TRANSFER WITH LENS MODEL TRAINING

Derick Otis Steinmann

A Dissertation

Submitted to the Graduate School of Bowling Green
State University in partial fulfillment of
the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 1972

Approved by Doctoral Committee

Michael E. Doherty Advisor
Department of Psychology

Richard L. Ecker
Graduate School Representative

Donald M. Ryznar

Charles J. Cronney

BOWLING GREEN STATE UNIVERSITY LIBRARY

ABSTRACT

The effects of training on the ability to learn probabilistic cue-criterion relationships as a function of differential feedback and previous training sessions were assessed.

Naive female subjects were trained to use three uncorrelated cues in the same manner as a hypothetical subject (target), and received either outcome feedback (OC group), outcome feedback and feedforward (FC group), or outcome feedback and lens model feedback (LMF group).

Outcome feedback consisted of the target's actual judgment; feedforward consisted of statistical information about the target's use of the cues, and lens model feedback permitted the subject to compare the statistical properties of the target's use of the cues with her own use of the three cues.

There were three questions of interest: 1) What are the effects of differential feedback? 2) Is lens model feedback used more effectively as a function of previous problems? 3) Do judges use outcome feedback more effectively after previous training with lens model feedback?

A task was defined by the statistical characteristics relating each cue to the criterion (eight tasks were used); a problem specified the order of the tasks, and within each problem, there were two sessions: all judges received outcome feedback during Session 1, and differential feedback, depending on group, during Session 2.

Performance was measured by r_a , the correlation between the target's judgments and the subject's judgments; G , reflecting the extent to which the subject's use of the cues was proportional to the target's use of the cues, and R_s , measuring the consistency with which the subject used the cues.

Performance increased from Session 1 to Session 2 for the LMF and FC groups, but did not change for the OC group. There was no systematic effect of problems on increased effectiveness of feedforward or lens model feedback. Judges in the LMF group used outcome feedback more effectively as a function of problems; the increases in r_a were due to increases in R_s , and not G .

ACKNOWLEDGMENTS

Dr. Michael Doherty's invaluable and extensive contributions to all aspects of my graduate education are deeply appreciated. I wish to thank Dr. Charles Cranny, Dr. Richard Eakin, and Dr. Donald Ragusa for serving as members of my doctoral committee.

I would like to express my gratitude to my wife, Marcia, for her encouragement.

This research was supported by PHS Research Grant No. MH-17487-02 from the National Institute of Mental Health awarded to Dr. Michael Doherty and Dr. Stuart Keeley.

TABLE OF CONTENTS

	Page
INTRODUCTION	1
Multiple Cue Probability Learning Tasks	1
The Lens Model	2
Types of Feedback	6
Transfer of Training	11
The Present Research	12
METHOD	13
MCPL Paradigm	13
Definitions of Terms	13
Task	13
Problem	13
Trial	13
Session	16
Subjects	16
Apparatus	16
Procedure	17
Lens Model Feedback Group	17
Feedforward Control Group	17
Outcome Feedback Control Group	17
Task Characteristics and Generation of Tasks	21
Design	22
Between-groups Analysis	22
Within-group Analysis	22

	Page
RESULTS	23
Dependent Measures	23
The Effects of Differential Feedback	24
Between-groups Analyses	24
Z_a	24
Z_s	26
Z_G	29
Within-groups Analyses	29
Summary of the Effects of Differential Feedback	32
Changes Over Problems	32
Learning to Use Lens Model Feedback	32
Between-groups Analyses	32
Within-groups Analyses	33
Learning to Use Outcome Feedback	33
The Effects of Cue Standard Deviations	36
Summary of Results	36
DISCUSSION	38
Differences among Groups	38
Changes over Problems	40
REFERENCES	43
APPENDICES	47
Appendix A	49
Appendix B	56
Appendix C	60

LIST OF TABLES

TABLE	Page
1. Task Characteristics	14
2. General Design	15
3. Summary of ANOVAS (Between-groups Analyses) . .	25
4. Summary of ANOVAS for Simple Effects (Differences between Sessions: Between- groups Analyses)	27
5. Groups x Sessions Summary Tables (Between- groups Analyses)	28
6. Summary of ANOVAS (Within-group Analyses: LMF Group)	30
7. Problems x Sessions Summary Tables (Within- group Analyses: LMF Group)	31
8. Groups x Problems Summary Tables (Between- groups Analyses)	34
9. Summary of ANOVAS for Simple Effects Dif- ferences among Problems: Between-groups Analyses)	35
10. Correlations of Cue Standard Deviations and Ecological Validities with Utilization Co- efficients for LMF and FF Groups (N=76)	37

LIST OF FIGURES

FIGURE	Page
1. The Lens Model	4
2. Apparatus	19

INTRODUCTION

The general purpose of this study was to assess the effects of training on judges' ability to learn probabilistic cue-criterion relationships. The type of information made available to the judges, and the number of previous training problems were the primary independent variables; performance on multiple cue probability learning (MCPL) tasks was assessed by several dependent measures.

Multiple Cue Probability Learning Tasks

On each trial in a MCPL task, the judge observes values of a specified number of cues and makes a judgment with respect to an environmental criterion. The environmental criterion may be a linear combination of the cues, typically with an error component. There are many trials, and the judge receives some measure of how well his judgments approximate the environmental criterion values. The specific measure may consist of the value of the environmental criterion (outcome feedback) on each trial, and/or more summary information, computed over blocks of trials.

For example, using MCPL techniques, judges have been trained to predict the socio-economic development of hypothetical nations on the basis of four cues: (a) amount

of foreign investment; (b) level of governmental control over private industry; (c) quantity of foreign aid, and (d) number of socialists in the nation's Parliament (Summers, Taliaferro, and Fletcher, 1970). The judge is trained to use the cues in the same way as another subject, called a target. Outcome feedback consists of presenting the judge with the target's actual response; more summary information consists of the correlations of each cue with socio-economic development.

Multiple cue probability learning tasks may be constructed which abstract many important characteristics inherent in "real life" judgmental tasks (Hammond and Summers, 1972), with Brunswik's lens model (Figure 1) providing a theoretical basis for the statistical analysis of the relationship between the judgments and the environmental criterion values. The MCPL task also permits the manipulation of the statistical environmental dimensions of interest; dimensions which are often not measurable in real life situations, for lack of an objective, environmental criterion.

The Lens Model

The theoretical framework of the research is the lens model (Brunswik, 1955), as adapted by Hammond (1966). The statistical basis of the lens model (Hursch, Hammond, and Hursch, 1964) has been reformulated (Tucker, 1964;

Figure 1. The Lens Model.

Dudycha and Naylor, 1966; Castellan, 1971a; Rozeboom, 1971), generalized to multiple criteria (Castellan, 1971b), and associated with independent psychological processes (Hammond and Summers, 1972). The statistical notation of Tucker (1964) will be used in this paper.

Given a set of standardized random variables, Y_e , Y_s , $X_1 \dots X_n$, where Y_e is a criterion value, and Y_s is the judge's estimate (judgment or prediction) of the criterion, and $X_1 \dots X_n$ are uncorrelated cues, the predicted criterion value is

$$\hat{Y}_e = r_{1e} X_1 + \dots + r_{ne} X_n \quad (1)$$

and the predicted judgment is

$$\hat{Y}_s = r_{1s} X_1 + \dots + r_{ns} X_n \quad (2)$$

where r_{1e} is the correlation of the first cue with criterion (ecological validity) and r_{1s} is the correlation of the first cue with the judgment (utilization coefficient). The residual, $Y_e - \hat{Y}_e$ is designated Z_e , and $Y_s - \hat{Y}_s$ is Z_s .

The lens model equation is

$$r_a = GR_e R_s + C(1 - R_e^2)^{1/2} (1 - R_s^2)^{1/2} \quad (3)$$

where

$r_a = r_{Y_e Y_s}$: the validity coefficient of the judge;
the correlation between the judge's predictions and the criterion values.

- $G = r_{\hat{Y}_e \hat{Y}_s}$: the correlation between the predicted values of the environment and the predicted values of the judge.
- $R_e = r_{Y_e \hat{Y}_e}$: the linear predictability of the criterion; the multiple correlation between the cue values and the criterion values.
- $R_s = r_{Y_s \hat{Y}_s}$: the linear predictability of the judge; the multiple correlation between the cue values and the judge's predictions.
- $C = r_{Z_e Z_s}$: the nonlinear component of r_a ; the correlation between the residual values of the criterion and the residual values of the judge's predictions.

