from those with multiple cues available. My focus here will be on multiple cue learning which seems to have relevance to a variety of "real world" situations in which a decision maker must integrate several sources of information. In the majority of these studies we find the lens model is used as a base for conceptualizing the problem. Although Bayesian studies pretty much parallel the lens model research, they seem rather uninterested in learning. There are, however, a handful of Bayesian studies that do consider learning, notably; Martin and Gettys (1968) who compared performance with nominal feedback or probabilistic feedback; Phillips and Edwards (1966) who studied performance under different payoff conditions; and Peterson, Ducharme and Edwards (1968) and Wheeler and Beach (1968) whose studies were "oriented to the misperception explanation of conservatism" (Slovic and Lichtenstein, 1971). In using the lens model, typically the independent variables are, 1) number of cues; 2) r_{ei} values and the multiple correlation R_e; 3) the various forms of the functional relationships between cues and the distal variable; 4) intercorrelation between cues; 5) feedback; and 6) a verbal context for the task. Most likely the subject will be given certain stimuli or cues, asked to respond with a judgment and then be given some form of feedback. So far major results have shown us (according to Slovic and Lichtenstein, 1971):

> 1) Subjects can learn to use linear cues appropriately (Lee and Tucker, 1962; Summers, 1962; and Uhl, 1963); 2) Learning of non-linear functions occurs but is slower than learning of linear relationships (Brehmer, 1969; Hammond and Summers, 1965; Sheets and Miller, in press; Summers, 1967; and Summers, Summers and Karkau, 1969) and is especially difficult if the subjects are not forwarned that the relations may not be nonlinear (Earle, 1970); Hammond and Summers, 1965; and Summers and Hammond, 1966); 3) Subjects can learn to detect changes in relative cue weights over time although they do so slowly (Peterson, Hammond and Summers, 1965; Summers, 1969); 4) It is easier for subjects to learn which cue to use than to discover which functional rule relates a known valid cue to the criterion; learning both of these simultaneously is especially difficult (Summers, 1967, 1969); 5) In a two cue task, pairing a cue of low or medium validity with one of high validity is detrimental to performance (a distraction effect), while pairing a cue of low validity with another of medium or

low validity is facilitative (Dudycha and Naylor, 1966); and 6) Subjects can learn to use valid cues even when they are not perceived with perfect reliability (Brehmer, 1970).

Feedback Manipulations

To date the majority of correlational studies have dealt with only two types of feedback, either outcome feedback where a subject is given the correct answer after every trial, or lens model feedback where the subject is provided information about the problem structure, or his information usage. From these studies emerged several findings that showed that lens model feedback was a more effective aid to learning than outcome feedback (Hammond and Boyle, 1970; and Todd and Hammond, 1965). In these particular studies, however, the ecological structure or criterion information feedback did not include random error. The single available study of the effects of error-prone feedback on learning and performance is that of Miklausich (1974). He argues that in most "real world" judgment tasks, feedback will typically be error-prone. His study showed that as misinformation increases in a multiple-cue probability learning task under error-prone feedback the achievement correlation will decrease.

Additionally one can argue that while the studies on lens model and outcome feedback showed the superiority of lens model feedback, it is unrealistic to assume that lens model feedback is readily available to the "real world" decision maker. Most individuals who make decisions concerning complex tasks do not have the handy access to a computer as does the scientist or the clinical analyst. It seems likely that "real world" judgment skills are commonly aquired under conditions of degraded feedback - i.e. error-prone, delayed, intermittent, evaluative, or even non-existent. There does seem to be, therefore, a whole set of decision makers who utilize multiple information systems and receive feedback that does not provide either the ecological information or the exact information concerning the task. Instead they receive vague evaluations of how well they are doing rather than a quantified feedback such as in outcome or lens model feedback. Because of the evaluative nature of this type of feedback I chose to call it "Evaluative Feedback". In the academic world evaluative feedback is evidenced by a pass/fail grading system in which a student is evaluated as either passing or failing. Since many factors could affect this evaluation, there seem to be many pieces of information represented by one small piece of information. In the business world a pay raise or refusal of a pay raise indicates an evaluation of sorts to the individual seeking the raise. If he gets the raise he knows he is doing well; if he does not get the raise he may think he is doing poorly. A production manager might look at the rise or fall of the production rate as an evaluation of his policies. The point is that in all these situations the subject is receiving only a vague evaluation of how well he is doing rather than some quantified numerical feedback. Since evaluative feedback appears to be a common "real world" learning situation, research can be of value in this area.

Dimensions of Evaluative Feedback

Evaluative feedback, then is any feedback which provides the subject information on how well he is performing rather than receiving some type of numerical feedback. Although within the framework of the lens model several aspects of evaluative feedback may be varied, the dimension of interest here is the number of classes of evaluative feedback available to the subject.

Research Question

"What are the effects of evaluative feedback in learning and performance in a multiple-cue probability learning problem?"

Hypothesis

In a multiple-cue probability learning task under evaluative feedback, achievement will increase as the number of classes of evaluative feedback are increased.

A laboratory experiment designed to test this hypothesis is described in the following chapter.

Summary

In summary, then, we conceive human judgment as a setting in which decision makers base their decisions on multiple error-prone information sources. Doctors evaluate a patient's multiple symptoms, stock brokers evaluate several conditions of the stock market and college administrators must evaluate the several facets of a potential student. In each case, the central process is one of reducing a number of error-prone informational inputs into a single estimate, judgement, or evaluation. Brunswik's lens model has proven to be a valuable tool in conceptualizing this type of judgment process and it is within the framework of this model that the present study is conceived.

In laboratory studies of judgment processes, it is usual to study the subject's learning when provided with outcome feedback over a series of trials. As a variant of this procedure, some researchers have provided the subjects with information about the problem structure, or their information usage. However, no reported study has examined the effect of what could be called evaluative feedback, where the subject is told only how well he is doing rather than receiving a numerical feedback. Since evaluative feedback seems a common "real world" learning situation, it appears that research would be of value. The proposed research will undertake a laboratory study of the effects of evaluative feedback on learning of, and performance in, a multiple-cue inference task. The findings of this study will, it is hoped, contribute both to our understanding of the processes involved in learning of such complex tasks, and to the generation of insights of possible value for management practice.

CHAPTER II

EXPERIMENTAL DESIGN

Purpose

The purpose of this chapter is to explain to the reader what criteria were considered in designing this experiment, how it was designed and what methods or procedures were used in administering the actual experiment and collection of data.

Task Design

In order to collect the data a hypothetical situation had to be constructed so the subjects would have a reasonable task to perform. A reasonable task is one that is neither too easy (a task in which all subjects achieve and hold the same level of learning after only a few trials) nor too hard (a task in which no subject is able to display any learning despite the number of trials). As expressed by Hoffman (1968), "It should [also] be recognized ... that certain patterns or configurations of cues ought, collectively, to 'go together', to 'be congruent' or to be 'consistent' in order for them to identify meaningful objects or persons". This means that the cues, when presented to a subject in a specific environment, should be representative of an actual, familiar, known or believable stimulus or object (Hoffman, 1968). The factors, therefore, that the subject looks for to "make sense of the problem", such as verbal context, interrelationship of the cues, and relevance to any prior experience should be designed so as to form a consistent package (Connolly, 1972).

To test the hypothesis proposed in Chapter I it was necessary to design a moderately complex linear task with four treatments corresponding to four separate groups of subjects. Group I received a basic verbal context and was given outcome feedback; the purpose of this group was to act as a base line with which to compare the performances of the remaining groups. Group II received the basic verbal context plus they had 2 classes of evaluative feedback available. Group III had the basic verbal context plus 3 classes of evaluative feedback while Group IV had the basic verbal context plus 5 classes of feedback made available. All four groups received their first 40 trials with no feedback so that performance with feedback could be compared to performance without feedback.

Verbal Context

The basic verbal context which was common to all groups placed the subject in the position of someone desiring to estimate the future performance of the stock market (perhaps in fun or to determine when to invest). Before he makes this estimate, however, the subject is to solicit the opinions of three local stockbrokers on what they think the daily change will be. With three opinions in hand the subject was asked to make his own prediction. In addition to the verbal context each of the groups was told that after each of their predictions they would receive either a) the actual market performance for that day (Group I) or b) an evaluation of their predictions based on actual market performance (Groups II, III and IV).

Task Properties

The problem used for this experiment was a three cue task in which the distal variable (Y_e) represented the daily change of the stock market as expressed by the Dow Jones average. Y_e was designed to be normally distributed with an average change of ± 10 points arbitrarily chosen as the standard deviation and with a mean of 0. This generated a hypothetical set of stock market changes with 67% of the changes being within ± 10 points.

The cues, or opinions of the local experts, were generated by means of the following equations;

$$X_i = Y_e \pm e_i \qquad i = 1,2,3$$

where

$$e_1 \sim N(0,225),\ e_2 \sim N(0,25),\ e_3 \sim N(0,100).$$

With the use of e_i containing different variances, it was possible to generate three realistic error-prone stock market experts of varying accuracy. The sample ecological validities (r_{ei}) (the correlations of the experts predictions to the actual market changes) were:

$$r_{e1} = .63 \qquad r_{e2} = .90 \qquad r_{e3} = .67$$

This gave the subject one fairly competent expert (X_2) and two relatively poor experts (X_1 and X_3). By use of a random number generator a set of cues was generated that could be displayed to the subject in a random manner. R_e, or the limits to achievement, for the first 40 trials was

.938 while it dropped to .934 for the second 40 trials.

Evaluative Feedback Criteria

Once the basic problem was designed a means of evaluating subjects had to be determined. Since the subjects would be receiving evaluations based on the actual stock market performance, Y_e was used as the evaluation guide. The subjects would simply receive an evaluation of "good" if they were relatively close to Y_e or a bad evaluation if they were more remote. With this in mind the evaluations were formulated as seen in Table 1.

