

GEORGIA INSTITUTE OF TECHNOLOGY LIBRARY

Regulations for the Use of Theses

Unpublished theses submitted for the Master's and Doctor's degrees and deposited in the Georgia Institute of Technology Library are open for inspection and consultation, but must be used with due regard for the rights of the authors. Passages may be copied only with permission of the authors, and proper credit must be given in subsequent written or published work. Extensive copying or publication of the thesis in whole or in part requires the consent of the Dean of the Graduate Division of the Georgia Institute of Technology.

This thesis by JEFFERY BRYAN FREY has been used by the following persons, whose signatures attest their acceptance of the above restrictions.

A library which borrows this thesis for use by its patrons is expected to secure the signature of each user.

NAME AND ADDRESS OF USER	BORROWING LIBRARY	DATE
--------------------------	-------------------	------

INTERMEDIATE CONSTRUCTS IN THE  
BRUNSWIK LENS MODEL

A THESIS

Presented to

The Faculty of the Division of Graduate  
Studies and Research

By

Jeffery Bryan Frey

In Partial Fulfillment  
of the Requirements for the Degree  
Master of Science in Operations Research

Georgia Institute of Technology

December, 1973

INTERMEDIATE CONSTRUCTS IN THE  
BRUNSWIK LENS MODEL

Approved: \_\_\_\_\_

\_\_\_\_\_  
Dr. Terry Connolly, Chairman

\_\_\_\_\_  
Dr. Douglas C. Montgomery

\_\_\_\_\_  
Dr. Thomas L. Sadosky

Date approved by Chairman: \_\_\_\_\_

J  
9 NOV 73

## ACKNOWLEDGMENTS

The author wishes to acknowledge the efforts and instruction of his principal advisor, Dr. Terry Connolly, and of his reading committee, Dr. Douglas C. Montgomery, and Dr. Thomas L. Sadosky.

## TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS . . . . .	ii
LIST OF TABLES . . . . .	iv
LIST OF ILLUSTRATIONS . . . . .	v
SUMMARY . . . . .	vi
Chapter	
1. INTRODUCTION . . . . .	1
Shadows on the Wall of the Cave	
Historical Perspective	
Brunswik's Lens Model	
Definitions	
Empirical Studies	
The Lens Model Re-examined	
Research Question	
Hypotheses	
II. EXPERIMENTAL DESIGN . . . . .	17
Background	
Verbal Contexts	
Test Generation	
Equipment and Materials	
Subjects	
Procedure	
III. RESULTS AND DISCUSSION OF RESULTS . . . . .	22
Primary Analysis	
Secondary Analysis	
IV. CONCLUSIONS AND RECOMMENDATIONS . . . . .	33
Review of Results	
Relationship to Real World	
Extension and Expansion	
APPENDIX . . . . .	36
BIBLIOGRAPHY . . . . .	37

## LIST OF TABLES

Table	Page
1. Results and Analysis of Phase A . . . . .	24
2. Results and Analysis of Phase B . . . . .	27
3. Percent Achievement, Phase A . . . . .	31
4. Percent Achievement, Phase B . . . . .	31

## LIST OF ILLUSTRATIONS

Figure	Page
1. Brunswik's Lens Model . . . . .	3
2. Problem Structure . . . . .	18
3. Test Generation Flow Diagram, Phase A . . . . .	18
4. Display of Results, Phase A . . . . .	23
5. Display of Results, Phase B . . . . .	26
6. Percent Achievement Using $R_e$ , Phase A . . . . .	29
7. Percent Achievement Using $R_e$ , Phase B . . . . .	29
8. Percent Achievement Using $R_e^*$ , Phase A . . . . .	30
9. Percent Achievement Using $R_e^*$ , Phase B . . . . .	30

## SUMMARY

Information presented to a decision maker for interpretation may be in at least two forms: cues - multiple indicators of the state of some underlying variable or construct; and components - representative separable dimensions of the entity about which a decision is to be made. Generally, the information is presented as a mixed set of these forms; this implies a requirement for a two-stage processing strategy, separating the estimation and evaluation phases. If the decision maker is able to apply an appropriate set of task-structuring constructs in an orderly manner, he should be able to reduce both his estimation error and the information-processing demands of the task.

This research investigated the effects on performance, in such tasks, of suggesting, and forcing, an appropriate conceptual structure. The central model was Brunswik's Lens Model, while the conceptual structure was suggested, and imposed, by means of a verbal problem context. The outcome of the study contributed to understanding the impact of conceptual structures on successful problem solving, particularly by its attempt to bridge the gap between the slow learning of laboratory experiments, and the relatively fast learning of the real world.



## CHAPTER I

### INTRODUCTION

#### Shadows on the Wall of the Cave

##### Historical Perspective

Egon Brunswik, in 1956, published a most significant theory concerning the means of studying how man does his mental work. Brunswik was the first to notice that task modelling was different from people modelling. Moreover, he devised a task modelling tool that linked tasks and people (Edwards, 1971). His original work has been modified and enhanced by Brunswik himself, and by a great many other people, most notably by Kenneth R. Hammond, working with various other researchers. Hammond, the director of the Research Program on Human Judgment and Social Interaction, Institute of Behavioral Science at the University of Colorado, is perhaps the principal advocate of Brunswik's work. Let us then examine this task modelling tool which allows one to separate out the environmental elements' contributions to a judge's achievement (Hammond, 1966).

##### Brunswik's Lens Model

The central postulate of Brunswik's work is that our knowledge of our environment is acquired in the face of uncertainty. For example, his original research on perception stressed that human attainment of a percept, such as the distance of an object, is based on a number of error-prone indicators (or cues) such as apparent size, brightness,

perspective information, and so on. Each cue is, on its own, unreliable to some extent, although each is related to the environmental state being perceived. Successful performance in such situations thus requires the individual to process a number of unreliable cues into a single (and more reliable) percept. Subsequent extensions of this underlying idea include clinical judgment situations such as Rorschach test interpretation (Hammond, 1955), estimates of student grade averages (Hammond, Hursch, and Todd, 1964), and use of geometric cues to predict a numerical criterion (Todd and Hammond, 1965).

The Brunswik approach to such situations treats simultaneously: (a) a set of relationships between a distal variable and a set of informational cues; and (b) a set of relationships between these cues and the subject's response. The model provides a framework in which the subject (or inferring organism) is placed in a structured relationship with the ecology, thus providing a basis for the study of complex human inferential situations (Yntema and Torgerson, 1961). This set of relationships is portrayed in Figure 1.

### Definitions

$Y_e$  is the established (true) value of the distal variable on the ecology side of the model. This is the variable which the subject will estimate. Note that it may be extremely difficult, or even impossible, to gain access to the value of  $Y_e$ .

$[x_i]$  is a set of cues, or multiple indicators of the underlying distal variable.

$Y_s$  is the subject's response, his estimate of the value of the distal variable.

Ecology

Subject

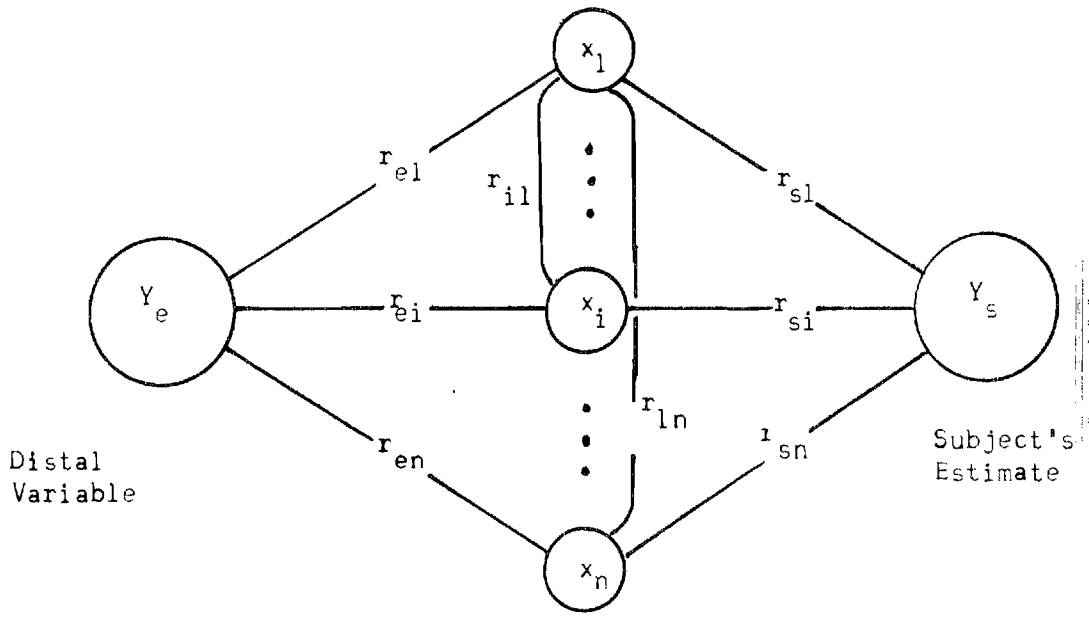


Figure 1. Brunswik's Lens Model.