Two additional correlations, not explicitly formulated in the above development, are the correlation of the judgments with the predicted criterion values, $r_{\hat{Y}_e Y_s}$, or r_{ce} , and the correlation of the criterion values with the predicted judgments, $r_{Y_e \hat{Y}_s}$, or r_{cs} . r_{ce} is equal to GR_s and r_{cs} is equal to GR_e (Dudycha and Naylor, 1966). (In Dudycha and Naylor's notation, r_e is equal to R_e , r_s is equal to R_s , and r_m is equal to G .)

Two important dimensions are not depicted in Figure 1: a feedback dimension, reflecting the type of information available to the judge, and a time dimension, reflecting potential transfer effects.

Types of Feedback

Slovic and Lichtenstein (1971) present an extensive review of the research done within the framework of the lens model. While the focus of the present study was the transfer of training from problem to problem, the research

dealing with training on a specific problem is relevant and will be reviewed briefly.

The term outcome feedback refers to the general procedure in which the judge observes the cue values, makes a judgment about the criterion, and immediately receives the criterion value as feedback. Hammond (1971, p. 904) designates the kind of learning attempted with outcome feedback as "stupid" learning, because "... outcomes are related to the cues in a complex, multi-determined, and uncertain way." and "...requires a long series of trials in order to distinguish between that which is nearly regular and that which is wholly accidental." An outstanding example is reported by Goldberg and Rorer (1965): 4050 trials of outcome feedback with respect to a judgment of psychosis vs. neurosis did not result in any increase in judgmental accuracy. Similarly, Deane, Hammond, and Summers (1971) found that judges did not learn a differentially weighted three-cue task when the function forms relating the cues to the criterion were inverse-U shaped and R_e was .92 with 200 trials of outcome feedback.

"Feedforward is knowledge about the task transmitted from the experimenter to the subject." (Bjorkman, 1971, p. 8), and is typically in the form of ecological validities or relative weights (Hoffman, 1960) which reflect the per cent of predictable criterion variance

accounted for by each cue. Magnusson and Nystedt (1969, p. 4) used judges who "... had taken part in two courses in test theory and were well acquainted with the concepts of correlation and prediction." These judges were asked to predict achievement scores for 30 subjects in an introductory psychology course on the basis of six cues with an R_e of .71. Three groups of four judges made all predictions twice: the first time with only the cue values, the second time with one group receiving ecological validities (feedforward), the second group receiving ecological validities and their own utilization coefficients (lens model feedback), and the third group served as a control. Note that lens model feedback is also feedforward on subsequent trials, but also permits the judge to compare his weights with the environmental weights. The lens model feedback group and the feedforward group did not differ from each other with respect to judgmental accuracy, and both groups were superior to the control group.

Newton (1965) used naive judges to predict grade point averages for 53 subjects on the basis of four cues with an R_e of .75. The general results were the same as those obtained in the Magnusson and Nystedt (1969) study: feedforward was sufficient to improve judgmental accuracy, and there was no difference between the group that received feedforward and the group that received lens model feedback.

However, there was an interval of three weeks between the first and second session for Newton's judges and two days for Magnusson and Nystedt's judges. Deane, Hammond, and Summers (1971) found that feedforward accompanied by outcome feedback was not so effective as lens model feedback presented immediately after each block of ten trials. In their task, there were three cues which were differentially related to the criterion in an inverse-U shaped manner, and R_e was .92. Judges in the lens model feedback group approximated the statistical limit of achievement, R_e , after 40 trials, judges presented with feedforward and outcome feedback attained an r_a of .57 after 200 trials, while judges with outcome feedback did not learn the task.

Summers, Taliaferro, and Fletcher (1970) found that feedforward accompanied by outcome feedback was more effective than outcome feedback alone. Todd and Hammond (1965) found that lens model feedback was more effective than outcome feedback, and that lens model feedback accompanied by outcome feedback was no more effective than lens model feedback alone.

In an attempt to enhance the efficacy of outcome feedback, Slovic, Rosov, Hoffman, Goetsch, Earle, and Geiseman (1972) replicated the Deane et. al. study, and observed four groups of judges under different feedback conditions. One group of judges received standard outcome

feedback, and a second group received outcome feedback with the cues ordered (five different orders) in such a manner as to make more clear to the judge the relationships of the cues to the criterion. For example, only one level of a specific cue was varied with the levels of the two other cues held constant for a specified number of trials. A third group could select the three cue values on each trial; in effect, judges could generate hypotheses with respect to the cue-criterion relationships, select cue values to test their hypotheses, and immediately receive outcome feedback. A fourth group received lens model feedback.

The average r_a for the group which received outcome feedback was .20; the group with outcome feedback ordered by the experimenter obtained an average r_a of .09, and the group which selected the cue values attained an average r_a of .11. These correlations reflect virtually no learning. Judges in the lens model feedback group obtained an average r_a of .72.

Despite the results of Slovic et. al.'s training procedure, one practical reason for continued attempts to train judges to use outcome feedback is that outcome feedback may be the only information available to a judge in an applied setting. Such outcome feedback may be degraded insofar as the cues may be unreliable and outcomes may be available for only some of the cases, perhaps after

long time intervals. Also, the same cues may not be available for all cases.

Transfer of Training

Transfer effects are generally not evaluated, since there are apparently no studies in which the same judges are trained on more than one task. The interpersonal learning (IPL) paradigm (Brehmer, 1971; Earle, 1970; Hammond, Wilkins, and Todd, 1966) is sometimes explicitly referred to as a transfer paradigm ("...the specification of the initial cognitive systems of the subjects not only shows what the subjects can learn from each other, it also shows what each subject might transfer to the IPL task." (Brehmer, 1971, p. 2)), but does not provide unequivocal information with respect to transfer effects.

The IPL paradigm is characterized by two stages, a training stage and an IPL stage. In the training stage, judges are trained to rely differentially on cues with respect to the cues' ecological validities and/or function forms. In the IPL stage, the relationship of the cues to the criterion changes, creating, in effect, a new task for both judges. The judges are generally not, however, informed that the task has changed. Each judge must now learn the new relationships of the cues to the criterion, the relationships of the cues to the other judge's predictions of the criterion, and decide whether to change

his policy on the basis of outcome feedback only from the task, outcome feedback from the other judge, or a verbal description of the other judge's policy. Consequently, even though the IPL may be considered a transfer task, it is difficult to disentangle changes specific to learning from those due to other factors.

The Present Research

There are no unequivocal data available with respect to changes in a judge's performance over different tasks. The primary purpose of this research was to investigate whether performance would change as a function of the number of previous problems to which a judge had been exposed and the type of feedback the judge had received. A determination of whether lens model feedback affected the judge's ability to utilize outcome feedback more effectively was an auxiliary goal of the research.

Specifically, information was obtained with respect to the following questions:

1. What are the differential effects of feedforward, outcome feedback, and lens model feedback?
2. Do repeated training sessions with lens model feedback result in learning to learn?
3. Does lens model training transfer to tasks in which outcome feedback only is available?

METHOD

MCPL Paradigm

A three cue probability learning paradigm in which judges are trained to produce the same decisions as a hypothetical, or target, judge was used (Summers, Taliaferro, and Fletcher, 1970). A content-free task was chosen to avoid pre-existing differences in judges' utilization coefficients, and to avoid the problem of unrepresentative combinations of cue values which would occur with an explicitly defined content area, given uncorrelated cues.

Definitions of Terms

Task. Eight tasks, which varied with respect to cue and criterion standard deviations, ecological validities, and R_e were used. A specific task is identified by the formal, statistical characteristics listed in Table 1. Judges were trained on the tasks in different orders.

Problem. The word problem specifies the order of the task on which a judge was trained. That is, Problem 1 refers to the first task on which a judge was trained; Problem 1 could be any of the eight tasks listed in Table 1. The order of task presentation for each judge is listed in Table 2.

Trial. On each trial, the judge observed the three cue values, and made a prediction of the target's

Table 1
Task Characteristics

Task Number	r_{1e}	r_{2e}	r_{33}	$SD_{X_{1e}}$	$SD_{X_{2e}}$	$SD_{X_{3e}}$	SD_{Y_e}	R_e
1	.50	.32	.22	5	5	5	15	.63
2	.32	.32	.59	5	5	15	15	.74
3	.55	.22	.50	15	15	5	30	.78
4	.45	.00	.59	15	5	15	15	.74
5	.22	.59	.00	15	15	15	30	.63
6	.45	.39	.55	15	5	5	30	.81
7	.55	.45	.50	5	15	5	15	.87
8	.00	.39	.39	5	15	15	30	.55

Note.--All cues had a mean of 50, and the criterion mean was 200.