Although the intervals for the evaluations were arbitrarily chosen, there was some consideration given as to how close to the distal variable the majority of the subjects would respond.

Procedure

An overhead projector was used to display the sets of cues to the subjects. At the same time the subjects were able to read the cues from their answer sheet. The transparency displayed the cues as seen in Figure 2. Samples of the answer sheets are displayed in Figures 3-5.

The answer sheets, which were reproduced by the Xerox copier, were placed at selected desks at the test site to preclude observation of another subject's work as much as possible. Pencils were also provided to the subjects.

For the first 40 trials, testing of each group was identical. The cues were displayed as seen in Figure 3 and no feedback was provided. In the second half of the experiment the different groups received the feedback for which they were designated. In the case of the outcome feedback

Table 1. Evaluation Given *vs* Respondent Error for 2, 3, and 5 Class Feedback Groups

	Response	Evaluation
Group II	$\lvert Y_s\rvert \leq \lvert Y_e \pm 2.0\rvert$	"good"
	$\lvert Y_e \pm 2.0\rvert < \lvert Y_s\rvert$	"bad"
Group III	$\lvert Y_s\rvert \leq \lvert Y_e \pm 1.0\rvert$	"good"
	$\lvert Y_e \pm 1.0\rvert < \lvert Y_s\rvert \leq \lvert Y_e \pm 2.5\rvert$	"fair"
	$\lvert Y_e \pm 2.5\rvert < \lvert Y_s\rvert$	"bad"
Group IV	$\lvert Y_s\rvert \leq \lvert Y_e \pm .5\rvert$	"very good"
	$\lvert Y_e \pm .5\rvert < \lvert Y_s\rvert \leq \lvert Y_e \pm 1.5\rvert$	"good"
	$\lvert Y_e \pm 1.5\rvert < \lvert Y_s\rvert \leq \lvert Y_e \pm 2.0\rvert$	"fair"
	$\lvert Y_e \pm 2.0\rvert < \lvert Y_s\rvert \leq \lvert Y_e \pm 4.0\rvert$	"bad"
	$\lvert Y_e \pm 4.0\rvert < \lvert Y_s\rvert$	"very bad"

	x_1	x_2	x_3
1.	-19.6	-4.0	-6.0
2.	39.4	25.8	30.9
3.	-10.5	1.0	-.1

Figure 2. Cues as Displayed on Screen.

	X_1	X_2	X_3	YOUR PREDICT.
1.	-19.6	-4.0	-6.0	
2.	39.4	25.8	30.9	
3.	-10.5	1.0	-.1	

Figure 3. Sample Answer Sheet for the 1st 40 Trials.

	X_1	X_2	X_3	YOUR PREDICT.	ACTUAL CHANGE
41.	-22.9	-22.4	-35.5		
42.	-11.3	1.9	5.4		
43.	-2.0	8.4	-3.6		

Figure 4. Sample Answer Sheet for Outcome Feedback Groups.

	X_1	X_2	X_3	YOUR PREDICT.	
41.	-22.9	-22.4	-35.5		G F B
42.	-11.3	1.9	5.4		G F B
43.	-2.0	8.4	-3.6		G F B

Figure 5. Sample Answer Sheet for Evaluative Feedback (Group III).

group (I), they were given the actual stock morket change after each trial and asked to record that value in the space provided. They were told that this could be used in making future predictions, however no scratch work was authorized. In the case of the evaluative feedback groups (II, III, IV) the tester went from subject to subject to individually evaluate each response. This was done so as not to bias other subjects' performances. Again no scratch work was allowed. The tester was able to evaluate each subject by merely taking a quick glance at a precomputed chart which listed the intervals for the appropriate evaluations. This chart is listed in the appendix.

At the conclusion of the experiment the subjects were asked to answer the following questions in an effort to gain some insight into their understanding of the problem.

1. How did you set about making your predictions?

2. Did you feel that any one of the "experts" was better or worse than any of the others? If so, which were better, which worse?

3. Did you feel the feedback you got in the second half of the experiment helped you make a better prediction?

Subjects

The subject pool consisted of unpaid students attending the Georgia Institute of Technology. Each member of the pool was assigned to a group (5 per cell) solely on the basis of their random volunteering.

CHAPTER III

RESULTS

Purpose

This chapter deals with the results of the experiment. The first section deals with the testing and analysis of the major hypothesis while the second section will deal with secondary analysis of the data.

Primary Analysis

The major hypothesis was previously stated as; Hypothesis - In a multiple-cue probability learning task under evaluative feedback, as the number of classes of evaluative feedback increases achievement will also increase.

The data in all four groups was analyzed using the STATPAC routine, RESTEM, to derive regression equations to fit the models of each subject's estimates and therefore providing the parameters $\hat{Y}_e$ and $\hat{Y}_s$. These intermediate parameters were then used to generate the necessary correlations, r_a, R_e, R_s, and G by means of the STATPAC routine CORAN. R.A. Fischer's Z-transformation was then applied to these correlations since the distribution of Z values is approximately normal (McNemar, 1969). This allowed averaging and subsequent analysis of variance.

The hypothesis concerned the comparison of achievements reached by the four groups and therefore, r_a was plotted and analyzed, first for the no feedback condition and then for the feedback condition. The values for R_e, limit of achievement, R_s, cognitive control, and G, knowledge of the task were also plotted to aid with interpretation. The data provided strong support for a direct contradiction to the major hypothesis. It indicated that as the number of classes of evaluative feedback are increased the level of performance deteriorates. Results and analysis are displayed in Figure 6 and Tables 2 and 3.

Notice that in Table 2, which analyzes the mean r_a across groups for trials 1-40, there was no significant difference in performance. But Table 3, which analyzes performance under feedback, indicates there is a strong deterioration effect due to feedback.

Secondary Analysis

Recall that the experiment was divided into two sets of 40 trials, the first 40 trials receiving no feedback and the second 40 receiving the appropriate feedback for that group. Because individual subjects in each group are completely different it was necessary to investigate the individual gains of each subject from a no feedback situation to a feedback situation. By doing this it was seen if the subjects' improvement or deterioration was caused by the effect of feedback. In all groups a change in achievement appears to be associated with the introduction of feedback. In the case of evaluative feedback the gain was a negative one thereby supporting the results of the primary analysis. This analysis also indicates that evaluative feedback is worse than no feedback, a very counter-intuitive finding indeed (it seems that some information should be better than no information at all). These results and analysis are displayed in Figure 7 and Table 4. It was also desired to gain insight into the problem by determining if the subject's performance was

Table 2. Mean r_a by Groups, Trials 1-40

	Mean r_a (based on Z-transformation
Group I (outcome feedback)(n=5)	.858
Group II (2 class feedback)(n=5)	.893
Group III (3 class feedback)(n=5)	.908
Group IV (5 class feedback)(n=5)	.869

ANOVA (on Z-transformations)

Source of Variation	D.F.	M.S.	F-ratio
Feedback	3	.05383	1.9116 (N.S.)
Error	16	.02816	
Total	19		

Table 3. Mean r_a by groups, Trials 41-80

	Mean r_a (based on Z-transformations)
Group I (outcome feedback)(n=5)	.891
Group II (2 class feedback)(n=5)	.833
Group III (3 class feedback)(n=5)	.786
Group IV (5 class feedback)(n=5)	.708

ANOVA (on Z-transformed r_a values)

Source of Variation	D.F.	M.S.	F-ratio
Feedback	3	.265	8.761 ($p \leq .001$)
Error	16	.0302	
Total	19		

Duncan's Multiple Range Test (Appendix, p. 65) shows

Gp I $\neq$ Gp III

Gp I $\neq$ Gp IV

Gp II $\neq$ Gp IV

Table 4. Gain Scores Between First and Second Blocks, by Groups

Note: Gain Score = d = $[r_a(\text{Trials } 41\text{-}80)] - [r_a(\text{Trials } 1\text{-}40)]$

	Mean Gain Score	t_o^*	p**
Group I (outcome feedback)(n=5)	.0336	2.806	($p \leq .05$)
Group II (2cl feedback)(n=5)	-.0570	-2.856	($p \leq .05$)
Group III (3cl feedback)(n=5)	-.1218	-4.912	($p \leq .01$)
Group IV (5cl feedback)(n=5)	-.1758	-3.849	($p \leq .025$)

* $t_o = \dfrac{\bar{d}}{S/\sqrt{n}} \sim t_{\alpha/2,n-1}$ (Hines and Montgomery, 1972)