$\hat{Y}_e$  is the best linear prediction of the distal variable based on the multiple indicators  $[x_i]$  of that variable.

$\hat{Y}_s$  is the best linear prediction of the subject's response, based on the cues.

$r_{ie}$  is the simple product-moment correlation between the value of the  $i^{\text{th}}$  cue and the true value of the distal variable. It is called the ecological validity of the  $i^{\text{th}}$  cue, since it is calculated for the ecology side of the model.

$r_{is}$ , the simple product-moment correlation between the value of the  $i^{\text{th}}$  cue and the subject's response, is called the subject's utilization coefficient for the  $i^{\text{th}}$  cue.

$r_a$ , the simple product-moment correlation between the distal variable and the subject's response, is referred to as the achievement index, since it provides a measure of the subject's success in his estimating task.

$R_e$ , the multiple correlation between the distal variable and its best linear prediction, measures the linear predictability of the distal variable from the cues. In laboratory tasks, it may be predetermined.

$R_s$  is the multiple correlation between the subject's response and its best linear prediction, on the subject side of the model. When limited to laboratory tasks, Hammond and Summers (1972) have defined  $R_s$  to be the cognitive control that the subject exerts over the use of his knowledge.

$G$  is the multiple correlation coefficient between the best linear prediction of the distal variable, and the best linear prediction of the

distal variable, and the best linear prediction of the response variable. It is called the matching index and measures the knowledge that the subject has obtained about the task properties (Hammond and Summers, 1972).

By definition, the task uncertainty measurement,  $R_e$ , sets the limit to achievement of any simple linear model, reached only if the subject has exact knowledge of the task structure and perfect cognitive control. As Hammond and Summers (1972) state:

$R_s$  is statistically independent of  $G$ . Such independence is critical, for it means that even should  $G$  reach unity (indicating perfect knowledge), if  $R_s$  were less than unity (indicating imperfect control), performance would be less than the limit of achievement ( $R_e$ ) would permit. Conversely,  $R_s$  might equal 1.00, thus indicating that the perfectly controlled cognitive system was not appropriate to the task system, thus preventing achievement  $r_a$  from reaching its upper limit  $R_e$ .

Two subjects, therefore, might have identical achievement indexes for different reasons; one because of perfect knowledge ( $G = 1.00$ ) but imperfect control ( $R_s < 1.00$ ), and the other because of perfect control ( $R_s = 1.00$ ), but imperfect knowledge ( $G < 1.00$ ). Variations between these extremes could also occur, of course.

$C$  is the simple product moment correlation between the non-linear residuals of  $Y_e$  ( $\delta_e$ ) and the non-linear residuals of  $Y_s$  ( $\delta_s$ ).

L. C. Tucker's 1964 paper presented an elegant alternative formulation to the important work of Hursch, Hammond, and Hursch (1964), and Hammond, Hursch, and Todd (1964). This formulation linked together the various elements defined above in a clear, concise manner. Tucker assumed that all variables were standardized, and that  $Y_e$  had a normal underlying distribution. He defined  $\delta_e$  and  $\delta_s$  as the residual errors of the standardized linear predictions of  $Y_e$  and  $Y_s$ . Thus,

$$\begin{aligned} Y_e &= \hat{Y}_e + \delta_e, \quad \text{and} \\ Y_s &= \hat{Y}_s + \delta_s. \end{aligned} \quad (1)$$

From multiple linear regression theory,

$$\begin{aligned} \text{Var}(\hat{Y}_e) &= R_e^2, \quad \text{and} \\ \text{Var}(\hat{Y}_s) &= R_s^2, \end{aligned} \quad (2)$$

so that, from (1):

$$\begin{aligned} \text{Var}(\delta_e) &= 1 - R_e^2, \quad \text{and} \\ \text{Var}(\delta_s) &= 1 - R_s^2. \end{aligned} \quad (3)$$

Thus:

$$r_a = \text{Cov}(Y_e Y_s) = \text{Cov}(\hat{Y}_e \hat{Y}_s) + \text{Cov}(\delta_e \delta_s). \quad (4)$$

Substitution into equation (4) yields:

$$r_a = GR_e R_s + C \sqrt{(1 - R_e^2)(1 - R_s^2)}. \quad (5)$$

Observe that when any of the following occur: (a) cognitive control reaches one; (b) the linear predictability goes to one; or when (c) C reaches zero, the right hand portion of equation (5) vanishes, and Tucker's equation reduces to:

$$r_a = G R_e R_s.$$

Therefore, Tucker has shown that under any of these restrictions, the achievement correlation is a function of the knowledge the subject has of the task, the properties of the environment and the subject's response

system, and the extent to which the non-linear variance elements of the ecology side of the model are correlated with the non-linear variance elements of the subject side of the model (Slovic and Lichtenstein, 1971).

### Empirical Studies

The Brunswik model provides an excellent model for research, some of the results of which might be used to program a machine to aid a decision maker. After all, a machine is consistent, and does not suffer from fatigue or prejudice. Clearly, a machine programmed with a simple linear model will at least equal a man's performance in tasks for which such a model is optimal.

Man, however, has been thought of as a superior decision maker because of some expectation of non-linear or configural skills. D. B. Yntema and W. S. Torgerson (1961), suggested that consistent use of a simple linear model resulted in better judgments than when humans had free reign in arriving at a judgment, to include the ability to consider non-linearities in the situation (Slovic and Lichtenstein, 1971),

L. W. Dudycha and J. C. Naylor (1966) went even further. They observed that their subjects, working on complex-structured multiple cue inferential tasks, were assigning proper weights to the cues, but were causing inaccuracies due to their inconsistencies. They concluded that, once the judgment pattern had been established by the subjects, the subjects should allow the strategy to persist unaltered (Slovic and Lichtenstein, 1971). In fact, Dawes (1971, 1972) later argued that any linear model will, on the average, out-perform subjects. Other studies have been conducted by Yntema, Torgerson and Lee (1961) and Raiffa (1970).

With an observation that certainly resembles solid experimental design, Hoffman (1960) noticed that statistical difficulties present themselves when the search for configural information use is made after conducting the experiment. Green (in Slovic and Lichtenstein, 1971) agreed and suggested that the first step be to advance the hypothesis of non-linear behavior, and then search for support for it. Slovic and Lichtenstein, 1971, found that differences in subjects' combination strategies were a function of whether two cues were in conflict or not. When both the cues were congruent, the subjects used both. When they were not, the subjects discounted one or the other or both, and turned to other information. Slovic's work as well as experiments by Hoffman (1960) and Anderson and Jacobson (1968) show that the linear model may need a term to account for the level of incompatibility among the cues.

In the modern world, decisions have become more necessary, more frequent, and more important than ever before. Ancient man had little to do but keep himself and his family alive. Life was remarkably simple then, albeit brutal. Technology has changed that; individual men now effect decisions that affect not only themselves and their families, but hundreds or millions of others.

The aftermath of a bad decision is usually to complain that not enough accurate information was available at the time the decision had to be made. Often, however, the problem is not that the information was not available, but that it had to be sorted, interpreted and integrated with other items of information to provide a proper base for the decision (Slovic and Lichtenstein, 1971).



One must find some way of solving this problem, and judging by the way the world keeps moving along, human beings have become very proficient at it. However, when an attempt is made to find out just how this is done, the issue becomes exceedingly complex. Although people have the facility for solving real life problems, and are doing a good job at it, attempts to present laboratory problems under controlled conditions meet with little successful comprehension by the subjects (Hammond and Summers, 1972). For example, some of Brehmer's (1973) subjects labored over 360 trials of line estimation, only to have their performance rating equal little better than chance.

The poor performance in a laboratory setting alluded to above, is often explained away by saying that the subject did not fully understand what he was to do. One can, however, using the correlations and formulas described above, discover the individual's level of knowledge and of control, and show (as was discussed on page 5) that perfect knowledge cannot guarantee perfect achievement, when there is imperfect cognitive control Hammond and Summers (1972).