Table 2
General Design

Group	Judge	Problem							
		1	2	3	4	5	6	7	8
LMF1	1	1	2	3	4	5	6	7	8
	2	2	3	4	5	6	7	8	1
	3	3	4	5	6	7	8	1	2
	4	4	5	6	7	8	1	2	3
LMF2	5	5	6	7	8	1	2	3	4
	6	6	7	8	1	2	3	4	5
	7	7	8	1	2	3	4	5	6
	8	8	1	2	3	4	5	6	7
FC	9	1	2	3	4	5			
	10	2	3	4	5	6			
	11	3	4	5	6	7			
	12	4	5	6	7	8			
OC	13	1	2	3	4	5			
	14	2	3	4	5	6			
	15	3	4	5	6	7			
	16	4	5	6	7	8			

Note.--The numbers refer to specific tasks, the characteristics of which are specified in Table 1.

judgment. The target's judgments were constructed to have a mean of 200, and a standard deviation of either fifteen or thirty, depending on the specific task. Outcome feedback was provided immediately after the judge's prediction of the target's judgment.

Session. Each problem consisted of two sessions of 100 trials on a given task, presented on successive days. All judges received outcome feedback during Session 1. Judges received differential feedback during Session 2, depending on the group to which they were assigned.

Subjects

Sixteen female university sophomores, who had not had a course in statistics, served as subjects. Subjects were paid \$1.60 an hour, and those enrolled in the Introductory Psychology course also received two hours of experimental credit. There were eight judges in the lens model feedback group (LMF), four judges in the feed-forward control group (FC), and four judges in the outcome control group (OC). An additional OC judge's responses were not used because she looked at the criterion values before she made her judgments.

Apparatus

Each trial was printed on an IBM card with the three cues and the criterion value on the same card. Each card

was presented in an apparatus which permitted the judge to observe the three cues, record her judgment directly on the card, and immediately receive outcome feedback (Figure 2).

Procedure

All judges received outcome feedback during Session 1 and Session 2 of each problem. Differences between groups consisted of differential feedback at the beginning of Session 2.

Lens model feedback group. Each judge in the LMF group received her three utilization coefficients squared, which were computed from the 100 trials of Session 1. Each LMF judge also received relative weights, her overall agreement with the target (r_a^2), and the statistical limit of achievement (R_e), which served as an index of task difficulty. On the first problem only, judges received the squared utilization coefficients as feedback, and R_e^2 as the statistical limit of achievement. In order to encourage optimizing strategies, as opposed to matching strategies, all subsequent problems utilized the feedback described above.

• Feedforward control group. Judges in the FC group received the same information as judges in the LMF group, with the exception of utilization coefficient feedback.

Outcome feedback control group. Each judge in the

Figure 2. Apparatus

OC group received r_a^2 and R_e primarily for motivational reasons, since these indices provide no specific information with respect to the target's policy.

LMF judges were trained on eight tasks (sixteen sessions), and FC and OC judges were trained on five tasks (ten sessions).

Each judge was assigned to an individual cubicle, and appeared at the same time of day for each session. Judges were trained on two tasks (four sessions) a week on Mondays through Thursdays.

The instructions for the three groups are presented in Appendix A, and the feedback forms are shown in Appendix B. (In the instructions, and on the feedback forms, the word model is used instead of target.) Judges were informed that there would be two sessions for each task, and were told each time a new task was introduced.

There were twenty practice trials, and 100 test trials during each session. Consequently, each judge in the LMF group made a total of 1920 judgments, and each judge in the FC and OC groups made a total of 1200 judgments.

After each session, the judgments were punched directly on the IBM cards which also contained the three cue values and the criterion value. A computer program then yielded an individual correlation matrix between the cue values, the criterion values, and the judgments,

for each judge's performance during the session. The correlation matrix permitted all information contained on the feedback forms, shown in Appendix B, to be obtained with the aid of a hand calculator.

Task Characteristics and Generation of Tasks

The relationship of the cues to the criterion was positive and linear, and the cues were not intercorrelated. It was deemed necessary to have a set of relatively heterogeneous tasks. The tasks were created to differ with respect to r_{1e} , R_e , and cue and criterion standard deviations. There were eight different ecological validities, such that each cue accounted for from zero per cent to thirty-five per cent of the criterion variance, in increments of five per cent. The standard deviation of each cue was either five or fifteen, and the standard deviation of the criterion was either fifteen or thirty. The levels of ecological validities, and the levels of cue and criterion standard deviations, were randomly assigned to the eight tasks with the sole restriction that each level appear the same number of times. The statistical characteristics of the eight tasks are presented in Table 1.

The cue and criterion values for each task were generated by using the following equation:

$$Z_{\text{criterion}_1} = r_{1e} f_{11} + r_{2e} f_{21} + r_{3e} f_{31} + (1 - R_e^2)^{1/2} f_{41} \quad (4)$$

This equation yields cue-criterion matrices with the desired statistical characteristics if $f_1 \dots f_4$ are normally distributed, uncorrelated, and in standard form. A random number generator produced sets of four z scores that were normally distributed, but correlated. Consequently, for each session of each task, factor score coefficients were obtained and applied to the generated z scores to yield factor scores (f's) which were uncorrelated and in standard form. The cue and criterion z scores obtained by this procedure were transformed to have the means and standard deviations listed in Table 1.

Design

There were two sets of analyses performed: a between-groups analysis designed to reveal the effects of differential feedback, and a within-groups analysis designed to assess more completely changes over problems.

Between-groups analysis. Three groups of four judges each were trained on five problems. The LMF1 (a subset of the LMF group), FC, and OC judges were trained on the tasks listed in Table 2. Although each judge within a group received a different order of tasks, the order across groups was identical. The average R_e , from Problem 1 to Problem 5, was .73, .73, .75, .78, and .74.

Within-group analysis. The LMF group included the LMF1 judges, and three other judges (Table 2). Judge 5

did not learn. In a postexperimental interview, she reported using a strategy in which her predictions of the target's judgments were based on whether the majority of the cue values were odd or even numbers. Since the primary purpose of the analysis was to assess changes over problems, given that the feedback was effective, Judge 5's data were excluded from the analyses. The average R_e , from Problem 1 to Problem 8, was .75, .72, .71, .75, .75, .73, .73, and .73.

RESULTS

Dependent Measures

The dependent measures were r_a , R_s , and G , computed over the last 100 trials within each session. The analyses reported below were also performed on r_{ce} and r_{cs} . These are redundant indices ($r_{ce} = GR_s$ and $r_{cs} = GR_e$), and since the analyses of these indices shed no additional light on the judges' performance, the results of these analyses are not presented.

Since each judge was observed only intermittently, and since the apparatus permitted the judge to remove a card and deliberately look at the criterion value prior to recording her judgment, C coefficients were examined as an index of cheating. One judge in the OC group was replaced; no other judge had consistently significant C coefficients, and there were no differences in C values

as a function of groups, problems, or sessions.

The data for each judge are presented in Appendix C. Each correlation coefficient was transformed to Fisher's Z coefficient prior to the analyses reported below.

The Effects of Differential Feedback

Between-groups Analyses

The observations for the first set of analyses were obtained from the first five problems of the LMFl, FC, and OC groups (Table 2). A 3 (Groups) x 5 (Problems) x 2 (Sessions) ANOVA with repeated measures on the last two factors (Winer, 1962, p. 319) was performed on each of the values Z_a , Z_s , and Z_G . Z_a , Z_s , and Z_G are transformations of r_a , R_s , and G to Fisher's Z coefficient, respectively.

The differences between Session 1 and Session 2 as a function of type of feedback (Groups) are presented in this section. The effects of problems pertain to a distinctly separate issue, and will be presented in a subsequent section.

Z_a . The summary of the ANOVA is presented in Table 3. A significant Sessions main effect and a Groups x Sessions interaction effect were obtained. An ANOVA of the simple main effects (differences between Session 1

Table 3
Summary of ANOVAS
(Between-groups Analyses)

Source	df	Z_a		Z_s		Z_G	
		MS	F	MS	F	MS	F
<u>Between subjects</u>	<u>11</u>						
A (Groups)	2	.027	.82	.405	1.90	.007	.01
Subj w. groups	9	.033		.213		.706	
<u>Within subjects</u>	<u>108</u>						
B (Problems)	4	.132	2.81*	.640	4.10**	.138	.13
AB	8	.056	1.19	.285	1.83	.765	.72
B x subj w. groups	36	.047		.156		1.068	
C (Sessions)	1	.310	16.32**	.865	11.09**	5.639	14.84**
AC	2	.101	5.32**	.250	3.21	1.515	3.99
C x subj w. groups	9	.019		.078		.380	
BC	4	.020	1.43	.107	1.62	.409	1.65
ABC	8	.014	1.00	.029	.44	.534	2.15
BC x subj w. groups	36	.014		.066		.248	

* $p < .05$

** $p < .01$

and Session 2 for each group separately) was performed; a summary of the ANOVA is presented in Table 4, and the means obtained from the Groups x Sessions summary tables are presented in Table 5 (Winer, 1962, p. 323).