** Null Hypothesis: Gain Scores and distributed normally, with mean zero.

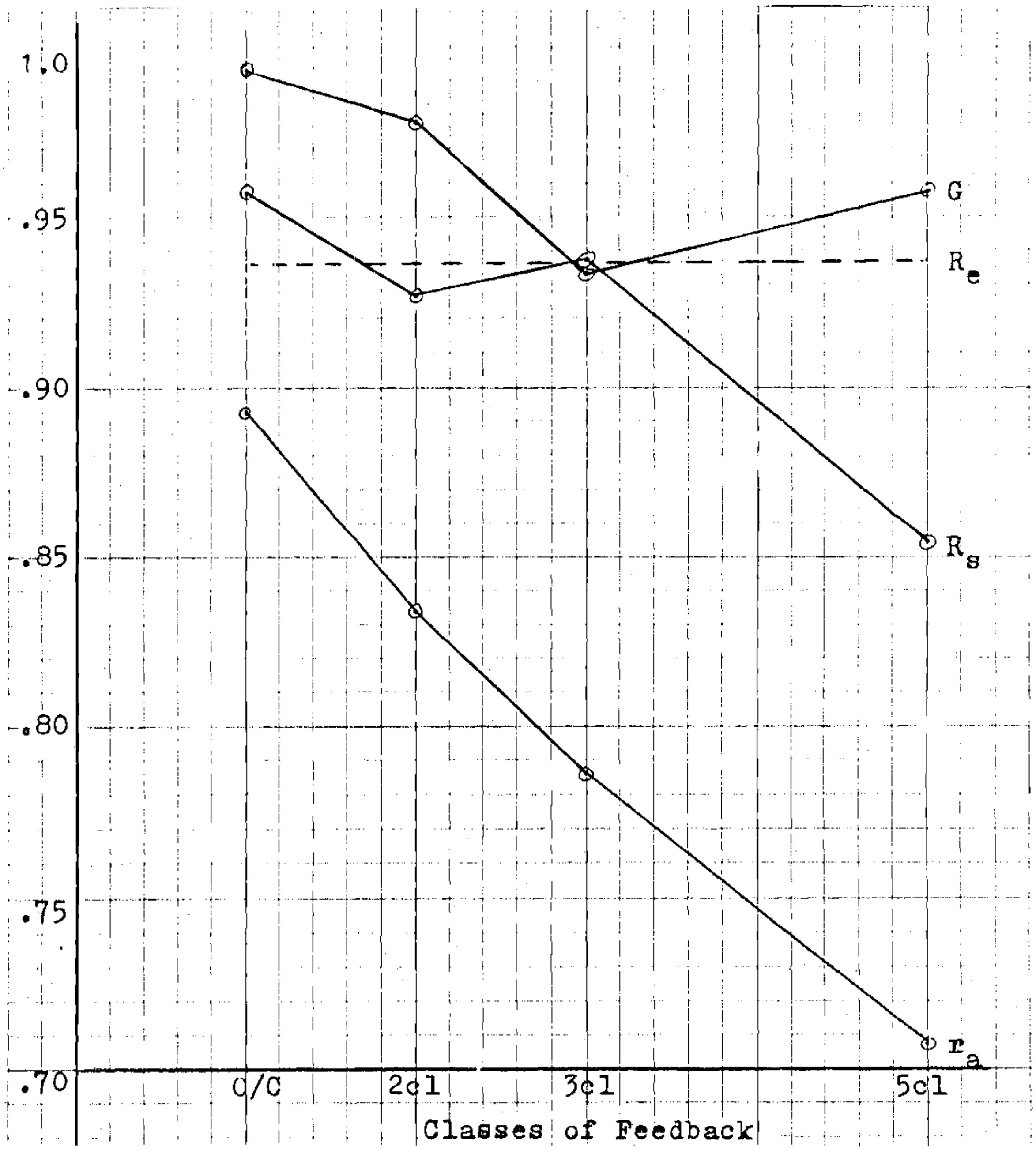

Figure 6. Mean r_a, R_s, and G by Groups, Trials 41-80.

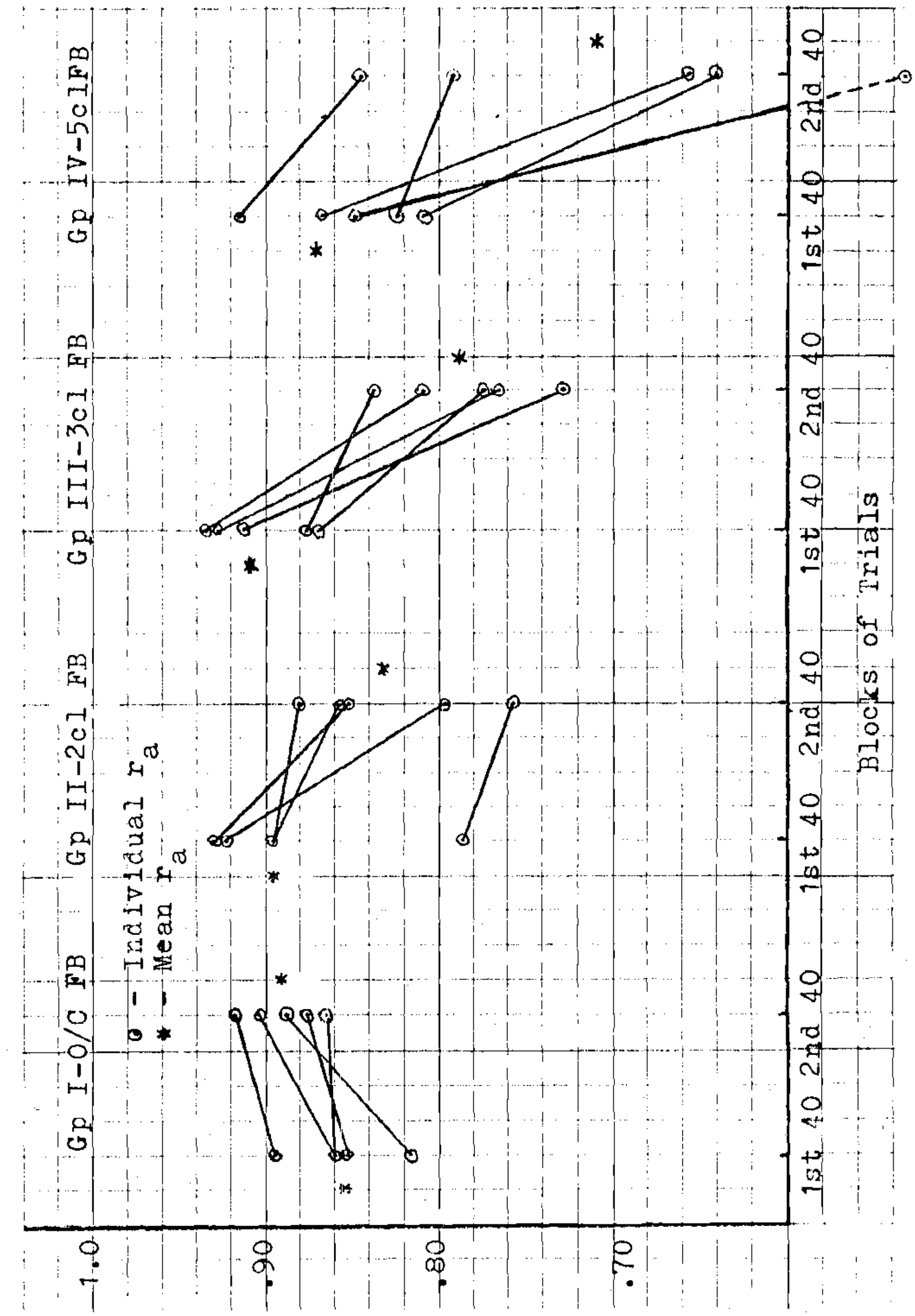

Figure 7. Individual Gain Scores Between First and Second Blocks, by Groups.

due primarily to R_s, cognitive control, or G, knowledge of the problem. Hammond and Summers (1972) suggest that "aquisition of knowledge and application of knowledge are independent components" and that they can be separated empirically. It is of interest, then, to determine whether changes in G or changes in R_s are more strongly associated with the deterioration in performance, r_a.

To examine this matter, we calculate:

$r_{(r_a,R_s)}$ - the correlation between achievement, r_a, and the correlation of the subject's linear regression equation estimates and his actual estimates of the distal variable; and

$r_{(r_a,G)}$ - the correlation between achievement and the correlation of the subject's linear regression equation estimates and the linear regression equation estimates of the distal variable.

The data yielded the following values:

$$r_{(r_a,R_s)} = .8993 \qquad r_{(r_a,G)} = .3270$$

suggesting that R_s, rather than G, is primarily responsible for the observed change in r_a.

In order to test the statistical significance of the difference, the statistic W suggested by Graybill (1961) was used on Z-transformed values of these correlations. The results are summarized in Table 5.

Table 5. Test of Statistical Significance of Difference Between $r_{(r_a,R_s)}$ and $r_{(r_a,G)}$.

	$r_{(r_a,R_s)}$	$r_{(r_a,G)}$
Z	1.4685	.3395
n-3	37	37
Z(n-3)	54.3345	12.5615

$$W = \sum_{i=1}^{2} (n_i - 3)(Z_i - Z)^2 \sim \chi^2 \text{ with 1 df}$$

$$W = 23.5808 \quad (p \leq .0001)$$

This indicates that there is in fact a highly significant difference between $r_{(r_a,R_s)}$ and $r_{(r_a,G)}$ which in turn implies that the subject's achievement failure was due primarily to the failure of the subject to consistently apply his own regression equation model. In other words it seems that although the subject's potential for application (knowledge of the task) was relatively high, his inability to apply this knowledge (cognitive control) prevented a potentially high achievement (Hammond and Summers, 1972). Using Hammond and Summers theory then, the poor performance seems to be due primarily to loss of cognitive control.

A final analysis was made to determine if there were any positive change in performance over the second 40 trials. If there was such a change then this might indicate that more trials were needed to uncover the real effect of evaluative feedback. To do this r_a and R_e were recomputed for trials 41-60 and trials 61-80 and a ratio r_a/R_e was defined as percent achievement. This ratio was formed because there was a reasonably large change in R_e for the last 20 trials (R_e for trials 41-60 was .958, R_e for trials 61-80 was .885), making it possible for a subject's performance to appear to deteriorate while in fact his percent achievement would be increasing. This analysis (Table 6, Figure 8) found no significant learning within the second block of 40 trials.

Post-Experimental Questionnaire Results

Recall that each subject was asked to answer three questions at the termination of the experiment. After reviewing these comments it was discovered that the subjects did indeed become more confused as the number of classes of evaluative feedback increased. Of the subjects exposed to outcome feedback (n=5), four said that they felt expert no. 2 (r_{e_2} = .9) was the most accurate. Of the subjects who received 2 class evaluative feedback (n=5) three said that they too felt expert no. 2 was the most accurate. Of the subjects under 3 class and 5 class evaluative feedback, however, only one from each group noticed the more accurate expert. Similarly, in response to question no. three (Did the feedback help you any way?), as the number of classes of evaluative feedback increased, the comments indicated that help decreased (See Table 7). In response to question one (Describe the strategy used to arrive at a prediction) there seemed to be the common strategy across all four groups of straight averaging. Although these comments do support the major hypothesis, it should be noted that self-insight has been shown to be rather untrustworthy in explaining individual strategies.