#### The Lens Model Re-examined

Thus far, this paper has presented the fundamental concepts of Brunswik's Lens Model, its statistical developments, and some illustrative research suggesting the power of the approach to understanding complex judgment and decision making skills. The remainder of this chapter will re-examine these ideas from a slightly different perspective, with a view to identifying the research question investigated in the empirical part of this study.

The lens model may be usefully considered as two interacting subsystems: a task subsystem, consisting of the relationships between the distal variable and the cues, and a response subsystem, consisting of the relationships between the cues and the subject's response. From this viewpoint, successful performance requires the subject to make some appropriate matching between the characteristics of his response subsystem and those of the task subsystem. As Hoffman (1960) notes, isomorphism between the two subsystems is not necessary for good performance: paramorphism will suffice.

In this view, then, the subject's task is to use whatever information he has available to adjust his response subsystem so as to match the characteristics of the task subsystem. The bulk of the reported laboratory research has provided the subject only with "outcome feedback" (e.g.: Hammond and Summers (1972); Dudycha and Naylor 1966a) to assist him in this "tuning," though recent work by Hammond and others (Sorensen (1967); Newton (1965); Todd and Hammond 1965)), has shown the advantages of providing the subject with more detailed information on characteristics of either task or response subsystems, or both. Other types of information which would allow such "tuning" have not, apparently, been systematically studied. Two exceptions are a study by Miklausich (1973) demonstrating the detrimental effects of error introduced into outcome feedback on the accuracy with which subjects can match a given task subsystem; and a study by Rose (in progress) of the effects of evaluation feedback on such matching. The present study will examine the effects of providing the subject with a verbal context implying the structure of the task subsystem.

A previously neglected aspect of the verbal context which will be considered here was proposed by Miller (1971). He was concerned with whether the name put on the cue was congruent with the values of the ecological correlations ( $r_{ie}$ ) or incongruent with them. He also raised the question of how feedback of the true values of the distal variable might interact with the cue labels. Miller had clinicians, statisticians, freshman math-oriented students, and other freshman students predict examination results, given three pieces of information. In one of the nine testing conditions, cues were labeled with what they actually represented; in the other eight they were either falsely labeled or not labeled at all. Except for the practicing statisticians, the subjects did worse when the conditions logically conflicted with the task subsystem correlations (which they were given).

One area for extension of the existing research, then, is the examination of modes other than outcome or lens model feedback to allow the subject to tune. A second extension is the range of task properties considered. It appears that Brunswik's original formulation (1956) of the Lens Model was in terms of a single underlying variable type of task, in which the cues serve as a set of indicators of the value of the underlying variable. Much of the subsequent laboratory work, however, has examined a rather different task subsystem, in which a set of orthogonal cue values determines the distal variable value either in a linear relationship, (e.g., Dawes (1971); Hammond, Hursch, and Todd (1964)), or a non-linear relationship, (e.g. Summers and Hammond (1966); Brehmer (1969)). Connolly (1973) argues that such procedures are inconsistent with Brunswik's formulations, and proposes a distinction between cue-type

problems for the single underlying variable case, and component type problems for the other cases.

Without developing Connolly's argument here in detail, it does appear of value to consider problem subsystems other than those in which there is a direct relationship between cues and distal variables. Of particular interest are what might be called complex structured problems in which there is at least one level of intermediate constructs between distal and cue variables. A simple example would be the estimation of the distal variable "area" where the entire cue set would be partitioned into a subset relating to "width" and a second relating to "length." In such a case, one would expect the within-subset correlations to be fairly high, while the between-subset correlations could appropriately be made equal to zero.

The interest of such complex structures problems is two-fold. First, they appear to represent a range of complex real-world problems more closely than do the simple structured tasks generally used in laboratory studies. Second, they require the subject to adopt a two-stage information-processing strategy to achieve the highest possible performance. This suggests that insight into the task structure would be an important determinant of performance - an insight which may be readily provided by means of a suitable verbal context, as discussed earlier.

In summary, the present research is intended to advance the domain the domain of laboratory studies in two areas. First, the task subsystem will be complex-structured. Second, the subject will be provided with a verbal context suggesting the structure of the task

subsystem, as well as the usual outcome feedback.

An underlying question which is not often discussed in the literature is that subjects seem to learn number puzzles in the laboratory very slowly, and yet appear able to handle successfully the real-world learning of highly complex problems. This study will check to see if the verbal construct idea bridges this gap in learning performance. The research question of interest may be summarized as follows:

#### Research Question

"What are the effects on learning and performance in a complex structured inference task of suggesting, and of imposing, an appropriate conceptual structure?"

The approach taken will be a three-group design. All three groups will work numerically identical problems, under conditions of no structuring insight, suggested structuring insight, and imposed structuring insight. Two major hypotheses will be examined:

#### Hypotheses

H<sub>1</sub>: In a complex-structured multiple cue inference task, subjects provided with a task-structuring insight will outperform subjects not provided with such insight.

H<sub>2</sub>: In a complex-structured multiple cue inference task, subjects forced to use the task-structuring insight will outperform those who have had such insight merely suggested to them.

The null hypothesis, in each case, is that the performance fails to be significantly superior.

The term complex-structured, as used above, indicates a problem in which the cues and groups of cues are presented in such a manner

that the decision maker can apply some task-structuring constructs to reduce the demands of the task. The detailed experimental design and procedure used to test these hypotheses are reported in the following chapter.

## CHAPTER 11

### EXPERIMENTAL DESIGN

#### Background

The primary feature of a well-designed experiment is a direct path - starting with the information desired, and working backward to the proper construction of an experiment to get this information.

To test the hypotheses proposed in Chapter 1, it was necessary to design a complex-structured problem with three treatments, corresponding to the three groups of subjects. Group I received the basic verbal context; its purpose was to orient the subject and arouse his interest in the problem. Group II received both the basic context, and a structuring context which implied the existence of intermediate constructs. Group III received all the information given the Group II subjects, but were, in addition, asked to record their estimates of the intermediate constructs, as well as their overall estimates.

#### Verbal Contexts

The basic verbal context was common to all three groups; and it was the only information available to Group I. This basic context placed the subject in the position of the personnel director for a large corporation, responsible for the acceptance or rejection of job applicants. In an effort to do a good job for the corporation, the personnel director (subject) wanted to devise an efficiency score prediction scheme using only four expert ratings for each applicant

(provided by consultants to the corporation). The context explained that an assistant had gone into the corporation files and pulled out a sample of efficiency scores of current employees, as well as the expert ratings given at the time they were hired. Thus the basic procedure was for the subjects to take four cues (labeled Expert A, etc.), make their response (efficiency score prediction), and then be given feedback (the actual current efficiency score from the files). The context was specifically designed to exclude any requirement for expertise in any specific field.

Group II was given a structuring verbal context in addition to the basic context. These subjects were told that two of the four expert ratings were by psychologists, and that the other two were by successful businessmen. The cues were relabeled PSY 1,2 and BUS 1,2. Thus intermediate constructs (overall psychological rating and overall business rating) were implied.

Group III subjects were given all the information above, and then were told to actually estimate and write down the overall psychological and business ratings as well as their predicted efficiency scores. Obviously, this actually forced the subjects in Group III to use (or at least note the presence of) intermediate constructs.

In summary then, each group of subjects worked a numerically identical problem (within each phase). The difference in treatment was that Group I received no insight into the structure of the problem, Group II had the structure implied, while Group III had the structure imposed.

#### Test Generation

Given the sensitivity of many Lens Model phenomena to slight



variations in problem characteristics (see, for example, Miklausich, 1973) it was decided to investigate two different versions (phases) of an essentially identical problem. The problem used throughout was a two-component, four-cue problem in which the distal variable ("efficiency score") was calculated as the product of two components ("psychological rating," P, and "business rating," B. Each component had two associated cues. The structure of the problem is shown in Figure 2. As may be seen, the problem is "complex-structured" in the sense discussed earlier.

For Phase A, the two (1 X 60) component matrices, named P and B, were generated from a normal distribution with mean = 5, and variance = 4; this provided a range (within two standard deviations) of 1 to 9. The scale was truncated at 1 and 9 to eliminate unreal values (in particular, the multiplication by zero in arriving at the distal variable). These matrices were then manipulated to provide the four cues (expert ratings), by adding induced error into the component array. The (4 X 60) error matrix was generated using a normal distribution with mean = 0 and variance = 1. The error scale was truncated at -1 and +1 (one standard deviation).