There was an increase in Z_a from Session 1 to Session 2 for the LMFl and FC groups, but no change for the OC group. While the differences were not assessed statistically, every judge's mean in Session 2 was greater than the mean Z_a in Session 1 for the LMFl and FC groups; this was true of only one judge in the OC group (Judge 16).

Z_s . The analyses performed on Z_a were also performed on Z_s , and the results of the analyses are also presented in Tables 3, 4, and 5. There was a significant Sessions main effect, but the Groups x Sessions interaction effect only approached significance at the .10 level. However, the ANOVA for simple main effects revealed that Z_s increased from Session 1 to Session 2 for judges in the LMFl and FC groups, but did not increase for judges in the OC group. This obtained for all judges except Judge 2 and Judge 8 in the LMFl and FC groups, while only Judge 16's mean Z_s was higher during Session 2 than Session 1 among the OC group judges. Performing an analysis of simple effects in the absence of a significant interaction increases the probability of Type 1 error, but permits a more complete description of the data.

Table 4
Summary of ANOVAS for Simple Effects
(Differences between Sessions: Between-groups Analyses)

		Z _a		Z _s		Z _G	
Source	df	MS	F	MS	F	MS	F
Simple effects of sessions:							
For LMFl	1	.235	12.37**	.546	7.00*	5.218	13.73**
For FC	1	.274	14.42**	.818	10.49*	3.450	9.08*
For OC	1	.002	.10	.001	.01	.001	.00
Sessions x subj w. groups	9	.019		.078		.380	

* p < .05

** p < .01

Table 5
Groups x Sessions Summary Tables
(Between-groups Analyses)

Group	Z_a			Z_s			Z_G		
	Session 1	Session 2	S2-S1	Session 1	Session 2	S2-S1	Session 1	Session 2	
LMF1	.49	.64	.15	.99	1.23	.24	1.38	2.10	.72
FC	.44	.60	.16	.83	1.11	.28	1.42	2.01	.59
OC	.58	.56	-.02	1.17	1.16	.01	1.74	1.73	- .01

Z_G. The results of the analyses on Z_G are also presented in Tables 3, 4, and 5. The overall results are similar to the results obtained with Z_S: there was a significant Sessions main effect, and the Groups x Sessions interaction effect approached significance at the .10 level. Again, Z_G increased from Session 1 to Session 2 for the feedback groups (LMF1 and FC), but not for the OC group. Each judge in the feedback groups had a higher mean Z_G in Session 2 than in Session 1; all judges in the OC group, except Judge 15, had lower mean Z_Gs in Session 2 than in Session 1.

Within-group Analyses

The analysis of the entire LMF group consisted of a 7 (Judges) x 8 (Problems) x 2 (Sessions) ANOVA with repeated measures on the last two factors (Winer, 1962, p. 368). One judge in the LMF group, Judge 5, did not learn. Her data are presented in Appendix C, but were not included in the analyses.

Table 6 (summary of ANOVAs) and Table 7 (Problems x Sessions summary tables) reveal that performance during Session 2 was greater than performance during Session 1 for all measures. Judge 2 had a lower mean R_S for Session 2 than for Session 1; no other judge had a lower mean, on any measure, for Session 2 than for Session 1.

Table 6
Summary of ANOVAS
(Within-group Analyses: LMF Group)

Source	df	Z_a		Z_e		Z_G	
		MS	F	MS	F	MS	F
<u>Between subjects</u>	<u>6</u>						
Subjects	6	.135		1.180		1.116	
<u>Within subjects</u>	<u>105</u>						
A (Problems)	7	.159	3.61**	1.028	9.02**	.628	.90
Linear	1	.602	9.26*	4.057	54.82**	1.337	1.50
A x subj	42	.044		.114		.701	
Linear	6	.065		.074		.892	
B (Sessions)	1	.356	44.50**	1.320	30.70**	9.389	17.62**
B x subj	6	.008		.043		.533	
AB	7	.014	1.08	.104	1.32	.631	1.37
AB x subj	42	.013		.079		.461	

* $p < .05$
** $p < .01$

Table 7
Problems x Sessions Summary Tables
(Within-group Analyses: LMF Group)

Problem	Z_a			Z_s			Z_G		
	Session 1	Session 2	S2-S1	Session 1	Session 2	S2-S1	Session 1	Session 2	S2-S1
1	.26	.47	.21	.40	.85	.45	1.41	1.70	.29
2	.41	.60	.19	1.05	1.38	.33	1.05	2.03	.98
3	.51	.63	.12	.97	1.34	.37	1.53	2.08	.55
4	.61	.73	.12	1.16	1.44	.28	1.58	2.13	.55
5	.64	.73	.09	1.46	1.39	-.07	1.60	2.00	.40
6	.59	.68	.09	1.32	1.46	.14	1.44	2.87	1.43
7	.65	.66	.01	1.48	1.53	.05	1.86	2.16	.30
8	.58	.66	.08	1.22	1.39	.17	1.59	1.74	.15

Summary of the Effects of Differential Feedback

Performance, as reflected by Z_a , Z_s , and Z_G , increased from Session 1 to Session 2 for the LMF and FC groups, but did not change for the OC group. Although the differences were not significant, the means in Table 5 suggest that the groups were not homogeneous. The differences among the groups during Session 1 are explicable only on the basis of random assignment. Although the mean performance of judges in the OC group seemed greater than the mean performance of judges in the LMF1 and FC groups during Session 1, there was no change over sessions for judges in the OC group.

Changes Over Problems

Learning to Use Lens Model Feedback

Between-groups analyses. There was a significant Problems main effect for Z_a and Z_s (Table 3). The Groups x Problems interaction effect was not significant; however, an ANOVA of simple main effects revealed that the Problems effect was significant only for the LMF1 group (Table 9). The linear component of the Problems main effect for the LMF1 group was significant, reflecting an increase over problems for Z_a and Z_s .

If lens model feedback were to become more effective as a function of previous problems, then the difference

between Session 1 and Session 2 should increase over problems. This clearly did not occur: there was no significant Problems x Sessions interaction effect for either Z_a , Z_s , or Z_G . The differences between Session 1 and Session 2 as a function of problems are presented in Table 8. The magnitude of the S_2-S_1 values did not increase across problems.

Within-group analyses. There was a significant Problems main effect for Z_a and Z_s (Table 6), and the linear component was also significant. There were no significant Problems x Sessions interaction effects: the S_2-S_1 values presented in Table 7 did not increase as a function of problems.

The basic conclusion is that learning to use lens model feedback did not occur: there was no indication of increased effectiveness of feedback as a function of problems.

Learning to Use Outcome Feedback

The significant Problems main effect for the LMF group, and the absence of the Problems x Sessions interaction effects (Table 6) for Z_a and Z_s reflected that judges utilized outcome feedback more effectively over problems. The increases were greatest for the first four problems (Table 7). The increase in Z_a was a function only of an increase in Z_s , since there was no significant Problems main effect for Z_G , and the mean R_e was virtually

Table 8
Groups x Problems Summary Tables
(Between-groups Analyses)

Measure	Problem	LMF	S2-S1	FC	S2-S1	OC	S2-S1
Z_a	1	.34	.31	.45	.24	.59	-.06
	2	.49	.21	.54	.13	.58	-.07
	3	.61	.09	.51	.27	.51	.09
	4	.72	.10	.61	.11	.66	.05
	5	.65	.06	.50	.08	.51	-.08
Z_s	1	.61	.59	.79	.32	1.07	.05
	2	1.06	.18	1.06	.43	1.33	-.08
	3	1.11	.65	1.01	.31	.98	.15
	4	1.28	.13	1.34	.23	1.30	.01
	5	1.50	-.03	.86	.13	1.16	-.19
Z_G	1	1.74	.65	1.43	.31	2.06	-.46
	2	1.72	1.17	1.69	-.17	2.01	-.51
	3	1.80	.88	1.57	.86	1.45	.26
	4	1.89	.76	1.61	.61	1.71	.76
	5	1.55	.14	2.28	1.32	1.46	-.10

Table 9
Summary of ANOVAS for Simple Effects
(Differences among Problems: Between-groups Analyses)

		Z_a		Z_s		Z_G	
Source	df	MS	F	MS	F	MS	F
Simple effects of problems:							
For LMF1	4	.184	3.91*	.871	5.58**	.124	.33
Linear	1	.595	8.38*	3.223	13.21**	.038	.05
For FC	4	.029	.62	.162	1.04	.867	2.28
Linear	1	.026	.37	.037	.15	2.085	2.55
For OC	4	.032	.68	.178	1.14	.677	1.78
Linear	1	.004	.06	.019	.08	1.804	2.21
Problems x subj							
w. groups	36	.047		.156		.380	
Linear	9	.071		.224		.817	

* p < .05

** p < .01

constant across problems.