Table 6. Percent Achievement by Groups, Trials 41-60 and 61-80

Groups	Percent Achievement ($\bar{r}_a/R_e$)	
	Trials 41-60	Trials 61-80
I (O/C FB)	.940	.990
II (2cl FB)	.880	.920
III (3cl FB)	.840	.860
IV (5cl FB)	.850	.740

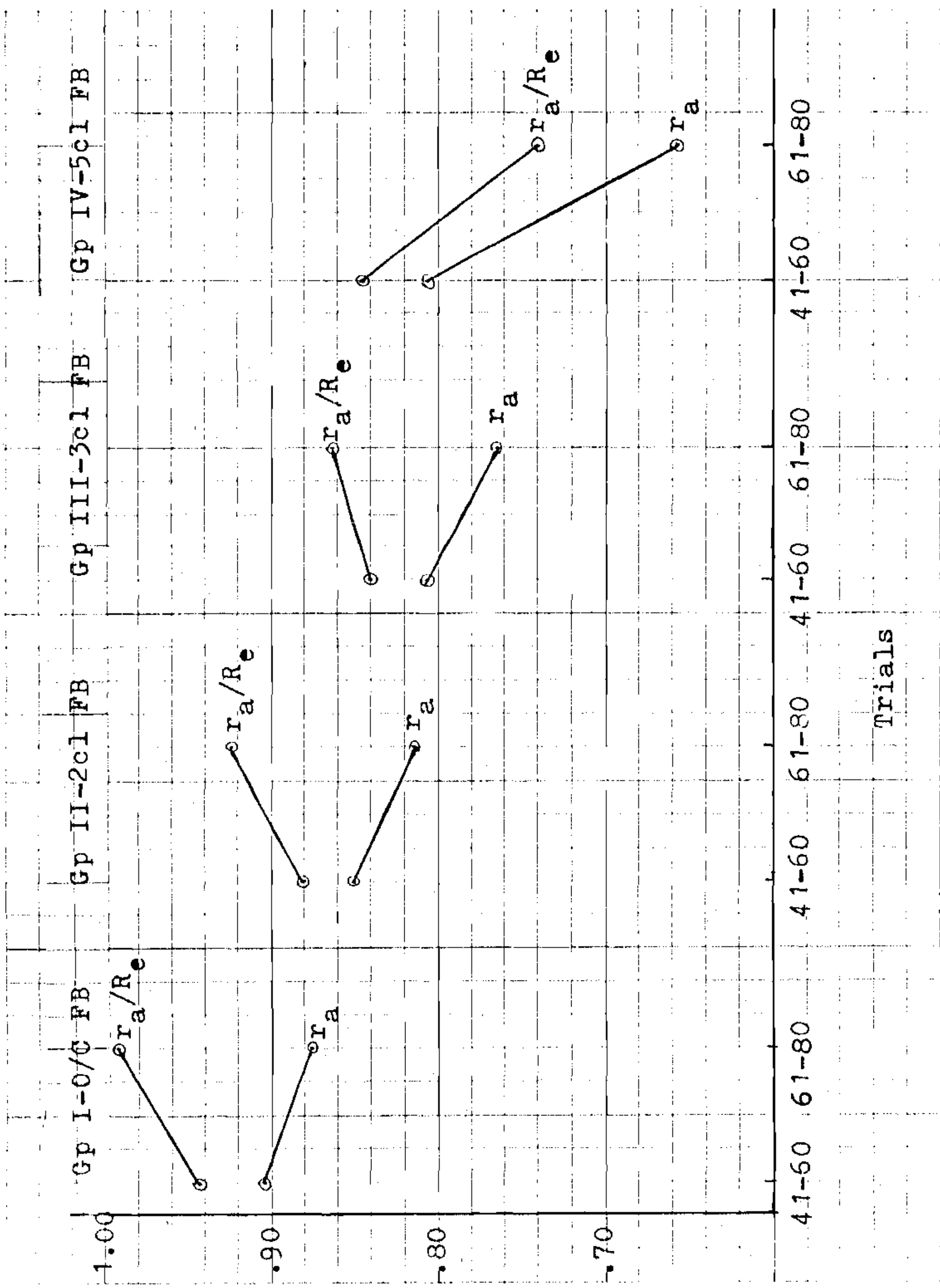

Figure 8. Percent Achievement by Groups, Trials 41-60 and 61-80

Table 7. Tabulation of Answers to Post-Experiment Questions

	Picked X_2 as the Best Expert	Said Feedback was of Assistance
O/C FB Gp I	4	4
2cl FB Gp II	3	3
3cl FB Gp III	1	2
5cl FB Gp IV	1	0

Analysis of Treatment Effects

As noted earlier, the cut off scores for the various evaluative feedback categories were set arbitrarily (Table 1). Table 8 below, shows the breakdown of feedback actually received for groups I, II, and III and shows clearly that the levels chosen were too severe. In fact, the evaluations which would have been received by an imaginary subject using the optimal (regression equation) strategy would have been only moderate (i.e. about 50% "good" evaluations and 50% "bad" evaluations in Group II). This suggests that the present results may apply only to "severe" evaluative feedback. Replication with less stringent cut-off points is clearly desireable.

Summary

In summary then, the data provide fairly strong support for a direct reversal of the major hypothesis. It was shown that achievement deteriorates as the number of classes of evaluative feedback increases. Secondary analysis indicated that not only was achievement under evaluative feedback worse than that under outcome feedback, but it was worse than performance with no feedback at all. It appears that this deterioration is due largely to the inability of the subjects to apply their knowledge of the task.

Table 8. Breakdown of Actual Feedback Received by Groups II, III, and IV.

Group II (2cl FB) Subject		G		B	
6		8		32	
7		11		29	
8		10		30	
9		12		28	
10		12		28	
Mean		10.6(26.5%)		29.4(73.5%)	
Group III (3cl FB)		G	F	B	
11		4	5	31	
12		5	9	26	
13		5	7	28	
14		3	5	32	
15		4	7	29	
Mean		4.2(10.5%)	6.6(16.6%)	29.2(73.5%)	
Group IV (5cl FB)	VG	G	F	B	VB
16	4	6	1	5	24
17	2	4	8	4	22
18	1	8	4	7	20
19	2	3	6	12	17
20	2	7	5	5	21
Mean	2.2(5.5%)	5.6(16%)	4.8(12%)	6.6(16.5%)	20.8(52%)

CHAPTER IV

CONCLUSIONS AND RECOMMENDATIONS

Purpose

The purpose of this chapter is to discuss the results as presented in the previous chapter and their implications, both theoretical and practical. Finally recommendations will be made as to what future research should be conducted to verify and expand the results of this paper.

Discussion

The major hypothesis was directly contradicted by the data as presented. In this experiment all groups that received evaluative feedback not only performed at a lower achievement level then the control group (outcome feedback group), but their performance seemed to deteriorate in direct proportion to the number of classes of evaluative feedback available. The performance level was also lower than performance under a no feedback condition. This deterioration seems to be caused primarily by a loss of cognitive control, R_s.

Let us discuss now some possible artifactual explanations of this performance. The first of these that comes to mind is that the subjects may have gotten tired or even bored thereby causing the poor performance. If this was the case one would expect to find no significant difference in deterioration between groups. As was seen, however, there was indeed a significant difference across groups due to feedback. It would seem therefore that boredom did not cause any deterioration of performance. It is possible, however, that the subjects became discouraged due to a preponderance of bad evaluations. It could be argued further that because a subject receives a low evaluation out of many evaluations rather than a low evaluation out of a few evaluations, he would become more discouraged. This would explain the inverse proportion of performance to number of classes of evaluative feedback.

There might also be some real effect explanations. Because of the nature of sampling error it was quite possible for a subject to have an appropriate model yet still receive bad evaluations to several responses. Basically one could be close without being good. This might have led to early false abandonment of an appropriate model. Note that all subjects in the evaluative feedback groups had relatively good models for their first 40 trials but when feedback was introduced they apparently abandoned these models for worse ones. The outcome feedback group on the other hand apparently substituted better models for the second 40 trials, as was expected.

Outcome feedback is also superior in that it provides the subject with two possible alternatives. First he can check his current model to see if it is consistent and second he may use the outcome feedback to provide some insight into the actual task model and therefore assist him in making his next prediction. With evaluative feedback, however, the subject can only check his current model. Because he knows nothing of the evaluation criteria he gains no insight to the model.

An immediate, though tentative, implication of these data is that if one can not give at least outcome feedback then do not give any feedback at all, at least where other aids such as verbal context are available. Since more empirical study needs to be done across a range of problems and settings, this is only tentative.

These findings can be applied to several areas, a few of which will be discussed here. First consider the "classroom" setting where evaluative feedback commonly exists in the form of grades. It was seen in this experiment that the more evaluative feedback available to the subject the worse his performance was. This would seem to indicate that rather than an evaluation the student might learn more if after tests he were given the correct answers and the proper methods of arriving at them. An evaluation would apparently only serve to confuse the student.

These findings could also apply to the managerial field where decision makers are currently evaluated, usually after blocks of several decisions. The subjects of this experiment became confused when feedback was given after each trial, a fact that could become magnified if the same feedback were given after more than one trial. Managers could probably obtain better performance from their subordinates if they explained exactly what is expected and if the job is done improperly, what was wrong about and how it should have been done. Any type of evaluation should be disgarded as it appears they show no positive contribution to learning for individuals.

Recommendations

The results of this experiment raised some interesting questions concerning the role of feedback in a multiple-cue probability learning task. Obviously more experiments need to be conducted to both confirm and supplement the findings presented in this paper. A particular replication under less "severe" evaluation cut-points is obviously required. By increasing the chances for a good evaluation would discouragement be eliminated, thereby possibly raising the level of achievement? Would harsher or milder evaluative adjectives affect performance in any different way? It would also be interesting to find out what effect, if any, a combination of evaluative and outcome feedback would have on performance. These are questions that should be addressed in future research as well as conducting replications of this experiment.