The Cue Matrix was generated according to the following formulae:

$$P_{1,i} = P_i + E_{1,i}, \quad i = 1, \dots, 60$$

$$P_{2,i} = P_i + E_{2,i} ;$$

$$B_{1,j} = B_j + E_{3,j}, \quad j = 1, \dots, 60$$

$$B_{2,j} = B_j + E_{4,j} .$$

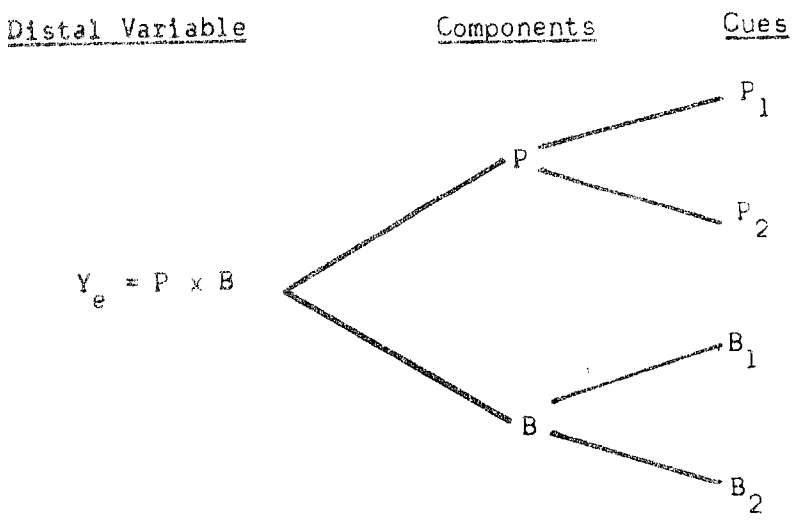


Figure 2. Problem Structure.

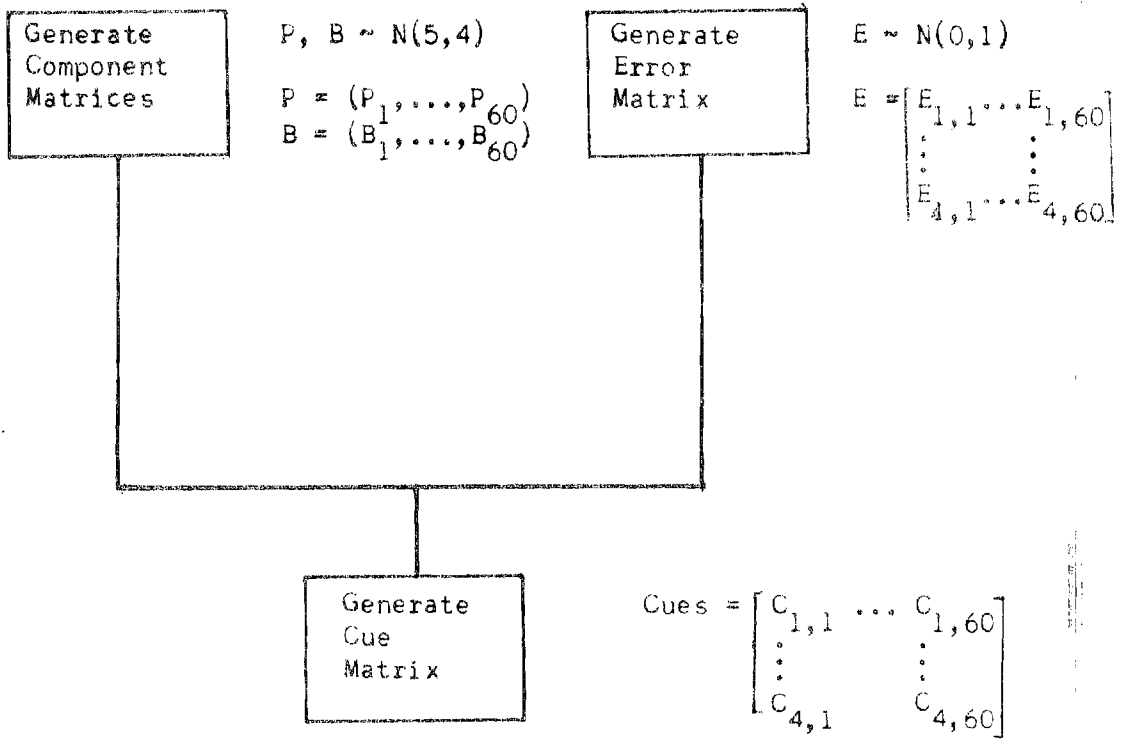


Figure 3. Test Generation Flow Diagram, Phase A.

and were arrayed for presentation to the subjects as  $P_{1,i}$ ;  $B_{1,j}$ ;  $P_{2,i}$ ;  $B_{2,j}$ , so as to avoid the intimation of the common origins of the P cues, and B cues, by their juxtaposition.

The same two-component, four cue task was used for both phases A and B, with  $Y_e$  equal to the product of the components in each case. In Phase A, the distribution of  $Y_e$  values followed that suggested by the histogram shown in the Appendix. In Phase B, the distribution of  $Y_e$  was approximately uniform across the range 1 to 60. For the sample of trials used, Phase A had an  $R_e$  value of 0.94, Phase B of 0.96. From the subjects' viewpoint, the major difference between Phase A and Phase B was expected to be increased frequency of extreme  $Y_e$  values in the latter. Since extreme values are probably highly diagnostic of problem structure, performance in Phase B was expected to exceed that in Phase A.

#### Equipment and Materials

An overhead projector was fitted with a roll of acetate which ran over a mounted transparency. The transparency displayed the headings appropriate to the subject group, e.g., for Group I:

EMPLOYEE NUMBER	EXPERT A	EXPERT B	EXPERT C	EXPERT D	ACTUAL EFFICIENCY SCORE
--------------------	-------------	-------------	-------------	-------------	-------------------------------

The cues and distal variable sets were written on the acetate roll, and were then rolled over the transparency. The actual efficiency score was covered by a card until the subjects had responded.

The answer booklet used by the subjects was reproduced by the XEROX copier, and was stapled together. The subjects were instructed to pull out the staple, and to manipulate the individual pages of the

booklet in any way they saw fit.

### Subjects

The members of the subject pool were unpaid juniors, seniors and graduate students attending the Georgia Institute of Technology. Each member of the pool was assigned to a phase, and to a particular group (five subjects per cell), solely on the basis of his random arrival at the testing site.

### Procedure

The subject groups were physically separated while the test instructions were given, and during the test itself, so that no member of one group could hear and compare the instructions being given to another group.

Each group was established in a comfortable room in standard institutional furniture. An effort was made to make the test site and all materials to appear neat, clean, and orderly - so that each subject would feel that he was working with a substantial investigation, and was not wasting his time by taking the test.

The groups were seated, and each subject was asked if he could see the projection screen clearly. When all were settled, the scenario (the appropriate verbal contexts) was read carefully and distinctly, with a request for questions after its reading. Immediately following, four examples were considered. The example data sets were shown on the projection screen, as well as on an example sheet in the subject's answer booklet. The subjects were told to tear out their example sheet

and keep it handy to their work area. Additionally, the several key points from the scenario (including the cue and response scales), printed at the top of the example sheet, were read aloud.

After answering all questions about the scenario and examples, the test data sets of four cues were shown, and read aloud, one set at a time. After all the subjects had written down their predictions for each four cue set, and had so indicated by looking up from their work, the "Actual Efficiency Score" (distal variable feedback) for that set was uncovered by the monitor, and shown to, read aloud to, and recorded by the subjects.

The data sets were shown sequentially until all 60 test data sets had been covered. The subjects in each group were given two one-minute breaks between each block of twenty test data sets to review their work, and to modify their approaches if they desired.

Therefore there were established, for each phase, three groups which corresponded to the three levels of information (insight) described above. The subtle change between the phases of the experiment is caused solely by the source of the numbers that were used by the subjects. That is, the underlying distribution of Phase A numbers was normal, while that of Phase B numbers was uniform.

## CHAPTER III

## RESULTS AND DISCUSSION OF RESULTS

Primary Analysis

As explained previously, the experiment consisted of two phases. The only difference between phases was that Phase A had a normal underlying distribution, while Phase B's distribution was uniform.

The cues and distal variables for both phases, with their associated subject responses, were processed through routines of the Univac STAT-PACK. The RESTEM routine was used to obtain the linear predictions  $\hat{Y}_e$  and  $\hat{Y}_s$ , and then the CORAN routine was used for the correlations  $r_a$ ,  $R_e$ ,  $R_s$ , and  $G$ . Prior to further analysis, the correlations were transformed to Fisher's  $z$  values, since the distribution of the  $z$  values is approximately normal. The transformation allowed averaging and analysis of the variance. See McNemar (1969) for further discussion of Fisher's transformation.