There was no significant Problems main effect for the FC and OC groups (Table 9).

The Effects of Cue Standard Deviations

Although the present study was not designed to assess the effects of cue standard deviations, it was apparent that judges' predictions were related to the standard deviations of the cues. Statistically, there should be no relationship between these variables. The correlations of cue standard deviations with the utilization coefficients of the judges, for the feedback groups, presented in Table 10, reveal that even after feedback, cue standard deviations were significantly related to the judges' utilization coefficients.

Summary of Results

With respect to the three questions stated in the introduction:

1. Lens model feedback and feedforward led to increased performance, as reflected by higher Z_a , Z_s , and Z_G values in Session 2 than in Session 1. There were no differences between feedforward and lens model feedback. There were no changes over sessions for judges in the OC group.

2. Judges did not use feedforward or lens model

Table 10
Correlations of Cue Standard Deviations and Ecological
Validities with Utilization Coefficients
for LMF and FC Groups (N=76)

	Session 1			Session 2		
	r_{1s}	r_{2s}	r_{3s}	r_{1s}	r_{2s}	r_{3s}
r_{1e}	.35**			.77**		
r_{2e}		.28**			.71**	
r_{3e}			.24**			.62**
SD_{X_1}	.40**			.35**		
SD_{X_2}		.47**			.64**	
SD_{X_3}			.39**			.32**

** $p < .01$

feedback more effectively as a function of the number of previous problems.

3. Performance, as measured by Z_a , increased as a function of previous problems during Session 1 for the LMF group; the increase in Z_a was due only to an increase in Z_s . Judges used outcome feedback more effectively as a function of problems by being more consistent.

Unexpectedly, cue standard deviations, even after feedforward and lens model feedback, had an effect on judges' utilization coefficients.

DISCUSSION

The hypothesis of major interest was not confirmed: judges did not learn to learn across the set of policy formation tasks used. The study did, however, show the effects usually associated with the use of feedback in MCPL tasks.

Differences among Groups

Feedforward and lens model feedback again were shown to be effective in their ability to control certain aspects of behavior. This was reflected by increased performance on all measures for the FC and LMF groups in Session 2, and no change for judges in the OC group. There were no differences between the LMFl and FC groups.

Although the feedback was effective, the magnitude

of the effect was less than that obtained with interactive graphic devices, such as Cognograph (Hammond, 1971). The feedback used in this study differed on several important dimensions from feedback with interactive graphic devices. For example, there was a twenty-four hour interval between Session 1 and Session 2, feedback was presented only once for each task, and the utilization coefficients were computed over blocks of 100 trials. Not only does Cognograph permit the feedback to be presented immediately after each block of ten trials, but the feedback indices including the cue-criterion function forms, are displayed graphically on a cathode ray tube. Judges also received outcome feedback in conjunction with feedforward and lens model feedback.

The MCPL task used in this study differed from previous MCPL tasks in that the cue standard deviations varied within a task, which may make it more difficult for judges to utilize lens model feedback efficiently. When cues are not on the same scale, the appropriate kind of lens model feedback is unclear. One alternative, supplying the judges with the regression coefficients, reduces the task to a mechanical arithmetic operation. Another alternative may be to train judges on tasks on which the ecological validities are constant and the standard deviations vary from task to task, and on tasks on which the standard deviations are constant, and the

ecological validities vary, providing feedback with respect to both sets of indices.

There were no differences between the LMFl and FC groups. The tasks were relatively easy, since the cues were uncorrelated and positively related to the criterion in a linear manner. It may not be necessary to provide the judge with his utilization coefficients in this type of task. It is also possible that judges did not accept the utilization coefficients as reflecting their weightings of the cues. The correlations of the cue standard deviations with the utilization coefficients suggest that the judges may have weighted the cues as if they were on the same scale.

Changes over Problems

Judges did not learn to use lens model feedback or feedforward more effectively as a function of problems. The only effect of lens model feedback on the use of outcome feedback was that judges became more consistent as a function of problems. There was no effect of feedforward or outcome feedback on subsequent use of outcome feedback.

There are no published data with which to compare these results. The tasks were relatively homogeneous, varying with respect to ecological validities and cue and criterion standard deviations only. It may be fruitful

to look for transfer effects with more sophisticated feedback devices, and with sets of tasks with more complexity than the strictly linear tasks used in the present study. For example, tasks with nonlinear-function forms should be evaluated. Although it seems essential, on the first task, for a judge to be informed of the function-forms when the relationships between the cues and the criterion are nonlinear, judges may subsequently entertain and test the hypothesis of nonlinearity with outcome feedback only.

Research pursuing the possibility that policy formation may transfer, under some conditions, seems essential; heterogeneity of judgmental conditions is characteristic of real life situations, and a consideration of potential changes across tasks is a step towards more representative judgmental research.

REFERENCES

REFERENCES

- Bjorkman, M. Feedforward and feedback as determiners of knowledge and policy: Notes on a neglected issue. Umea Psychological Reports, 1971, No. 50.
- Brehmer, B. Effects of cue validity and function form upon interpersonal learning to make inductive inferences. Umea Psychological Reports, Department of Psychology, University of Umea, No. 34, 1971.
- Brunswik, E. Representative design and probabilistic theory in a functional psychology. Psychological Review, 1955, 62, 193-217.
- Castellan, N. J., Jr. Comments on the "Lens Model" equation and the analysis of multiple cue judgment tasks. Indiana Mathematical Psychology Program, Report No. 71-1, 1971a.
- Castellan, N. J., Jr. The analysis of multiple criteria in multiple-cue judgment tasks. Indiana Mathematical Psychology Program. Report No. 71-6, 1971b.
- Deane, D. H., Hammond, K. R., and Summers, D. A. Acquisition and application of knowledge in complex inference tasks. University of Colorado, Institute of Behavioral Science, Program on Human Judgment and Social Interaction, Report No. 130, 1971.
- Dudycha, L. W., and Naylor, J. C. Characteristics of the human inference process in complex choice behavior

- situations. Organizational Behavior and Human Performance, 1966, 1, 110-128.
- Earle, T. C. Task learning, interpersonal learning, and cognitive complexity. Oregon Research Institute Research Bulletin, Vol. 10, No. 2, 1970.
- Goldberg, L. R., and Rorer, L. G. Learning clinical inference: The results of intensive training on clinicians' ability to diagnose psychosis vs. neurosis from the MMPI. Paper presented at the meeting of the Western Psychological Association, Honolulu, June, 1965.
- Hammond, K. R. Probabilistic functionalism: Egon Brunswik's integration of the history, theory, and method of psychology. In K. R. Hammond (Ed.), The psychology of Egon Brunswik. New York: Holt, Rinehart and Winston, 1966, Pp. 15-80.
- Hammond, K. R. Computer graphics as an aid to learning. Science, 1971, 172, 903-908.
- Hammond, K. R., and Summers, D. A. Cognitive control. Psychological Review, 1972, 79, 58-67.
- Hammond, K. R., Wilkins, M. M., and Todd, F. J. A research paradigm for the study of interpersonal learning. Psychological Bulletin, 1966, 65, 221-232.
- Hoffman, P. J. The paramorphic representation of clinical judgment. Psychological Bulletin, 1960, 47, 116-131.
- Hursch, C., Hammond, K. R., and Hursch, J. L. Some

methodological considerations in multiple-cue probability studies. Psychological Review, 1964, 71, 42-60.

Magnusson, D., and Nystedt, L. Cue relevance and feedback in a clinical prediction task. Report from the Psychological Laboratories of the University of Stockholm, No. 272, 1969.

Newton, J. R. Judgment and feedback in a quasi-clinical situation. Journal of Personality and Social Psychology, 1965, 1, 336-342.

Rozeboom, W. Comments on Professor Hammond's paper. In Royce, J., and Rozeboom, W. (Eds.), Psychology of knowing. London: Gordon & Breach, 1971.

Slovic, P., and Lichtenstein, S. Comparison of Bayesian and regression approaches to the study of information processing in judgment. Organizational Behavior and Human Performance, 1971, 6, 649-744.

Slovic, P., Rosov, B., Hoffman, P., Goetsch, J., Earle, T., and Geiseman, R. Effects of feedback in a multiple-cue learning study. Paper presented at the Boulder Conference on Human Judgment, Boulder, March, 1972.