Within the framework of the Brunswikian lens model we have seen studies dealing with lens model feedback, outcome feedback, error-prone feedback and evaluative feedback. These studies indicate that lens model feedback generates the best performance with error-prone and evaluative feedback generating the least desireable performance. Within a "real world" setting, however, it seems unlikely that lens model feedback would normally be available. Outcome feedback is probably the best feedback that a decision maker could hope for and even then, learning was shown to be extremely slow for complex tasks (Goldberg, 1968). Yet in a setting such as the "real world" where most tasks are considered to be of a complex nature, and in all likelihood the feedback received is either error prone or evaluative, decision makers are having apparent rapid success in learning these complex tasks. What is it then that aids these decision makers, since it is obviously not the feedback received? Ideally more research should be conducted under more realistic settings or in the "real world" setting itself. Since this is not always easily done, future research should consider the total environment from which the subject "makes sense of the problem", such as prior training and experience, error within feedback, and the outside pressures normally associated with decision making. It could be as Yntema and Torgerson (1961) suggested that "good judgment may turn out to be a simpler matter than we would like to think it is," or it could be that training and prior experiences enables individuals to make sense of the problem quickly even without good feedback.

APPENDICES

Computer Program Used to Generate Cues and Distal Variable

```
5 FØR I=1 TØ 80
10 Y=10*(RND+RND+RND+RND+RND+RND+RND+RND+RND+RND+RND+RND-6)
15 A=Y+(15*(RND+RND+RND+RND+RND+RND+RND+RND+RND+RND+RND+RND-6))
20 B=Y+(5*(RND+RND+RND+RND+RND+RND+RND+RND+RND+RND+RND+RND-6))
25 C=Y+(10*(RND+RND+RND+RND+RND+RND+RND+RND+RND+RND+RND+RND-6))
30 PRINT 'X1= ';A,'X2= ';B,'X3= ';C,'Y= ';Y
35 NEXT I
40 END
```

```
X1= -19.624168  X2= -3.9989311  X3= -6.2498182  Y=  2.334801
X1=  39.39876   X2=  25.813257  X3=  30.905101  Y=  30.027993
X1= -10.48587   X2=  1.0544106  X3= -.0978148   Y= -1.4093173
X1=  33.663068  X2=  .16153008  X3=  2.4814129  Y=  .95938623
X1= -5.9556642  X2= -5.4998028  X3= -3.766765   Y= -6.1476237
X1=  26.771213  X2=  4.1116992  X3= -.33698797  Y=  6.8935078
X1=  .76346546  X2= -3.2658505  X3= -9.0839618  Y= -1.6603994
X1= -4.7747409  X2= -8.9307525  X3= -21.638207  Y= -6.003713
X1=  27.384548  X2=  10.331969  X3= -3.6585015  Y=  1.9414842
X1=  7.2315493  X2= -12.815439  X3= -26.758162  Y= -10.808224
X1=  9.9832722  X2= -.02785981  X3= -7.7220023  Y=  .23245633
X1=  19.709723  X2=  3.3838397  X3=  9.0110648  Y=  1.4965516
X1=  8.7869191  X2=  8.9501905  X3=  9.3556225  Y=  9.8828912
X1= -11.22158   X2= -4.9530998  X3= -10.334577  Y= -6.5925014
X1=  .32041609  X2= -6.1006871  X3= -9.0824062  Y= -4.95094
X1= -8.2612618  X2=  .28280348  X3= -19.566303  Y= -2.0802552
X1= -27.190607  X2= -7.5991276  X3=  12.123707  Y= -4.1671908
X1=  26.304834  X2=  4.8412994  X3=  11.100978  Y=  8.4078473
X1=  .09909064  X2=  6.4872962  X3=  .23993611  Y=  1.6438937
X1=  13.382414  X2=  6.5888349  X3=  21.666641  Y=  11.185948
X1= -4.2664383  X2= -11.510548  X3= -20.889497  Y= -13.619338
X1= -24.619184  X2= -6.7410949  X3= -12.75806   Y= -5.514313
X1=  7.8944012  X2= -3.3834443  X3= -11.186538  Y= -5.622372
X1= -40.287601  X2= -17.682472  X3= -24.979992  Y= -20.225407
X1=  23.088141  X2= -2.5407988  X3=  1.7705774  Y=  4.6776414
X1= -1.3540468  X2= -.26171654  X3=  .79470694  Y=  4.3311501
X1=  9.4384476  X2=  9.4922712  X3=  6.5470797  Y=  6.875478
X1=  12.480887  X2=  3.6011219  X3= -9.5226097  Y= -2.8863931
X1= -1.2235525  X2= -11.968052  X3= -13.855501  Y= -6.4261985
X1=  1.766735   X2=  1.8474868  X3= -9.7351396  Y= -1.1970907
X1=  2.4248537  X2= -5.3025303  X3= -18.801529  Y= -9.2069399
X1= -17.75145   X2= -6.6832084  X3= -6.9518095  Y= -4.3757081
X1= -31.020753  X2= -10.492779  X3=  6.569823   Y= -10.944034
X1= -18.135506  X2= -5.9660229  X3= -.31270385  Y= -7.3421222
X1=  1.2875384  X2= -11.926949  X3= -24.23545   Y= -12.069287
X1= -12.131298  X2=  1.3302064  X3=  3.358078   Y=  3.8293123
```

X1=	-12.350362	X2=	-3.325575	X3=	5.2750468	Y=	-.01010358
X1=	42.440817	X2=	7.6883385	X3=	12.706037	Y=	10.678547
X1=	6.3172221	X2=	-11.963981	X3=	-17.198094	Y=	-6.9640881
X1=	-4.7283822	X2=	-9.2630132	X3=	4.8495519	Y=	-3.8294703
X1=	-22.863425	X2=	-22.362831	X3=	-35.496489	Y=	-19.458119
X1=	-11.637117	X2=	1.8595853	X3=	5.4396546	Y=	-1.638639
X1=	-2.0171478	X2=	8.3916357	X3=	-3.5796231	Y=	4.6224147
X1=	21.706068	X2=	10.541081	X3=	9.1486532	Y=	10.272042
X1=	-7.7373	X2=	4.4997907	X3=	-4.1612005	Y=	1.1364156
X1=	7.0474563	X2=	8.6840466	X3=	.56296587	Y=	9.4522399
X1=	-24.241482	X2=	-.75478852	X3=	-11.375631	Y=	-3.4443295
X1=	-2.7841389	X2=	13.147792	X3=	30.606506	Y=	10.802426
X1=	-37.231195	X2=	8.3945531	X3=	-29.995469	Y=	-2.3432821
X1=	-3.1258619	X2=	11.347094	X3=	21.036687	Y=	5.7465226
X1=	-18.924691	X2=	-12.438569	X3=	-7.0393765	Y=	-8.9706987
X1=	-12.193689	X2=	-16.371207	X3=	2.8125739	Y=	-15.50912
X1=	-1.3563627	X2=	-3.8587019	X3=	-7.3796207	Y=	-5.2416748
X1=	14.18179	X2=	-.88867188	X3=	3.4610951	Y=	6.9834536
X1=	8.7933996	X2=	-11.099741	X3=	-25.776783	Y=	-15.466592
X1=	.08275777	X2=	-13.369293	X3=	-15.47818	Y=	-7.0074427
X1=	34.707518	X2=	22.12854	X3=	9.3559492	Y=	24.886591
X1=	14.196219	X2=	-.85469306	X3=	14.6759	Y=	3.3188897
X1=	-1.6819164	X2=	-1.7652914	X3=	-11.067027	Y=	-5.8248091
X1=	14.916297	X2=	13.885227	X3=	26.036229	Y=	13.460225
X1=	-32.877031	X2=	-7.2486613	X3=	-10.769261	Y=	-9.4951361
X1=	3.2350281	X2=	10.196771	X3=	-1.7028976	Y=	-1.3868374
X1=	-8.0227342	X2=	4.0506086	X3=	9.1016978	Y=	5.5884296
X1=	16.025306	X2=	-.7600385	X3=	26.230796	Y=	5.8736545
X1=	10.921787	X2=	12.130628	X3=	17.06534	Y=	3.5516918
X1=	-32.836519	X2=	-5.4397753	X3=	2.7025932	Y=	-2.9226512
X1=	-9.810739	X2=	10.767364	X3=	9.6300578	Y=	7.415818
X1=	11.695045	X2=	3.0919629	X3=	-10.332803	Y=	-.04776299
X1=	-37.778847	X2=	-13.144732	X3=	13.043953	Y=	-11.311852
X1=	12.005492	X2=	14.679686	X3=	6.0630053	Y=	11.205198
X1=	7.099123	X2=	10.141955	X3=	27.839253	Y=	7.3014635
X1=	6.9298998	X2=	.13694972	X3=	-4.4951558	Y=	2.5154191
X1=	2.9473421	X2=	6.35261	X3=	-17.883204	Y=	.04902482
X1=	7.5547767	X2=	-2.3233557	X3=	5.3063303	Y=	-.63830435
X1=	-18.777341	X2=	-15.747531	X3=	-2.111311	Y=	-10.57938
X1=	12.638908	X2=	1.9193354	X3=	13.842595	Y=	2.8330016
X1=	-14.227514	X2=	-12.230794	X3=	15.193923	Y=	-12.409328
X1=	-14.288996	X2=	5.8357105	X3=	10.380412	Y=	4.6172285
X1=	19.748926	X2=	5.4792357	X3=	-1.5421438	Y=	8.2649463
X1=	-1.1844149	X2=	4.2815679	X3=	-2.4796557	Y=	6.0239297