The experimental task of Phase A was performed by the three groups of subjects, corresponding to the three treatment levels defined by the verbal context (insights). Since the hypotheses concerned the improvement in performance between groups, the statistic  $r_a$  was plotted (Figure 4). The value of the limit of achievement, and the knowledge about the properties of the task were also plotted to help with interpretation. The data provide weak support ( $p < 0.10$ ) for Hypothesis 2: that the performance of the group of subjects forced to use the task-

$H_1$ : Subjects provided with a task-structuring insight (Group II) will outperform subjects not provided with such insight (Group I).

$H_2$ : Subjects forced to use the task-structuring insight (Group III) will outperform those who have had such insight merely suggested to them (Group II).

$H_0$ : Performance fails to be significantly superior.

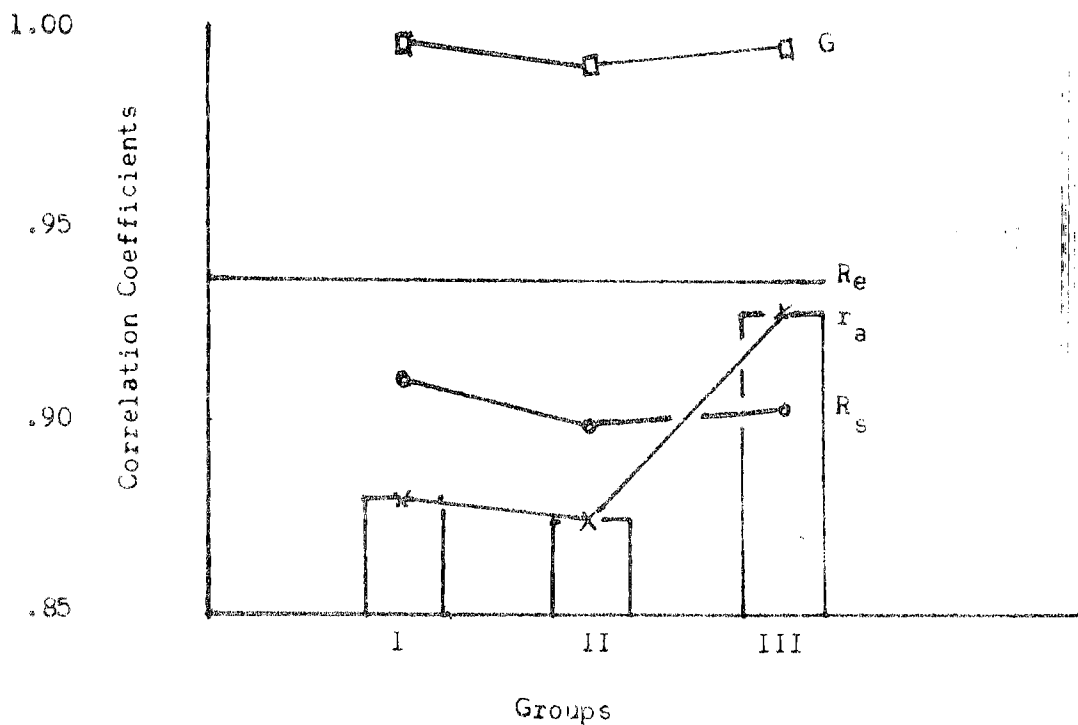


Figure 4. Display of Results, Phase A.

Table 1. Results and Analysis of Phase A ( $R_e = 0.94$ )

<u>Results</u> ( $r_a/z_a$ )		Groups		
		I	II	III
Question 1		0.775/1.034	0.822/1.163	0.837/1.214
Blocks	2	0.901/1.488	0.888/1.413	0.935/1.696
	3	0.925/1.625	0.903/1.487	0.965/2.019

ANOVA (on  $z_a$ )

Source	SS	df	MS	Ratio
Insight	0.1534	2	0.077	5.105 ( $p \leq 0.10$ )
Ques. Blks.	0.5156	2	0.258	17:161 ( $p \leq 0.025$ )
Error	0.0600	4	0.015	
Total	0.7290	8		

\*  $F(0.10, 2, 4) = 4.32$ ; reject the null hypothesis.

Duncan's Test

$s_e = 0.054812$ ;  $n_2 = 6$ ;  $\alpha = 0.05$ .

$p =$	2	3
Sig. Range	3.46	3.58
Least Sig. Rnd	0.190	0.196
$\mu_I = 1.380$	II <u>vs</u> III: 0.268 > 0.196	
$\mu_{II} = 1.355$	I <u>vs</u> III: 0.264 > 0.190	
$\mu_{III} = 1.643$	I <u>vs</u> II: -0.025 < 0.190	

\* Group III outperformed Group II;  $H_2$  is supported.  
Group II did not outperform Group I;  $H_1$  is not supported.



structuring insight was superior to that of the group who had the insight merely suggested to them.

Phase B was identical to Phase A (with the sole exception of the numbers in the problem), and was purposely conducted in the same manner as Phase A. The raw data was coded in the same manner, and transformed (by Fisher's transform) in the same manner as that of Phase A. To check on the group performance differences, the statistic  $r_a$  for each group (along with  $R_e$ ,  $R_s$ , and  $G$ ) was plotted (Figure 5). These data provide strong support ( $p < 0.005$ ) for both of the hypotheses (Table 2).

#### Secondary Analysis

Although the primary analysis was most satisfactory in supporting the hypotheses, some additional examinations were made. Recall that during the administration of the experiment, the subjects were given a break after the twentieth question and the fortieth question. These made natural divisions in the work, so that it was possible to investigate three blocks of questions, and check the groups' achievement across the blocks. Thereby, it was seen if the subject's improvement in achievement was caused by experience gained by taking the test, or by the verbal contexts tested. In both Phases A and B it appeared that the subject's performance improved by the verbal context, and not by the experience of taking the test.

As another method of getting at this question, a ratio, FA, was defined as percent achievement, or the ratio of average achievement with average possible achievement. The conclusion reached above that the

$H_1$ : Subjects provided with a task-structuring insight (Group II) will outperform subjects not provided with such insight (Group I).

$H_2$ : Subjects forced to use the task-structuring insight (Group III) will outperform those who have had such insight merely suggested to them (Group II).

$H_0$ : Performance fails to be significantly superior.

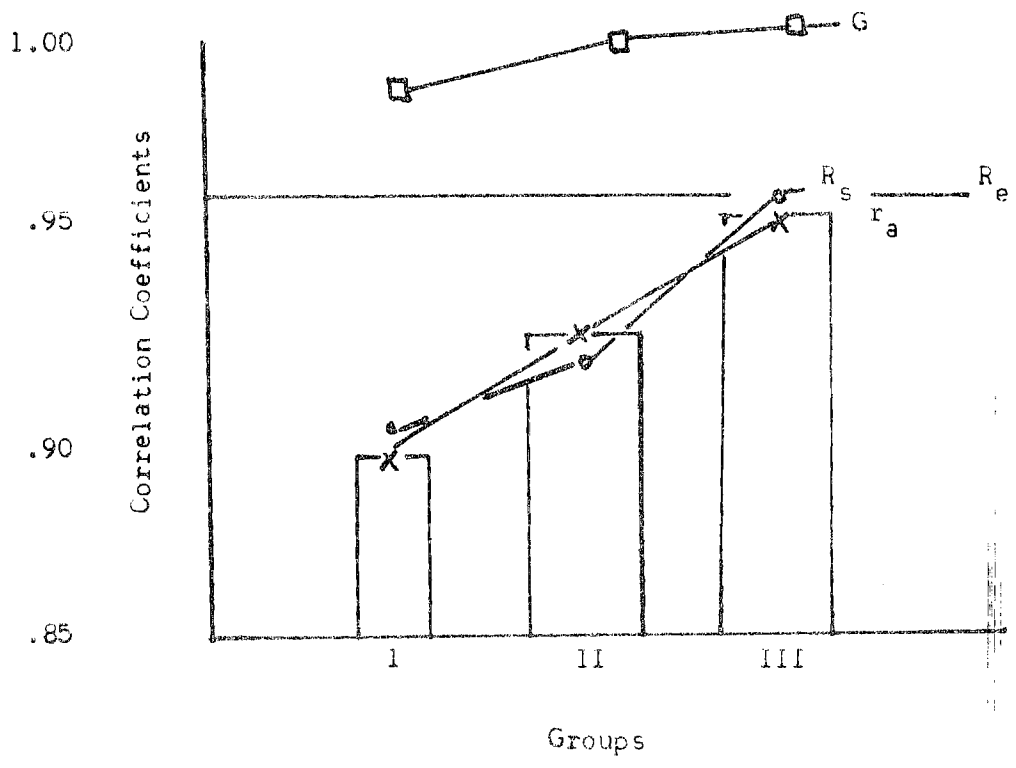


Figure 5. Display of Results, Phase B.