Summers, D. A., Taliaferro, J. D., and Fletcher, D. J. Judgment policy and interpersonal learning. Behavioral Science, 1970, 15, 514-521.

Todd, F. J., and Hammond, K. R. Differential feedback in

two multiple-cue probability learning tasks.

Behavioral Science, 1965, 10, 429-435.

Tucker, L. R. A suggested alternative formulation in the developments by Hursch, Hammond, & Hursch, and by Hammond, Hursch & Todd. Psychological Review, 1964, 71, 528-530.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

APPENDICES

APPENDIX A

LMF Group

Instructions

The general purpose of this research is to find out how people learn to use information to make judgments or decisions.

Your job will be to learn how another person, called the Model, uses information (cues) to make judgments. The judgments you will learn to make are related to how people learn to make important judgments in real life situations.

There will always be three cues and one judgment printed on each IBM card as pictured below.

Model's judgment	Cue 1	Cue 2	Cue 3
(Write your judgment here)			

You will be shown the three cues and the judgment that the Model has made on the first card. Try to find out how the Model has used the cues to make his judgment.

After the first card, you will see only the three cues. Write down the judgment that you think the Model has made in the space on the card outlined above. When you remove the card from the box, you will see the Model's actual judgment.

It is very important that you see only the cues before you make your judgment. If, for any reason, you happen to see the Model's judgment before you have made your own judgment, write down a 0 instead of a judgment.

After you have made a judgment, turn the card over, and don't look at it again. You should only be considering one card at a time. Please keep the cards your judgments are one in the same order that you got them.

These tasks are very difficult: it is not possible to be completely accurate. Even a computer program written to make the best judgment will rarely make exactly the same judgment as the Model. The basic reason for this is that people are not perfectly consistent, and the Model may be making errors in using the three cues.

Although the Model may not be using the cues in a perfectly consistent way, the degree to which he uses each cue can be determined. In the example below, the Model is using Cue 1 the most and Cue 3 the least in making his judgments. The Unknown portion in the example reflects that the Model's judgments are not perfectly related to the three cues.

Cue 1	40%
Cue 2	20%
Cue 3	10%
Unknown	30%

There will be two sessions for each Model whose judgments you will be learning. In the next session, you will see how the Model uses the cues. You will also be shown a similar representation of how you have used the cues. This will permit you to compare how you have used the cues with how the Model uses the cues. You will also be given an overall measure of how well you have succeeded in making the same judgments as the Model. The highest score possible is 100.

People are sometimes suspicious in psychological research projects, and are concerned that they are not told the real purpose of the study. This study is concerned only with training people to use information: it is not related to personality or intellectual assessment in any way whatsoever. Some students will get different types of information in order to find out what kinds of information are best for learning.

Please let me know if you have any questions.

FC Group

Instructions

The general purpose of this research is to find out how people learn to use information to make judgments or decisions.

Your job will be to learn how another person, called the Model, uses information (cues) to make judgments. The judgments you will learn to make are related to how people learn to make important judgments in real life situations.

There will always be three cues and one judgment printed on each IBM card as pictured below.

Model's judgment	Cue 1	Cue 2	Cue 3
	(Write your judgment here)		

You will be shown the three cues and the judgment that the Model has made on the first card. Try to find out how the Model has used the cues to make his judgment.

After the first card, you will see only the three cues. Write down the judgment that you think the Model has made in the space on the card outlined above. When you remove the card from the box, you will see the Model's actual judgment.

It is very important that you see only the cues before you make your judgment. If, for any reason, you happen to see the Model's judgment before you have made your own judgment, write down a 0 instead of a judgment.

After you have made a judgment, turn the card over, and don't look at it again. You should only be considering one card at a time. Please keep the cards your judgments are on in the same order that you got them.

These tasks are very difficult: it is not possible to be completely accurate. Even a computer program written to make the best judgment will rarely make exactly the same judgment as the Model. The basic reason for this is that people are not perfectly consistent, and the Model may be making errors in using the three cues.

Although the Model may not be using the cues in a perfectly consistent way, the degree to which he uses each cue can be determined. In the example below, the Model is using Cue 1 the most and Cue 3 the least in making his judgments. The Unknown portion in the example reflects that the Model's judgments are not perfectly related to the three cues.

Cue 1	40%
Cue 2	20%
Cue 3	10%
Unknown	30%

There will be two sessions for each Model whose judgments you will be learning. In the next session, you will see how the Model uses the cues.

You will also be given an overall measure of how well you have succeeded in making the same judgments as the Model. The highest score possible is 100.

People are sometimes suspicious in psychological research projects, and are concerned that they are not told the real purpose of the study. This study is concerned only with training people to use information: it is not related to personality or intellectual assessment in any way whatsoever. Some students will get different types of information in order to find out what kinds of information are best for learning.

Please let me know if you have any questions.

OC Group

Instructions

The general purpose of this research is to find out how people learn to use information to make judgments or decisions.

Your job will be to learn how another person, called the Model, uses information (cues) to make judgments. The judgments you will learn to make are related to how people learn to make important judgments in real life situations.

There will always be three cues and one judgment printed on each IBM card as pictured below.

Model's judgment	Cue 1	Cue 2	Cue 3
(Write your judgment here)			

You will be shown the three cues and the judgment that the Model has made on the first card. Try to find out how the Model has used the cues to make his judgment.

After the first card, you will see only the three cues. Write down the judgment that you think the Model has made in the space on the card outlined above. When you remove the card from the box, you will see the Model's actual judgment.

It is very important that you see only the cues before you make your judgment. If, for any reason, you happen to see the Model's judgment before you have made your own judgment, write down a 0 instead of a judgment.

After you have made a judgment, turn the card over, and don't look at it again. You should only be considering one card at a time. Please keep the cards your judgments are on in the same order that you got them.

These tasks are very difficult: it is not possible to be completely accurate. Even a computer program written to make the best judgment will rarely make exactly the same judgment as the Model. The basic reason for this is that people are not perfectly consistent, and the Model may be making errors in using the three cues.

There will be two sessions for each Model whose judgments you will be learning.

You will be given an overall measure of how well you have succeeded in making the same judgments as the Model. The highest score possible is 100.

People are sometimes suspicious in psychological research projects, and are concerned that they are not told the real purpose of the study. This study is concerned only with training people to use information: it is not related to personality or intellectual assessment in any way whatsoever. Some students will get different types of information in order to find out what kinds of information are best for learning.

Please let me know if you have any questions.

APPENDIX B

LMF Group

Name _____

Date _____

Session _____

Room number _____

Task number _____

On this task, only _____ of the relationship can
be considered to be known. _____ is unknown.

The known relationship can be broken down as indicated
in the Model's judgments below.

	Model's judgments	Your judgments	Your judgments
Cue 1			
Cue 2			
Cue 3			
		First 100 cards	Second 100 cards
Overall agreement			

FC Group

Name _____

Date _____

Session _____

Room number _____

Task number _____

On this task, only _____ of the relationship can be considered to be known. _____ is unknown.

The known relationship can be broken down as indicated in the Model's judgments below.

	Model's judgments
Cue 1	<input type="text"/>
Cue 2	<input type="text"/>
Cue 3	<input type="text"/>

First 100 cards

Second 100 cards

Overall
agreement

OC Group

Name _____

Date _____

Session _____

Room number _____

Task number _____

First
100 cardsSecond
100 cardsOverall
agreement

APPENDIX C

The correlation coefficients are listed in the following order: r_a , R_s , G , C , r_{1s} , r_{2s} , r_{3s} . The first four numbers identify, respectively, the judge, the task, the problem, and the session.