Cues As Displayed On Screen

	X_1	X_2	X_3
1.	-19.6	-4.0	-6.0
2.	39.4	25.8	30.9
3.	-10.5	1.0	-.1
4.	33.7	.2	2.5
5.	-6.0	-5.5	-3.8
6.	26.8	4.1	-.3
7.	.8	-3.3	-9.1
8.	-4.8	-8.9	-21.6
9.	27.4	10.3	-3.7
10.	7.2	-12.8	-26.8
11.	10.0	0.0	-7.7
12.	19.7	3.4	9.0
13.	8.8	9.0	9.4
14.	-11.2	-5.0	-10.3
15.	.3	-6.1	-9.1
16.	-8.3	.3	-20.0
17.	-27.2	-7.6	12.1
18.	26.3	4.8	11.1
19.	.1	6.5	.2
20.	13.4	6.6	21.7

	X_1	X_2	X_3
21.	-4.3	-11.5	-20.9
22.	-24.6	-6.7	-12.8
23.	7.9	-3.4	-11.2
24.	-40.3	-17.7	-25.0
25.	23.1	-2.5	1.8
26.	-1.4	-.3	.8
27.	9.4	9.5	6.5
28.	12.5	3.6	-9.5
29.	-1.2	-12.0	-13.9
30.	1.8	1.8	-9.7
31.	2.4	-5.3	-18.8
32.	-17.8	-6.7	-7.0
33.	-31.0	-10.5	6.6
34.	-18.1	-6.0	-.3
35.	1.3	-11.9	-24.2
36.	-12.1	1.3	3.4
37.	-12.4	-3.3	5.3
38.	42.4	7.7	12.7
39.	6.3	-12.0	-17.2
40.	-4.7	-9.3	4.8

	X_1	X_2	X_3
41.	-22.9	-22.4	-35.5
42.	-11.3	1.9	5.4
43.	-2.0	8.4	-3.6
44.	21.7	10.5	9.1
45.	-7.7	4.5	-4.2
46.	7.0	8.7	.6
47.	-24.2	-.8	-11.4
48.	-2.8	13.1	30.6
49.	-37.2	8.4	-30.0
50.	-3.1	11.3	21.0
51.	-18.9	-12.4	-7.0
52.	-12.2	-16.4	2.8
53.	-1.4	-3.9	-7.4
54.	14.2	-.9	3.5
55.	8.8	-11.1	-25.8
56.	.1	-13.4	-15.5
57.	34.7	22.1	9.4
58.	14.2	-.9	14.7
59.	-1.7	-1.8	-11.1
60.	14.9	13.9	26.0

	X_1	X_2	X_3
61.	-32.9	-7.2	-10.8
62.	3.2	10.2	-1.7
63.	-8.0	4.1	9.1
64.	16.0	-.8	26.2
65.	10.9	12.1	17.1
66.	-37.8	-5.4	2.7
67.	-9.8	10.8	9.6
68.	11.7	3.1	-10.3
69.	-37.8	-13.1	13.0
70.	12.0	14.7	6.1
71.	7.1	10.1	27.8
72.	6.9	.1	-4.5
73.	2.9	6.4	-17.9
74.	7.6	-2.3	5.3
75.	-18.8	-15.7	-2.1
76.	12.6	1.9	13.8
77.	-14.2	-12.2	15.2
78.	-14.3	5.8	10.4
79.	19.7	5.5	-1.5
80.	-1.2	4.3	-2.5

Group II Evaluations - Good/Bad

Evaluation = Good if,

41.	$(-21.5 \leq Y_s \leq -17.5)$		61.	$(-11.5 \leq Y_s \leq -7.5)$	
42.	(-3.6	0.4)	62.	(-3.4	0.6)
43.	(2.6	6.6)	63.	(3.6	7.6)
44.	(8.3	12.3)	64.	(3.8	7.8)
45.	(-.9	3.1)	65.	(1.6	5.6)
46.	(7.5	11.5)	66.	(-4.9	-.9)
47.	(-5.4	-1.4)	67.	(5.4	9.4)
48.	(8.8	12.8)	68.	(-2.0	2.0)
49.	(-4.3	-0.3)	69.	(-13.3	-9.3)
50.	(3.7	7.7)	70.	(9.2	13.2)
51.	(-11.0	-7.0)	71.	(5.3	9.3)
52.	(-17.5	-13.5)	72.	(0.5	4.5)
53.	(-7.2	-3.2)	73.	(-2.0	2.0)
54.	(5.0	9.0)	74.	(-2.6	1.4)
55.	(-17.5	-13.5)	75.	(-12.6	-8.6)
56.	(-9.0	-5.0)	76.	(0.8	4.8)
57.	(22.9	26.9)	77.	(-14.4	-10.4)
58.	(1.3	5.3)	78.	(2.6	6.6)
59.	(-7.8	-3.8)	79.	(6.3	10.3)
60.	(11.5	15.5)	80.	(4.0	8.0)

Bad otherwise

Group III Evaluations - Good/Fair/Bad

	Evaluation = Good if;		Fair if;	
41.	$(-20.5 \leq Y_s \leq$	$-18.5)$	$(-22.0 \leq Y_s \leq$	$-17.0)$
42.	(-2.6	-0.6)	(-4.1	0.0)
43.	(3.6	5.6)	(2.1	7.1)
44.	(9.3	11.3)	(7.8	12.8)
45.	(0.1	2.1)	(-1.4	3.6)
46.	(8.5	10.5)	(7.0	12.0)
47.	(-4.4	-2.4)	(-5.9	-.9)
48.	(9.8	11.8)	(8.3	13.3)
49.	(-3.3	-1.3)	(-4.8	0.2)
50.	(4.7	6.7)	(3.2	8.2)
51.	(-10.0	-8.0)	(-11.5	-6.5)
52.	(-16.5	-14.5)	(-18.0	-13.0)
53.	(-6.2	-4.2)	(-7.7	-2.7)
54.	(6.0	8.0)	(4.5	9.5)
55.	(-16.5	-14.5)	(-18.0	-13.0)
56.	(-8.0	-6.0)	(-9.5	-4.5)
57.	(23.9	25.9)	(22.4	27.4)
58.	(2.3	4.3)	(0.8	5.8)
59.	(-6.8	-4.8)	(-8.3	-3.3)
60.	(12.5	14.5)	(11.0	16.0)
61.	(-10.5	-8.5)	(-12.0	-7.0)
62.	(-2.4	-0.4)	(-3.9	1.1)
63.	(4.6	6.6)	(3.1	8.1)
64.	(4.8	6.8)	(3.3	8.3)
65.	(2.6	4.6)	(1.1	6.1)
66.	(-3.9	-1.9)	(-5.4	-0.4)
67.	(6.4	8.4)	(4.9	9.9)
68.	(-1.0	1.0)	(-2.5	2.5)
69.	(-12.3	-10.3)	(-13.8	-8.8)
70.	(10.2	12.2)	(8.7	13.7)
71.	(6.3	8.3)	(4.8	9.8)
72.	(1.5	3.5)	(0.0	5.0)
73.	(-1.0	1.0)	(-2.5	2.5)
74.	(-1.6	0.4)	(-3.1	1.9)
75.	(-11.6	-9.6)	(-13.1	-8.1)
76.	(1.8	3.8)	(0.3	5.3)
77.	(-13.4	-11.4)	(-14.9	-9.9)
78.	(3.6	5.6)	(2.1	7.1)
79.	(7.3	9.3)	(5.8	10.8)
80.	(5.0	7.0)	(3.5	8.5)

Bad otherwise

Group IV Evaluations - Very Good/Good/Fair/Bad/Very Bad

41.	$(-20 \leq Y_s \leq -19)$	$(-21 \leq Y_s \leq -18)$	$(-22 \leq Y_s \leq -17)$	$(-23.5 \leq Y_s \leq -15.5)$
42.	(-2.1 -1.1)	(-3.1 -0.1)	(-4.1 0.9)	(-5.6 2.4)
43.	(4.1 5.1)	(3.1 6.1)	(2.1 7.1)	(0.6 8.6)
44.	(9.8 10.8)	(8.8 11.8)	(7.8 12.8)	(6.3 14.3)
45.	(0.6 1.6)	(-0.4 2.6)	(-1.4 3.6)	(-2.9 5.1)
46.	(9 10)	(8 11)	(7 12)	(5.5 13.5)
47.	(-3.9 -2.9)	(-4.9 -1.9)	(-5.9 -0.9)	(-7.4 0.6)
48.	(10.3 11.3)	(9.3 12.3)	(8.3 13.3)	(6.8 14.8)
49.	(-2.8 -1.8)	(-3.8 -0.8)	(-4.8 0.2)	(-6.3 1.7)
50.	(5.2 6.2)	(4.2 7.2)	(3.2 8.2)	(1.7 9.7)
51.	(-9.5 -8.5)	(-10.5 -7.5)	(-11.5 -6.5)	(-13 -5)
52.	(-16 -15)	(-17 -14)	(-18 -13)	(-19.5 -11.5)
53.	(-5.7 -4.7)	(-6.7 -3.7)	(-7.7 -2.7)	(-9.2 -1.2)
54.	(6.5 7.5)	(5.5 8.5)	(4.5 9.5)	(3 11)
55.	(-16 -15)	(-17 -14)	(-18 -13)	(-19.5 -11.5)
56.	(-7.5 -6.5)	(-8.5 -5.5)	(-9.5 -4.5)	(-11 -3)
57.	(24.4 25.4)	(23.4 26.4)	(22.4 27.4)	(20.9 28.9)
58.	(2.8 3.8)	(1.8 4.8)	(0.8 5.8)	(-0.7 7.3)
59.	(-6.3 -5.3)	(-7.3 -4.3)	(-8.3 -3.3)	(-9.8 -1.8)
60.	(13 14)	(12 15)	(11 16)	(9.5 17.5)
61.	(-10 -9)	(-11 -8)	(-12 -7)	(-13.5 -5.5)
62.	(-1.9 -0.9)	(-2.9 0.1)	(-3.9 1.1)	(-5.4 2.6)
63.	(5.1 6.1)	(4.1 7.1)	(3.1 8.1)	(1.6 9.6)
64.	(5.3 6.3)	(4.3 7.3)	(3.3 8.3)	(1.8 9.8)
65.	(3.1 4.1)	(2.1 5.1)	(1.1 6.1)	(-0.4 7.6)
66.	(-3.4 -2.4)	(-4.4 -1.4)	(-5.4 -0.4)	(-6.9 1.1)
67.	(6.9 7.9)	(5.9 8.9)	(4.9 9.9)	(3.4 11.4)
68.	(-0.5 0.5)	(-1.5 1.5)	(-2.5 2.5)	(-4 4)
69.	(-11.8 -10.8)	(-12.8 -9.8)	(-13.8 -8.8)	(-15.3 -7.3)
70.	(10.7 11.7)	(9.7 12.7)	(8.7 13.7)	(7.2 15.2)
71.	(6.8 7.8)	(5.8 8.8)	(4.8 9.8)	(3.3 11.3)
72.	(2 3)	(1 4)	(0 5)	(-1.5 6.5)
73.	(-0.5 0.5)	(-1.5 1.5)	(-2.5 2.5)	(-4 4)
74.	(-1.1 -0.1)	(-2.1 0.9)	(-3.1 1.9)	(-4.6 3.4)
75.	(-11.1 -10.1)	(-12.1 -9.1)	(-13.1 -8.1)	(-14.6 -6.6)
76.	(2.3 3.3)	(1.3 4.3)	(0.3 5.3)	(-1.2 6.8)
77.	(-12.9 -11.9)	(-13.9 -10.9)	(-14.9 -9.9)	(-16.4 -8.4)
78.	(4.1 5.1)	(3.1 6.1)	(2.1 7.1)	(0.6 8.6)
79.	(7.8 8.8)	(6.8 9.8)	(5.8 10.8)	(4.3 12.3)
80.	(5.5 6.5)	(4.5 7.5)	(3.5 8.5)	(2 10)