Table 2. Results and Analysis of Phase B ( $R_e = 0.96$ )

Results ( $r_a/z_a$ )	Groups		
	I	II	III
Question 1	0.881/1.379	0.909/1.520	0.934/1.689
Blocks 2	0.929/1.654	0.946/1.794	0.969/2.079
3	0.867/1.320	0.917/1.571	0.951/1.842

ANOVA (on  $z_a$ )

Source	SS	df	MS	Ratio
Insight	0.266	2	0.133	41.325 (p = 0.005)
Ques Blks	0.170	2	0.085	26.488 (p = 0.010)
Error	0.013	4	0.003	
Total	0.489	8		

\*  $F(0.01, 2, 4) = 18.00$ ; reject the null hypothesis.

Duncan's Test

$s_e = 0.025$ ;  $n_2 = 6$ ;  $\alpha = 0.05$

	p = 2	3
Sig Range	5.24	5.51
Least Sig. Range	0.133	0.140
$\mu_I = 1.450$	$\mu_I$ vs $\mu_{III}$ : 0.421 > 0.140	
$\mu_{II} = 1.628$	$\mu_{II}$ vs $\mu_{III}$ : 0.242 > 0.133	
$\mu_{III} = 1.870$	$\mu_I$ vs $\mu_{II}$ : 0.179 > 0.133	

\* Group III outperformed Group II;  $H_2$  is supported.  
 Group II outperformed Group I;  $H_1$  is supported.

increased performance is caused by the verbal context rather than the experience of taking the test seems justified here as well. If each of the iso-group lines had a common origin, but varied at their conclusion, a different learning experience between groups would be shown. However, the iso-group lines, particularly in Phase B, are nearly parallel; this indicated the same learning experience, with the percent achievement varying only among contexts. See Figures 6 and 7 for plots of percent achievement versus question blocks.

A variation of the percent achievement ratio provided yet another analysis of the data. Based on the structure of the problem, a B and P were calculated as the average of their associated cues (recall that the cues were generated from  $\hat{B}$  and  $\hat{P}$ , rather than the opposite). Then, a variable called  $Y_e^*$  was defined as the product of  $\hat{B}$  and  $\hat{P}$ . Finally, the correlation  $R_e^* = r_{Y_e Y_e^*}$  was calculated, and  $PA^* = r_a / R_e^*$  was plotted in Figures 8 and 9, for Phases A and B respectively. It is felt that this variation is more illustrative than the previous method, since it recognizes that the subject can actually exceed the  $R_e$  from the conventional lens model calculations, i.e., that the best possible stable performance for this problem can actually exceed what was considered the upper limit of achievement. Not incidentally, this method leads one to exactly the same conclusions as above - that the hypotheses have been supported.

In summary, then, the data provide fairly strong support for both major hypotheses, with some indications as to the mechanisms involved. Phase B data strongly support the hypotheses of higher achievement of subjects provided with a structuring insight over those not so helped,

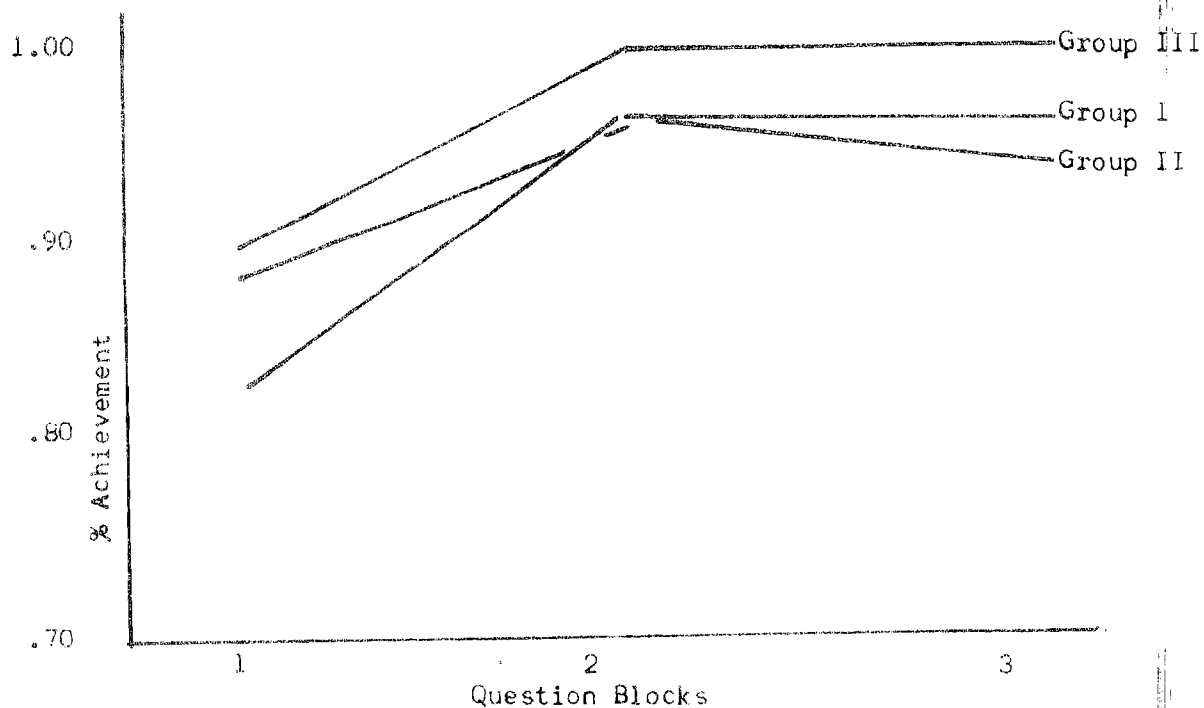


Figure 6. Percent Achievement Using  $R_e$ , Phase A.

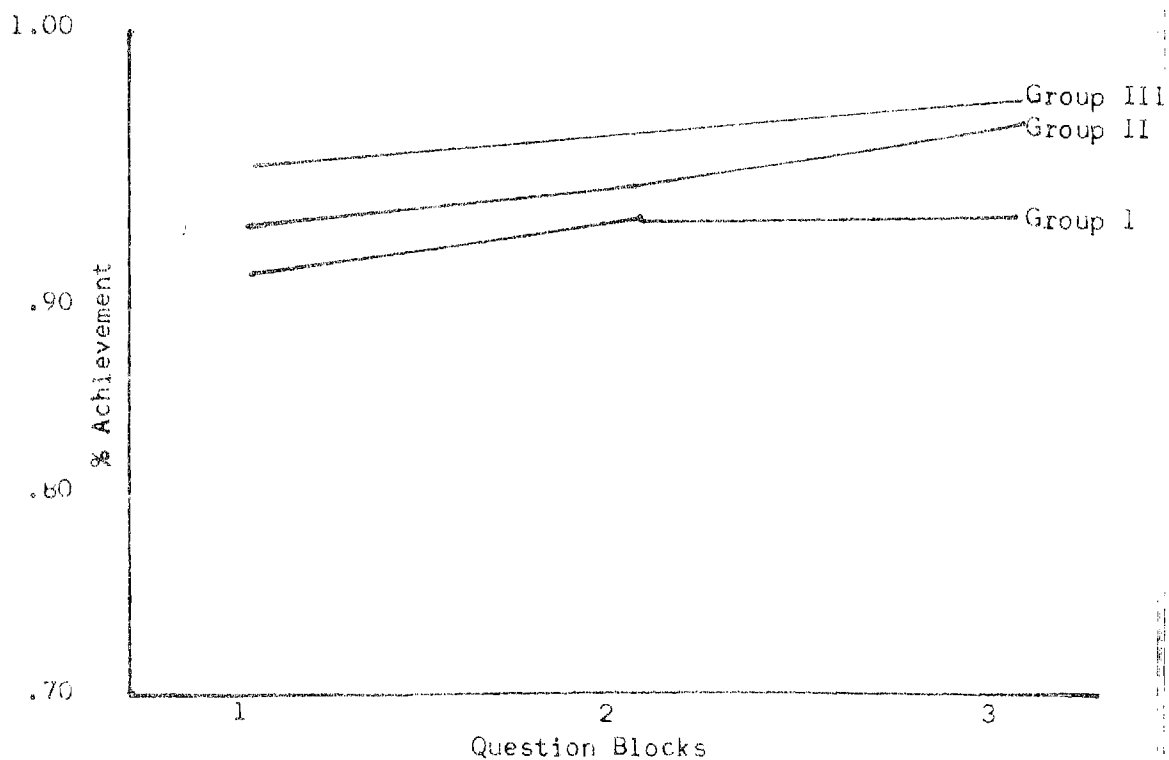


Figure 7. Percent Achievement Using  $R_e$ , Phase B.