4	4	1	2	0.556	0.767	0.987-0.010	0.359	0.006	0.678
4	5	2	1	0.401	0.575	0.940 0.081	0.218	0.497	0.190
4	5	2	2	0.574	0.924	0.996-0.028	0.278	0.880-0.053	
4	6	3	1	0.481	0.756	0.851-0.096	0.675	0.102	0.324
4	6	3	2	0.633	0.867	0.892 0.023	0.743	0.171	0.413
4	7	4	1	0.723	0.873	0.977-0.083	0.668	0.304	0.473
4	7	4	2	0.716	0.869	0.982-0.079	0.638	0.466	0.362
4	8	5	1	0.465	0.890	0.983-0.032	0.046	0.505	0.732
4	8	5	2	0.564	0.942	0.984 0.197	0.119	0.730	0.584
4	1	6	1	0.524	0.877	0.948 0.012	0.583	0.637	0.155
4	1	6	2	0.566	0.899	0.994 0.023	0.662	0.464	0.394
4	2	7	1	0.619	0.938	0.914-0.049	0.038	0.299	0.888
4	2	7	2	0.663	0.968	0.941-0.085	0.166	0.250	0.920
4	3	8	1	0.589	0.872	0.874-0.003	0.697	0.484	0.202
4	3	8	2	0.561	0.811	0.901-0.008	0.669	0.404	0.216
5	5	1	1	0.075	0.141	0.940-0.014	0.014	0.140	0.006
5	5	1	2	-0.010	0.025	0.596-0.025-	0.008	0.019	0.014
5	6	2	1	0.172	0.214	0.592 0.121	0.210-	0.024	0.032
5	6	2	2	0.187	0.226	0.960 0.020	0.158	0.122	0.105
5	7	3	1	0.276	0.246	0.832 0.206	0.231	0.079	0.028
5	7	3	2	0.100	0.157	0.903-0.044	0.046	0.124	0.084
5	8	4	1	0.106	0.195-	0.253 0.162	0.184-	0.013-	0.062
5	8	4	2	0.213	0.117	0.015 0.255	0.116	0.010-	0.006
5	1	5	1	0.209	0.251	0.944 0.081	0.169	0.093	0.160
5	1	5	2	0.093	0.151	0.878 0.013	0.120	0.012	0.091
5	2	6	1	0.088	0.213	0.925-0.086	0.129	0.129	0.110
5	2	6	2	0.327	0.324	0.729 0.239	0.027	0.299	0.123
5	3	7	1	0.192	0.258	0.709 0.082	0.015	0.023	0.257
5	3	7	2	0.151	0.259	0.930-0.057	0.133	0.014	0.222
5	4	8	1	0.264	0.275	0.950 0.109	0.098-	0.046	0.253
5	4	8	2	0.064	0.053	0.208 0.083	0.009	0.052	0.007
6	6	1	1	0.499	0.630	0.908 0.083	0.537	0.204	0.259
6	6	1	2	0.608	0.813	0.838 0.166	0.742	0.228	0.241
6	7	2	1	0.700	0.958	0.889-0.297	0.337	0.814	0.376
6	7	2	2	0.734	0.993	0.877-0.263	0.358	0.857	0.350
6	8	3	1	0.487	0.862	0.972 0.071	0.212	0.555	0.624
6	8	3	2	0.564	0.932	0.981 0.206	0.160	0.597	0.698
6	1	4	1	0.221	0.492	0.976-0.116	0.320	0.288	0.239
6	1	4	2	0.562	0.962	0.991-0.154	0.693	0.499	0.443
6	2	5	1	0.719	0.968	0.991 0.079	0.314	0.348	0.847
6	2	5	2	0.685	0.829	0.976 0.224	0.244	0.240	0.755
6	3	6	1	0.645	0.984	0.843 0.027	0.664	0.679	0.256
6	3	6	2	0.679	0.968	0.879 0.141	0.737	0.576	0.251
6	4	7	1	0.683	0.969	0.972-0.086	0.666	0.183	0.679
6	4	7	2	0.741	0.977	0.975 0.249	0.688	0.165	0.674
6	5	8	1	0.521	0.938	0.792 0.148	0.535	0.575	0.513
6	5	8	2	0.556	0.975	0.883 0.069	0.520	0.722	0.398
7	7	1	1	0.504	0.534	0.963 0.136	0.226	0.366	0.316
7	7	1	2	0.551	0.755	0.910-0.123	0.676	0.252	0.222
7	8	2	1	0.425	0.893	0.916-0.055	0.029	0.325	0.831
7	8	2	2	0.420	0.894	0.979-0.160	0.018	0.744	0.495

1	1	1	1	0.389	0.542	0.986	0.084	0.371	0.315	0.239
1	1	1	2	0.526	0.767	0.990	0.104	0.597	0.448	0.176
1	2	2	1	0.567	0.824	0.949	-0.020	0.119	0.261	0.772
1	2	2	2	0.537	0.812	0.992	-0.159	0.256	0.370	0.676
1	3	3	1	0.548	0.724	0.852	0.163	0.635	0.328	0.113
1	3	3	2	0.536	0.792	0.895	-0.027	0.668	0.378	0.194
1	4	4	1	0.584	0.795	0.909	0.120	0.659	0.173	0.410
1	4	4	2	0.634	0.854	0.990	0.025	0.599	0.050	0.606
1	5	5	1	0.534	0.841	0.880	0.130	0.675	0.500	0.043
1	5	5	2	0.503	0.822	0.948	0.023	0.523	0.634	-0.031
1	6	6	1	0.428	0.643	0.970	-0.165	0.471	0.287	0.331
1	6	6	2	0.581	0.843	0.841	0.022	0.763	0.121	0.338
1	7	7	1	0.647	0.817	0.948	-0.096	0.394	0.625	0.348
1	7	7	2	0.667	0.769	0.998	0.017	0.528	0.369	0.420
1	8	8	1	0.309	0.785	0.827	-0.086	0.452	0.423	0.482
1	8	8	2	0.571	0.856	0.994	0.240	0.067	0.636	0.569
2	2	1	1	0.273	0.331	0.968	0.058	0.066	0.117	0.303
2	2	1	2	0.383	0.676	0.962	-0.205	0.260	0.119	0.613
2	3	2	1	0.475	0.721	0.771	0.101	0.674	0.257	0.009
2	3	2	2	0.482	0.752	0.901	-0.099	0.658	0.313	0.187
2	4	3	1	0.591	0.820	0.912	0.097	0.411	0.328	0.629
2	4	3	2	0.541	0.791	0.968	-0.063	0.614	-0.054	0.496
2	5	4	1	0.405	0.844	0.830	-0.120	0.717	0.436	0.088
2	5	4	2	0.597	0.885	0.983	0.131	0.460	0.755	0.046
2	6	5	1	0.384	0.823	0.665	-0.170	0.811	0.129	0.052
2	6	5	2	0.504	0.857	0.693	0.077	0.838	0.025	0.177
2	7	6	1	0.594	0.888	0.790	-0.074	0.265	0.828	0.179
2	7	6	2	0.616	0.807	0.999	-0.266	0.504	0.398	0.489
2	8	7	1	0.450	0.909	0.959	-0.072	0.235	0.516	0.710
2	8	7	2	0.409	0.870	0.921	-0.074	0.330	0.588	0.549
2	1	6	1	0.497	0.737	0.985	0.083	0.554	0.312	0.373
2	1	8	2	0.510	0.802	0.961	0.062	0.493	0.451	0.444
3	3	1	1	0.131	0.120	0.774	0.094	0.083	0.085	0.015
3	3	1	2	0.345	0.645	0.815	-0.125	0.635	0.074	0.082
3	4	2	1	-0.032	0.813	-0.042	-0.017	-0.037	0.812	-0.008
3	4	2	2	0.536	0.690	0.946	0.108	0.522	-0.155	0.424
3	5	3	1	0.405	0.643	0.877	0.068	0.474	0.401	0.167
3	5	3	2	0.567	0.919	0.999	-0.045	0.321	0.861	0.032
3	6	4	1	0.594	0.832	0.797	0.181	0.012	0.244	0.795
3	6	4	2	0.640	0.878	0.938	-0.095	0.491	0.124	0.717
3	7	5	1	0.754	0.978	0.875	0.083	0.684	0.045	0.698
3	7	5	2	0.754	0.940	0.884	0.218	0.649	0.055	0.678
3	8	6	1	0.462	0.720	0.746	0.291	0.275	0.094	0.659
3	8	6	2	0.540	0.935	1.000	0.092	0.009	0.670	0.652
3	1	7	1	0.593	0.843	0.971	0.194	0.561	0.423	0.466
3	1	7	2	0.530	0.882	0.995	-0.048	0.680	0.412	0.381
3	2	8	1	0.505	0.706	0.984	-0.013	0.202	0.390	0.553
3	2	8	2	0.594	0.918	0.909	-0.098	0.134	0.146	0.896
4	4	1	1	0.159	0.188	0.908	0.050	0.050	0.042	0.176