Instructions Read to Group I

You are about to participate in an experiment to measure learning. You are asked to make decisions based on the information provided you on the three sheets you have before you. These numbers represent daily stock market predictions by three local "experts". Assume that each numbered trial is a new day and the predictions listed for each trial were made the previous day.

I will display each day's predictions on the board as I read them out loud. As each day's prediction is read, please record your prediction for that day on the answer sheet. Make your predictions based solely on the information before you. When 40 trials have been completed please put down your pencils and a 5 minute break will be given. I repeat, please do not make any predictions beyond #40!

BREAK

You will continue to make predictions as before, however, after each prediction I will tell you the actual stock market performance for that day. This may now be recorded next to your prediction in the space provided. You now have an additional piece of information with which to make your next prediction. This procedure will be followed until 80 trials have been completed.

Instructions Read to Group II

You are about to participate in an experiment to measure learning. You are asked to make decisions based on the information provided you on the three sheets you have before you. These numbers represent daily stock market predictions by three local "experts". Assume that each numbered trial is a new day and the predictions listed for each trial were made the previous day.

I will display each day's predictions on the board as I read them out loud. As each day's prediction is read, please record your prediction for that day on the answer sheet. Make your predictions based solely on the information before you. When 40 trials have been completed please put down your pencils and a 5 minute break will be given. I repeat, please do not make any predictions beyond #40!

BREAK

Please look at responses 41-80. You will notice that next to each set of numbers there is a blank space for your prediction and then the letters G and B. These letters stand for GOOD and BAD. Starting with #41 please record your prediction as before, but this time I will come around and evaluate each prediction by circling the appropriate letter before you go on to the next one. The evaluation is based on how close your prediction is to the actual market performance. You now have an additional piece of information with which to make your next decision.

Instructions Read to Group III

You are about to participate in an experiment to measure learning. You are asked to make decisions based on the information provided you on the three sheets you have before you. These numbers represent daily stock market predictions by three local "experts". Assume that each numbered trial is a new day and the predictions listed for each trial were made the previous day.

I will display each day's predictions on the board as I read them out loud. As each day's prediction is read, please record your prediction for that day on the answer sheet. Make your predictions based solely on the information before you. When 40 trials have been completed please put down your pencils and a 5 minute break will be given. I repeat, please do not make any predictions beyond #40!

BREAK

Please look at responses 41-80. You will notice that next to each set of numbers there is a blank space for your prediction and then the letters G, F and B. These letters stand for GOOD, FAIR and BAD. Starting with #41 please record your prediction as before, but this time I will come around and evaluate each prediction by circling the appropriate letter before you go on to the next one. This evaluation is based on how close your prediction is to the actual market performance. You now have an additional piece of information with which to make your next prediction.

Instructions Read to Group IV

You are about to participate in an experiment to measure learning. You are asked to make decisions based on the information provided you on the three sheets you have before you. These numbers represent daily stock market predictions by three local "experts". Assume that each numbered trial is a new day and the predictions listed for each trial were made the previous day.

I will display each day's predictions on the board as I read them out loud. As each day's prediction is read, please record your prediction for that day on the answer sheet. Make your predictions based solely on the information before you. When 40 trials have been completed please put down your pencils and a 5 minute break will be given. I repeat, please do not make any predictions beyond #40!

BREAK

Please look at responses 41-80. You will notice that next to each set of numbers is a blank space for your prediction and then the letters VG, G, F, B, VB. These letters stand for VERY GOOD, GOOD, FAIR, BAD, and VERY BAD. Starting with #41 please record your predictions as before, but this time I will come around and evaluate each prediction by circling the appropriate letter before you go on to the next one. This evaluation is based on how close your prediction was to the actual market performance. You now have an additional piece of information with which to make your next prediction.

Correlations and Z-transformations for
the outcome feedback group

	trials 1-40		trials 41-80	
R_e	.938		.934	
r_{a1}	.861		.865	
r_{a2}	.861		.906	
r_{a3}	.890	$\bar{r}_a$ = .858	.920	$\bar{r}_a$ = .891
r_{a4}	.854		.869	
r_{a5}	.814		.887	
R_{s1}	.958		.956	
R_{s2}	.942		.974	
R_{s3}	.957	$\bar{R}_s$ = .935	.979	$\bar{R}_s$ = .959
R_{s4}	.924		.937	
R_{s5}	.861		.921	
G_1	.976		.953	
G_2	.947		.979	
G_3	.997	$\bar{G}$ = .987	.997	$\bar{G}$ = .994
G_4	.992		.997	
G_5	.999		.999	
$Z_{r_{a1}}$	1.2972		1.3129	
$Z_{r_{a2}}$	1.2972		1.5047	
$Z_{r_{a3}}$	1.4219		1.5890	
$Z_{r_{a4}}$	1.2707		1.3290	
$Z_{r_{a5}}$	1.1388		1.4077	

Correlations and Z-transformations for
the 2c1 evaluative feedback group

	trials 1-40		trials 41-80	
R_e	.938		.934	
r_{a6}	.929		.854	
r_{a7}	.921		.795	
r_{a8}	.895	$\bar{r}_a$ = .893	.880	$\bar{r}_a$ = .833
r_{a9}	.784		.755	
r_{a10}	.895		.855	
R_{s6}	.980		.926	
R_{s7}	.988		.905	
R_{s8}	.915	$\bar{R}_s$ = .958	.966	$\bar{R}_s$ = .925
R_{s9}	.871		.902	
R_{s10}	.952		.898	
G_6	.994		.983	
G_7	.975		.940	
G_8	.990	$\bar{G}$ = .989	.993	$\bar{G}$ = .979
G_9	.967		.936	
G_{10}	.997		.991	
$Z_{r_{a6}}$	1.6510		1.2707	
$Z_{r_{a7}}$	1.5956		1.0849	
$Z_{r_{a8}}$	1.4465		1.3758	
$Z_{r_{a9}}$	1.0557		.9845	
$Z_{r_{a10}}$	1.4465		1.2745	

Correlations and Z-transforms for the
3cl evaluative feedback group

	trials 1-40		trials 41-80	
R_e	.938		.934	
r_{a11}	.874		.771	
r_{a12}	.879		.838	
r_{a13}	.911	$\bar{r}_a = .908$	.728	$\bar{r}_a = .786$
r_{a14}	.928		.766	
r_{a15}	.931		.811	
R_{s11}	.937		.952	
R_{s12}	.988		.974	
R_{s13}	.941	$\bar{R}_s = .965$	.921	$\bar{R}_s = .939$
R_{s14}	.948		.917	
R_{s15}	.978		.896	
G_{11}	.972		.931	
G_{12}	.956		.906	
G_{13}	.971	$\bar{G} = .974$	.913	$\bar{G} = .933$
G_{14}	.973		.921	
G_{15}	.988		.969	
$Z_{r_{a11}}$	1.3498		1.0228	
$Z_{r_{a12}}$	1.3713		1.2144	
$Z_{r_{a13}}$	1.5334		.9245	
$Z_{r_{a14}}$	1.6438		1.0106	
$Z_{r_{a15}}$	1.6658		1.1299	

Correlations and Z-transformations for the
5cl evaluative feedback group

	trials 1-40		trials 41-80	
R_e	.938		.934	
r_{a16}	.844		.522	
r_{a17}	.811		.640	
r_{a18}	.916	$\bar{r}_a$ = .869	.841	$\bar{r}_a$ = .708
r_{a19}	.822		.794	
r_{a20}	.870		.647	
R_{s16}	.960		.546	
R_{s17}	.836		.713	
R_{s18}	.971	$\bar{R}_s$ = .943	.976	$\bar{R}_s$ = .851
R_{s19}	.966		.929	
R_{s20}	.914		.735	
G_{16}	.986		.961	
G_{17}	.988		.981	
G_{18}	.973	$\bar{G}$ = .980	.936	$\bar{G}$ = .958
G_{19}	.970		.943	
G_{20}	.975		.953	
$Z_{r_{a16}}$	1.2349		.5791	
$Z_{r_{a17}}$	1.1299		.7582	
$Z_{r_{a18}}$	1.5636		1.2246	
$Z_{r_{a19}}$	1.3847		1.0822	
$Z_{r_{a20}}$	1.3331		.7701	