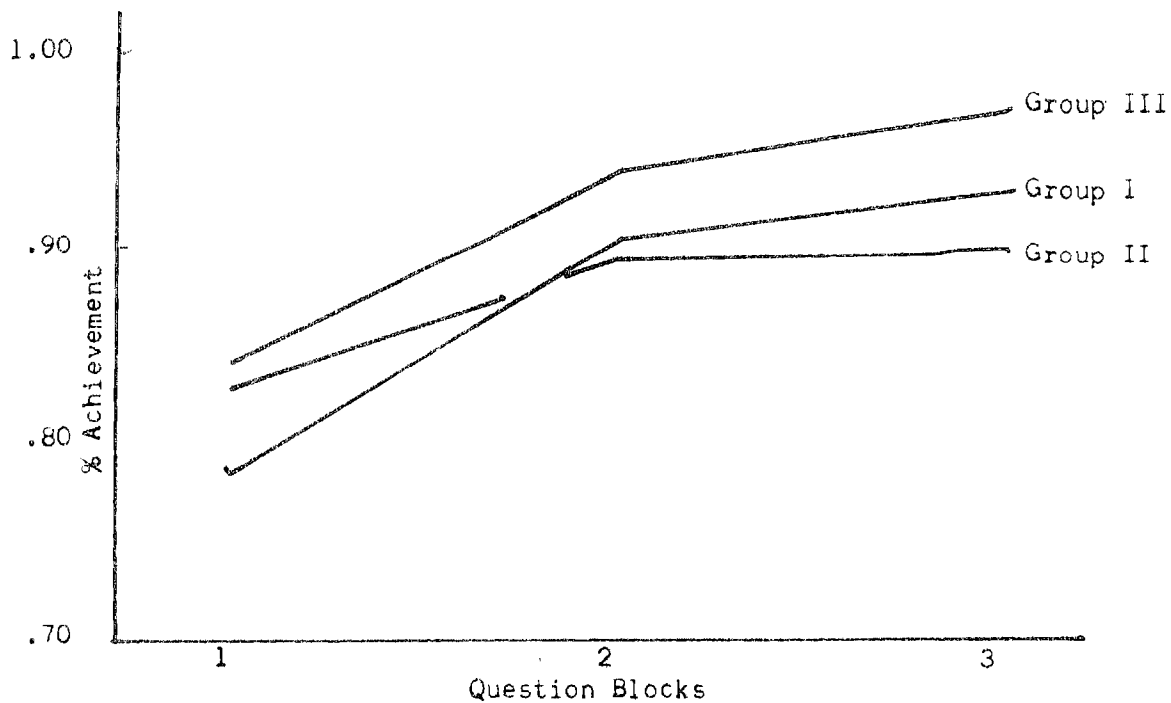


Figure 8. Percent Achievement Using  $R_e^*$ , Phase A.

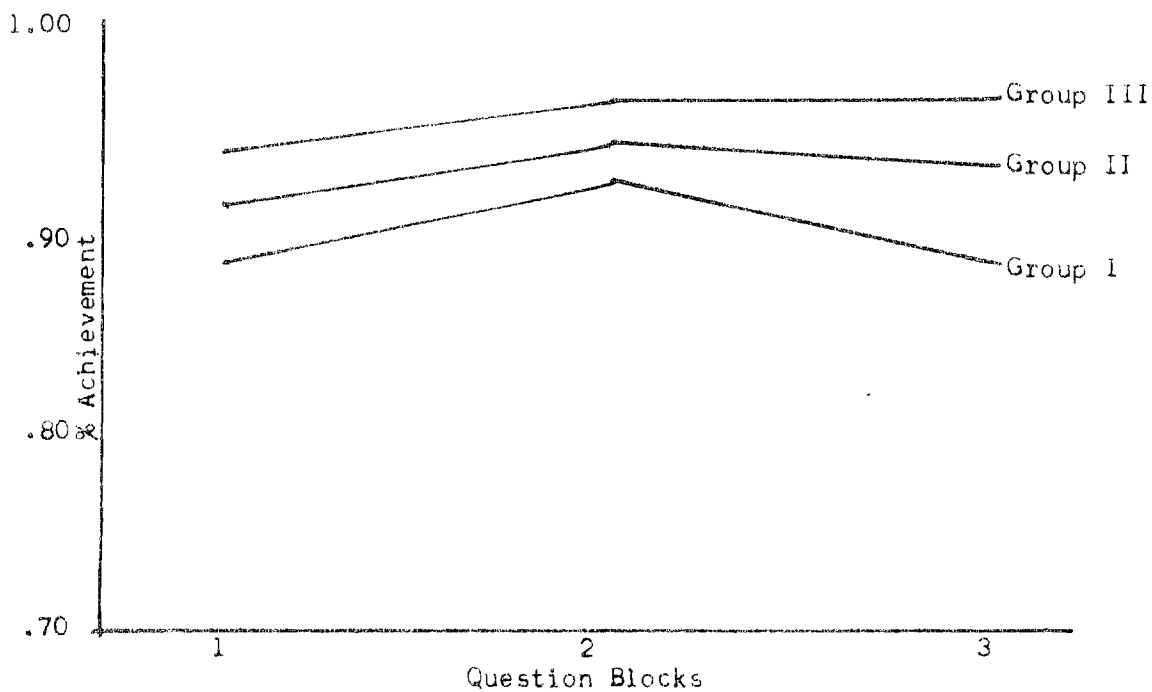


Figure 9. Percent Achievement Using  $R_e^*$ , Phase B.

Table 3. Percent Achievement, Phase A

Question Block	$R_e$	$R_e^*$
1	0.928	0.988
2	0.924	0.992
3	0.956	0.996
Overall	0.938	0.992

% Achievement Using $R_e^*$				
		Groups		
		I	II	III
Question 1	0.785		0.832	0.848
Blocks 2	0.908		0.895	0.942
3	0.929		0.907	0.969

% Achievement Using $R_e$				
		Groups		
		I	II	III
Question 1	0.835		0.886	0.902
Blocks 2	0.976		0.962	*1.012
3	0.968		0.944	*1.010

Table 4. Percent Achievement, Phase B

Question Block	$R_e$	$R_e^*$
1	0.962	0.939
2	0.985	0.996
3	0.919	0.994
Overall	0.963	0.975

% Achievement Using $R_e^*$				
		Groups		
		I	II	III
Question 1	0.892		0.920	0.946
Blocks 2	0.935		0.951	0.975
3	0.889		0.941	0.975

% Achievement Using $R_e$				
		Groups		
		I	II	III
Question 1	0.916		0.945	0.971
Blocks 2	0.944		0.961	0.984
3	0.943		0.998	*1.035

\*Note that % achievement exceeds 100% in some calculations with  $R_e$ .

and of subjects required to use such insight over those with insight merely available. Phase A data tends to confirm the latter hypothesis, though showing no evidence of the former.

Examination of Figures 6 - 9 suggests quite strongly that superior achievement here is not the result of faster learning for the insight, and forced insight groups. Rather, it appears that all groups learn at essentially the same rate, with the highest performing groups starting with an initial advantage, and maintaining it throughout the experiment.

A final comment should be added on the overall difference between Phase A and Phase B data. Recalling that the only difference between the Phases was in the underlying  $Y_e$  distribution, it should be noted, first, that the phenomena identified here are sensitive to apparently minor changes in problem characteristics, and second, that the anticipated higher diagnosticity of the uniform distribution used in Phase B appears to be confirmed.



## CHAPTER IV

### CONCLUSIONS AND RECOMMENDATIONS

#### Review of Results

The experiment was designed to test if people, given an insight into the structure of a task, could use the insight to assist them in performing the task. The task, in this experiment, was to integrate cue information into the estimation of a distal variable.

Each of the three groups of subjects was given a basic motivating verbal context; Group I received only that. Group II was given some insight when the grouping of pairs of cues was implied. This insight was reemphasized and actually enforced for Group III, which was required to write down its overall estimates of the factors represented by the grouped pairs.

As described and analyzed in detail above, it was demonstrated that under certain conditions, people can, and do, use verbal context successfully to establish their model.