7	1	3	1	0.503	0.795	0.970	0.045	0.704	0.211	0.303
7	1	3	2	0.572	0.899	0.971	0.078	0.811	0.362	0.142
7	2	4	1	0.680	0.885	0.951	0.192	0.229	0.175	0.837
7	2	4	2	0.674	0.898	0.958	0.117	0.259	0.189	0.839
7	3	5	1	0.504	0.809	0.865	-0.100	0.649	0.451	0.172
7	3	5	2	0.682	0.879	0.962	0.097	0.710	0.365	0.367
7	4	6	1	0.508	0.897	0.754	0.024	0.553	0.556	0.435
7	4	6	2	0.633	0.915	1.000	-0.163	0.541	-0.006	0.738
7	5	7	1	0.576	0.910	0.990	-0.026	0.410	0.804	0.113
7	5	7	2	0.595	0.946	0.989	0.014	0.215	0.920	-0.054
7	6	8	1	0.538	0.893	0.729	0.052	0.853	-0.003	0.264
7	6	8	2	0.645	0.889	0.858	0.104	0.663	-0.023	0.592
8	8	1	1	0.018	0.224	0.632	-0.072	0.176	0.100	0.095
8	8	1	2	0.002	0.244	0.712	-0.115	0.055	0.237	0.010
8	1	2	1	0.021	0.128	0.849	-0.060	0.055	0.112	0.028
8	1	2	2	0.412	0.686	0.878	0.063	0.673	0.053	0.122
8	2	3	1	0.230	0.535	0.924	-0.233	0.083	0.374	0.373
8	2	3	2	0.474	0.835	0.829	-0.110	0.053	0.012	0.833
8	3	4	1	0.491	0.861	0.748	-0.023	0.853	0.120	0.009
8	3	4	2	0.522	0.853	0.837	-0.087	0.834	0.031	0.174
8	4	5	1	0.472	0.738	0.852	0.014	0.123	0.213	0.696
8	4	5	2	0.620	0.851	0.998	-0.024	0.486	-0.048	0.697
8	5	6	1	0.524	0.723	0.959	0.142	0.111	0.714	-0.003
8	5	6	2	0.495	0.803	0.913	0.068	-0.023	0.795	0.114
8	6	7	1	0.375	0.810	0.668	-0.175	0.784	-0.054	0.195
8	6	7	2	0.334	0.709	0.513	0.095	0.707	-0.051	0.0
8	7	8	1	0.649	0.815	0.944	-0.075	0.286	0.494	0.582
8	7	8	2	0.579	0.771	0.928	-0.129	0.623	0.414	0.187
9	1	1	1	0.197	0.285	0.913	0.046	0.158	0.133	0.197
9	1	1	2	0.446	0.708	0.891	0.095	0.612	0.047	0.352
9	2	2	1	0.340	0.466	0.997	-0.003	0.198	0.166	0.388
9	2	2	2	0.565	0.916	0.862	-0.082	0.097	0.054	0.909
9	3	3	1	0.445	0.780	0.829	-0.141	0.482	0.573	0.220
9	3	3	2	0.586	0.881	0.883	-0.049	0.839	0.174	0.205
9	4	4	1	0.670	0.899	0.947	0.133	0.448	0.271	0.731
9	4	4	2	0.650	0.905	0.985	-0.036	0.417	-0.035	0.803
9	5	5	1	0.457	0.843	0.763	0.097	0.326	0.564	0.535
9	5	5	2	0.582	0.904	0.996	0.038	0.250	0.869	-0.010
10	2	1	1	0.527	0.782	0.901	0.020	0.063	0.150	0.765
10	2	1	2	0.576	0.815	0.919	0.050	0.184	0.097	0.788
10	3	2	1	0.421	0.770	0.674	0.048	0.392	0.657	0.083
10	3	2	2	0.483	0.818	0.849	-0.145	0.796	0.123	0.145
10	4	3	1	0.150	0.199	0.820	0.044	0.158	0.085	0.086
10	4	3	2	0.717	0.873	0.973	0.269	0.356	0.035	0.796
10	5	4	1	0.357	0.679	0.765	0.036	0.347	0.415	0.410
10	5	4	2	0.583	0.921	0.960	0.079	0.078	0.918	-0.003
10	6	5	1	0.365	0.465	0.938	-0.004	0.382	0.221	0.202
10	6	5	2	0.430	0.578	0.923	-0.004	0.463	0.114	0.327
11	3	1	1	0.191	0.507	0.826	-0.244	0.420	0.276	0.066
11	3	1	2	0.573	0.844	0.860	0.039	0.818	0.055	0.199
11	4	2	1	0.548	0.774	0.962	-0.009	0.288	0.065	0.716

11	4	2	2	0.695	0.949	0.977	0.041	0.717-0.039	0.620
11	5	3	1	0.395	0.893	0.678	0.008	0.786 0.300	0.298
11	5	3	2	0.440	0.828	0.964-0.149	0.082	0.824	0.005
11	6	4	1	0.408	0.661	0.908-0.171	0.543	0.302	0.227
11	6	4	2	0.556	0.774	0.919-0.054	0.644	0.217	0.371
11	7	5	1	0.487	0.523	0.893	0.191	0.304	0.413
11	7	5	2	0.512	0.573	0.988	0.057	0.324	0.368
12	4	1	1	0.306	0.545	0.860-0.072	0.367	0.254	0.312
12	4	1	2	0.446	0.501	0.968	0.150	0.393-0.010	0.310
12	5	2	1	0.440	0.676	0.790	0.167	0.404	0.362
12	5	2	2	0.375	0.511	0.928	0.113	0.348	0.374-0.008
12	6	3	1	0.429	0.620	0.881-0.025	0.543	0.127	0.272
12	6	3	2	0.486	0.640	0.985-0.053	0.421	0.213	0.432
12	7	4	1	0.542	0.758	0.709	0.231	0.259	0.712
12	7	4	2	0.521	0.699	0.933-0.116	0.413	0.524	0.210
12	8	5	1	0.410	0.679	0.980	0.077	0.113	0.406
12	8	5	2	0.442	0.716	1.000	0.084	0.009	0.507
13	1	1	1	0.410	0.688	0.922	0.024	0.377	0.358
13	1	1	2	0.351	0.639	0.954-0.049	0.367	0.440	0.282
13	2	2	1	0.570	0.822	1.000-0.090	0.351	0.348	0.657
13	2	2	2	0.551	0.816	0.997-0.138	0.339	0.292	0.682
13	3	3	1	0.543	0.831	0.844-0.001	0.643	0.502	0.160
13	3	3	2	0.549	0.863	0.778	0.097	0.452	0.697
13	4	4	1	0.638	0.866	0.904	0.174	0.617	0.318
13	4	4	2	0.541	0.758	0.964	0.001	0.545	0.155
13	5	5	1	0.373	0.815	0.796-0.109	0.505	0.484	0.419
13	5	5	2	0.403	0.850	0.792-0.056	0.477	0.538	0.454
14	2	1	1	0.521	0.693	0.994	0.030	0.348	0.304
14	2	1	2	0.487	0.673	0.987-0.013	0.191	0.270	0.586
14	3	2	1	0.399	0.763	0.744-0.099	0.537	0.540	0.052
14	3	2	2	0.250	0.638	0.794-0.288	0.481	0.413	0.071
14	4	3	1	0.552	0.864	0.914-0.096	0.600	0.309	0.539
14	4	3	2	0.567	0.821	0.953-0.032	0.630	0.152	0.504
14	5	4	1	0.384	0.714	0.768	0.053	0.434	0.406
14	5	4	2	0.443	0.740	0.740	0.186	0.498	0.395
14	6	5	1	0.366	0.636	0.868-0.172	0.566	0.203	0.206
14	6	5	2	0.366	0.697	0.689-0.054	0.669-0.039	0.190	
15	3	1	1	0.573	0.857	0.813	0.103	0.660	0.533
15	3	1	2	0.456	0.826	0.758-0.076	0.501	0.643	0.132
15	4	2	1	0.699	0.965	0.966	0.049	0.636	0.224
15	4	2	2	0.691	0.970	0.962	0.000	0.698	0.204
15	5	3	1	0.182	0.492	0.717-0.070	0.228	0.286	0.329
15	5	3	2	0.430	0.821	0.886-0.068	0.457	0.604	0.317
15	6	4	1	0.553	0.887	0.926	0.110	0.854	0.157
15	6	4	2	0.607	0.875	0.999	0.162	0.709	0.431
15	7	5	1	0.715	0.921	0.897-0.024	0.434	0.769	0.263
15	7	5	2	0.494	0.741	0.917-0.268	0.506	0.517	0.159
16	4	1	1	0.665	0.831	0.998	0.135	0.509	0.044
16	4	1	2	0.690	0.924	0.961	0.128	0.647	0.216
16	5	2	1	0.482	0.859	0.832	0.047	0.510	0.554
16	5	2	2	0.434	0.820	0.774	0.074	0.472	0.498

16	6	3	1	0.410	0.513	0.923	0.056	0.423	0.201	0.209
16	6	3	2	0.444	0.536	0.965	0.051	0.398	0.177	0.312
16	7	4	1	0.645	0.914	0.829	0.075	0.261	0.830	0.281
16	7	4	2	0.738	0.958	0.881	0.067	0.371	0.823	0.322
16	8	5	1	0.486	0.906	0.974	0.013	0.222	0.637	0.605
16	8	5	2	0.496	0.830	0.973	0.114	0.185	0.568	0.576