Program Used to Compute Z-transformations

```
RØSE-DØNALD*ZETE.MAIN
          DIMENSIØN R(50),N(50),Z(50)
          READ(5,1) K
          READ(5,1) (R(I),I=1,K)
          READ(5,1) (N(I),I=1,K)
        1 FØRMAT( )
          ZBAR=0
          ZBAR2=0
          DØ 10 I=1,K
          Z(I)=.5*ALØG((1+R(I))/(1-R(I)))
          ZBAR=ZBAR+(N(I)-3)*Z(I)
       10 ZBAR2=ZBAR2+N(I)-3
          ZBA=ZBAR/ZBAR2
          RBAR=(EXP(2.*ZBA)-1.)/(EXP(2.*ZBA)+1.)
          WRITE(6,15) ZBA
       15 FØRMAT(1X,'ZBAR IS',F10.6)
          DØ 25 I=1,K
       25 WRITE(6,20) Z(I)
       20 FØRMAT(1X,F10.6)
          WRITE(6,18) RBAR
       18 FØRMAT(1X,'RBAR IS',F10.6)
          END
END PRT
```

Analysis of Variance Across Classes of Feedback

POINT	VALUE	RESIDUAL
1	.13129000+01	-.11575998+00
2	.15047000+01	.76040015-01
3	.15890000+01	.16034001+00
4	.13290000+01	-.99659994-01
5	.14077000+01	-.20959988-01
6	.12707000+01	.72620019-01
7	.10849000+01	-.11317997+00
8	.13758000+01	.17772003+00
9	.98450000+00	-.21357998+00
10	.12745000+01	.76420024-01
11	.10228000+01	-.37639990-01
12	.12144000+01	.15396000+00
13	.92450000+00	-.13593999+00
14	.10106000+01	-.49839988-01
15	.11299000+01	.69460005-01
16	.57910000+00	-.30373999+00
17	.75820000+00	-.12463999+00
18	.12246000+01	.34176002+00
19	.10822000+01	.19936003+00
20	.77010000+00	-.11273999+00

TREAT	REPL.	EST OF EFFECTS
1	5	.28615506+00
2	5	.55575043-01
3	5	-.82064942-01
4	5	-.25966495+00

ANOVA Table

	SUM OF SQUARES	D.F.	MEAN SQUARE
TREATMENT	.79566921+00	3	.26522307+00
ERROR	.48434360+00	16	.30271475-01
TOTAL	.12800127+01	19	

F VALUE .87614849+01 PROBABILITY F IS EXCEEDED .00114255

Test of Means ($\bar{r}_a/R_e$) Between the Last
Two Blocks of 20 Trials

H_o: $\mu_1 - \mu_2 = 0$ μ_1 = Mean percent achievement for trials 41-60

H_1: $\mu_1 - \mu_2 \neq 0$ μ_2 = Mean percent achievement for trials 61-80

	Q/C FB	2cl FB	3cl FB	5cl FB
$\bar{X}_1$	.940	.880	.840	.850
$\bar{X}_2$	.990	.920	.860	.740
SS_1	.0126	.0070	.0042	.0708
SS_2	.0815	.0263	.0159	.5642
S_p	.0118	.0042	.0025	.0794
t_o	-.713 (N.S.)	-.781 (N.S.)	-.569 (N.S.)	.943 (N.S.)

where

$$t_o = \frac{\bar{X}_1 - \bar{X}_2}{S_p\sqrt{\frac{1}{n} + \frac{1}{n}}} \sim t_{2n-2}$$

Duncan's Multiple Range Test
for Major Hypothesis

$s_e = .0814$ $n_2 = 16$ $\alpha = .05$

	2	3	4
Sign. Range	3.00	3.15	3.23
Least Sign. Range	.2442	.2564	.2710

$\bar{Z}_I = 1.4287$ I vs IV; $.5459 > .2710$
I vs III; $.3683 > .2564$

$Z_{II} = 1.1980$ I vs II; $.2307 < .2442$

$\bar{Z}_{III} = 1.0604$ II vs IV; $.3152 > .2564$
II vs III; $.1376 < .2442$

$\bar{Z}_{IV} = .8828$ III vs IV; $.1776 < .3360$

BIBLIOGRAPHY

1. Brunswik, Egon, Perception and the Representative Design of Experiments, Berkeley: University of California Press, 1956.

2. Connolly, Terry, "Making Sense of the Problem:" The Inferential Base in Lens Model Research, Proposed Article.

3. Dawes, R. M., A Case Study of Graduate Admissions: Application of Three Principles of Human Decision Making, American Psychologist, 1971, 26, 180-188.

4. Edwards, W., Bayesian and Regression Models of Human Information Processing - a Myopic Perspective, Organizational Behavior and Human Performance, 1971, 6, 639-648.

5. Enborn, Hillel J., Use of Non-Linear, Non Compensating Models as a Function of Task and Amount of Information, Organization Behavior and Human Performance, 6, 1-27, (1971).

6. Goldberg, L. R., Simple Models or Simple Processes? Some Research on Clinical Judgments; American Psychologist, 23, 1968, 483-496.

7. Goldberg, L. R., Man Versus Model of Man: A Rationale Plus Some Evidence for a Method of Improving on Clinical Inferences, Psychological Bulletin, Vol. 73, 6, 422-432, 1970.

8. Graybill, F. A., An Introduction to Linear Statistical Models, Vol. 1, McGraw Hill (1961).

9. Hammond, K. R., and Boyle, P. J., Quasi-Rationality, Quarrels and New Conceptions of Feedback, Program on Cognitive Processes Report No. 130, Institute of Behavioral Science, University of Colorado, 1970.

10. Hammond, K. R., Hursch, C. J., and Todd, E. J., Analyzing the Components of Clinical Inference, Psychological Review, 1964, Vol. 71, No. 6, 438-456.

11. Hammond, K. R., and Summers, David A., Cognitive Dependence on Linear and Non-linear Cues, Psychological Review, 72, No. 3, 215-224 (1965).

12. Hammond, K. R., and Summers, David A., Cognitive Control, Psychological Review, Vol. 79, No. 1, 58-67 (1972).

13. Hayes, J. R., Human Data Processing Limits in Decision Making, In E. Bennett (Ed.), Information System Science and Engineering, Proceedings of the First Congress on Information Systems Sciences, New York, McGraw-Hill, 1964.

14. Hines, W. W., and Montgomery, D. C., Probability and Statistics In Engineering and Management Science, The Ronald Press Company, New York (1972).

15. Hoffman, Paul J., The Paramorphic Representation of Clinical Judgment, Psychological Bulletin, Vol. 57, No. 2, 1960.

16. Hoffman, Paul J., Cue Consistency and Configurality in Human Judgment, in Kleinmuntz, Benjamin (Ed), John Wiley and Sons, Inc., New York, 1968.

17. Hoffman, P. J., and Blanchard, W. A., A Study of the Effects of Varying Amounts of Predictor Information on Judgment, Oregon Research Institute Research Bulletin, 1961.

18. Hursch, C., Hammond, K. R., and Hursch, J. L., Some Methodological Considerations in Multiple Cue Probability Studies, Psychological Review, 1964, 71, 42-60.

19. Kleinmuntz, Benjamin, The processing of Clinical Information by Man and Machine, in Kleinmuntz, B (Ed), John Wiley and Sons, Inc., New York, 1968.

20. Martin and Gettys, Feedback and Response Mode in Performing a Bayesian Decision Task, Journal of Applied Psychology, 1969, 53, 413-418.

21. McNemar, Quinn, Psychological Statistics, John Wiley and Sons, Inc., New York, 1949.

22. Miklausich, V. M., The Impact of Error-Prone Feedback on Learning in a Multiple-Cue Dicision Problem, M. S. Thesis, Georgia Institute of Technology, Atlanta, Georgia, 1973.

23. Oskamp, S., Over Confidence in Case-Study Judgments, Journal of Consulting Psychology, 1965, 29, 261-265.

24. Peterson, C. R., Ducharme, W. M., and Edwards, W., Sampling Distributions and Probability Revisions, Journal of Experimental Psychology, 1968, 76, 236-243.

25. Phillips, L. D., and Edwards, W., Conservatism in a Simple Probability Inference Task, Journal of Experimental Psychology, 1966, 72, 346-357.

26. Slovic, P., Cue Consistency and Cue Utilization in Judgment, American Journal of Psychology, 1966, 79.

27. Slovic, P., Analyzing the Expert Judge: A Descriptive Study of a Stockbroker's Decision Processes, Journal of Applied Psychology, 1969, 53.

28. Slovic, S., Fleissner, D., and Bauman, W. S., Quantitative Analysis of Investment Decisions, Journal of Business, 1972.

29. Slovic, Paul and Lichtenstein, Sarah, Comparison of Bayesian and Regression Approaches to the Study of Information Process in Judgment, Organizational Behavior and Human Performance, 6, 649-744 (1971).

30. Todd, F. J., and Hammond, K. R., Differential Feedback in Two Multiple-Cue Probability Learning Tasks, Behavioral Science, 1965, 10, 429-435.

31. Tucker, L. R., A Suggested Alternative Formulation in the Development of Hursch, Hammond, and Hursch, and by Hammond, Hursch, and Todd, Psychological Review, 1964, 71, 528-530.

32. Wheeter, G., and Beach, L. R., Subjective Sampling Distributions and Conservatism, Organizational Behavior and Human Performance, 1968, 3, 36-46.

33. Wiggins, N., and Hoffman, P. J., Three Models of Clinical Judgment, Journal of Abnormal Psychology, 1968, 73, 70-77.

34. Yntema, D. B., and Torgerson, Man-Computer Cooperation in Decisions Requiring Common Sense, IRE Transactions on Human Factors in Electronics, HFE-2, pp. 20-26, (1961).