#### Relationship to Real World

The first and most obvious comparison to be made when comparing laboratory experiments with those of field observations is that the laboratory problems generally require no professional expertise (set of concepts) with which the subject is to work. Most laboratory experimental designs specifically engineer out any such requirement of the

subject, and provide any necessary concepts in the conduct of the experiment (verbal context, or outcome feedback). The laboratory subject may of course use combination rules which he has tried and found useful in some setting other than the laboratory - but this does not constitute professional experience.

The results in Chapter III show that the stated hypotheses are strongly supported for one set of data, less clearly for another. Thus the phenomena seem extremely sensitive to changes in the underlying problem characteristics. This sensitivity is not particularly helpful to an attempt to extend the thesis to real life decisions. Follow up experiments with an examination of a broader range of problems are clearly necessary.

Nonetheless, it was shown herein that, given the situations and accompanying restrictions described, conceptual structures can be suggested or imposed by means of an appropriate verbal construct, and that these conceptual structures impact on the degree of success in a complex-structured, multiple cue inference task.

#### Extension and Expansion

Despite the modest dimensions of the present effort, it is felt that the field of inquiry into the study of complex-structured tasks versus that of single underlying variable, has been opened. Obviously, replication and altered replication of this experiment should be made to firmly establish this area of the problem space. Immediately thereafter, the exploration of the range of available models and other elements of the inferential base and their relationship (already begun by

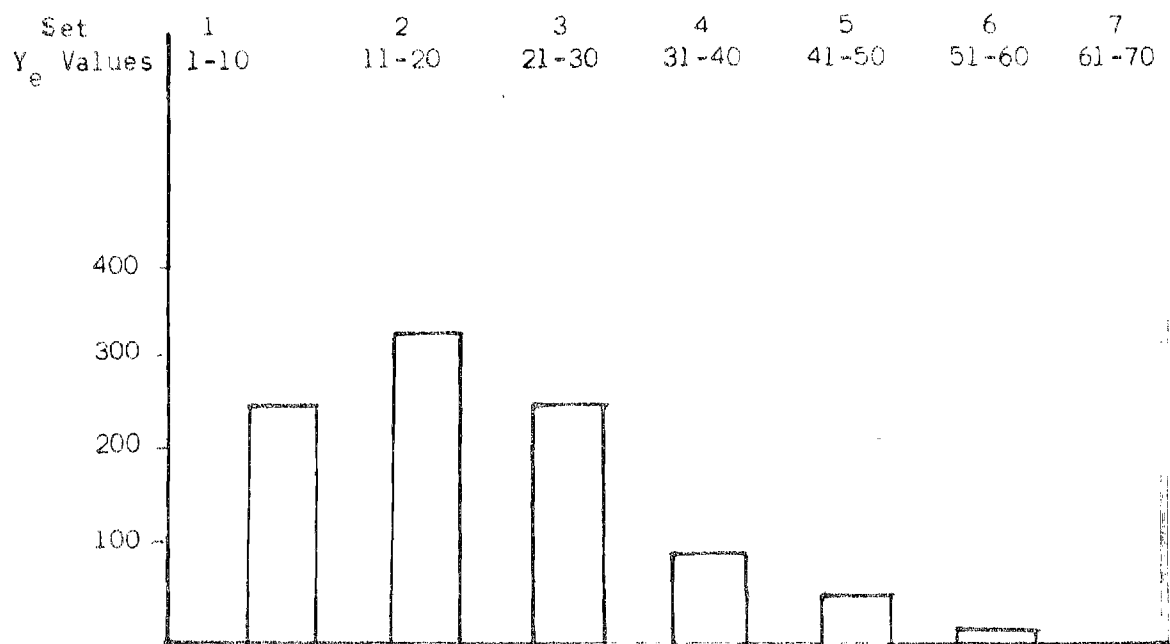
Miklausich and Rose, as previously noted), should be continued and expanded.

Another area deserving of attention is the study of teaching as providing a conceptual framework for learning (tested by the rate of learning).

When we understand the aggregate complex-structured task study, we may understand the relationship between the laboratory and the real world.

## APPENDIX

## HISTOGRAPH OF DISTRIBUTION



## BIBLIOGRAPHY

- Agnew, N. M., and Pyke, S. W. The science game: an introduction to research in the behavioral sciences. Englewood Cliffs, N.J.: Prentice-Hall, 1969.
- Anderson, N. H. and Jacobson, A. Further data on a weighted average model for judgment in a lifted weight task. Perception and Psychophysics, 1968, 4, 81-84.
- Beach, L. R. Multiple regression as a model for human information utilization. Org. Beh. and Hum. Perf., 1967, 2, 276-289.
- Brehmer, B. Inference behavior in a situation where the cues are not reliably perceived. Org. Beh. and Hum. Perf., 1970, 5, 330-347.
- Brehmer, B. Note on the relation between single-cue probability learning and multiple-cue probability learning. Org. Beh. and Hum. Perf., 1973, 9, 246-252.
- Brehmer, B. and Lindberg, L. The relation between cue dependency and cue validity in single-cue probability learning with scaled cue and criterion variables. Org. Beh. and Hum. Perf., 1970, 5, 542-
- Brunswik, E. Perception and the representative design of psychological experiments. Berkeley: Univ. of Cal. Press, 1956.
- Connolly, T. 'Making sense of the problem': the inferential base in lens model research. Unpub.: 1973.
- Dean, D. H., Hammond, K. R., and Summers, D. A. Acquisition and application of knowledge in complex inference tasks. J. Exp. Psych., 1972, 92, 20-26.
- Dawes, R. M. A case study of graduate admissions: application of three principles of human decision making. American Psychologist, 1971, 26, 180-188.
- Dawes, R. M. Objective optimization under multiple subjective functions. Unpub.: 1972.
- Dudycha, L. W., and Naylor, J. C. Characteristics of the human inference process in complex choice behavior situations. Org. Beh. and Hum. Perf., 1966, 1, 110-128.

- Dudycha, L. W., and Naylor, J. C. The effect of variations in the cue R matrix upon the obtained policy equations of judges. Education and Psychological Measurement, 1966b, 26, 583-603.
- Edwards, W. Bayesian and regression models of human information processing - a myopic perspective. Org. Beh. and Hum. Perf., 1971, 6, 639-648.
- Hammond, K. R. Computer graphics as an aid to learning. Science, 1971, 172, 903-908.
- Hammond, K. R. Probabilistic functionalism: Egon Brunswik's integration of the history, theory, and method of psychology. The Psychology of Egon Brunswik. New York: Holt, Rinehart, and Winston. 1966.
- Hammond, K. R. Probabilistic functioning and the clinical method. Psychological Review, 1955, 64, 255-262.
- Hammond, K. R., Hursch, C. J., and Todd, F. J. Analyzing the components of clinical inference. Psychology Review, 1964, 71, 438-456.
- Hammond, K. R., and Summers, D. A. Cognitive Control. Psychological Review, 1972, 79, 58-67.
- Hines, W. W., and Montgomery, D. C. Probability and statistics in engineering and management science. New York: The Ronald Press Company, 1972.
- Hoffman, P. J. The paramorphic representation of clinical judgment. Psychological Bulletin, 1960, 47, 116-131.
- Hursch, C. J., Hammond, K. R. and Hursch, J. L. Some methodological considerations in multiple-cue probability studies. Psychological Review, 1964, 71, 42-60.
- McClelland, D. D. The achievement motive. Harvard Business Review, July-Aug. 1962, 99-112.
- McNemar, Q. Psychological statistics. New York: John Wiley and Sons, Inc., 1969.
- Miklausich, V. M. The impact of error-prone feedback on learning in a multiple-cue decision problem. Georgia Institute of Technology. Atlanta, Ga.: 1973.
- Miller, P. McC. Do labels mislead? Org. Beh. and Hum. Perf., 1971, 6, 480-500.
- Newton, J. R. Judgment and feedback in a quasi-clinical situation. Journal of Personality and Social Psychology, 1965, 1, 336-342.

- Raiffa, H. Decision analysis. Reading, MA: Addison-Wesley, 1970.
- Rose, D. E. Thesis in progress. Georgia Institute of Technology, Atlanta, Ga.: 1973.
- Slovic, P., and Lichtenstein, S. Comparison of Bayesian and regression approaches to the study of information processing in judgment. Org. Beh. and Hum. Perf., 1971, 6, 649-744.
- Sorensen, J. D. Learning to use multiple cues in making clinical judgments: a multiple regression analysis. Dissertation Abstracts, 1967, 28, 1212B-1213B.
- 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.
- Tucker, L. C. A suggested alternative formulation.... Psychological Review, 1964, 71, 528-530.
- Yntema, D. B., and Torgerson, W. S. Man-computer cooperation in decisions requiring common sense. IRE Transactions on Human Factors in Electronics, 1961, 20-26.