INFORMATION TO USERS

This material was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.

2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in "sectioning" the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.

4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from "photographs" if essential to the understanding of the dissertation. Silver prints of "photographs" may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.

5. PLEASE NOTE: Some pages may have indistinct print. Filmed as received.

Xerox University Microfilms

300 North Zeeb Road
Ann Arbor, Michigan 48106

75-3329

RUFFNER, John Wesley, 1948-
THE EFFECT OF GROUP DISCUSSION AND SHIFTING CUE VALIDITIES ON HUMAN INFERENCE BEHAVIOR IN A MEANINGFUL ENVIRONMENT.

Iowa State University, Ph.D., 1974
Psychology, industrial

Xerox University Microfilms, Ann Arbor, Michigan 48106

THIS DISSERTATION HAS BEEN MICROFILMED EXACTLY AS RECEIVED.

The effect of group discussion and
shifting cue validities on human inference
behavior in a meaningful environment

by

John Wesley Ruffner

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of
The Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major: Psychology

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1974

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

In recent years there has developed a marked resurgence of interest in the scientific investigation of the human inference process. This resurgence has been predicated by the fact that many individuals in organizational contexts often must learn to use information from various sources in deriving estimates or predictions about some future state of affairs. The credit manager's job necessitates that he combine information from sources such as average monthly debt and number of creditors to estimate an individual's credit risk. Another example might be that of admissions personnel seeking the best way to combine information from entrance tests and high school grade point averages to predict success in college. The present study follows in this line of investigation by extending a formal model of decision-making in meaningful environments to include several variables not systematically studied in previous research.

A large portion of these investigative efforts have drawn on both the theoretical and methodological formulations of Egon Brunswik's probabilistic functionalism (Brunswik, 1952). Brunswik's initial theoretical interests centered on the manner in which individuals were able to attain perceptual achievement in an uncertain environment. Adjustment to the environment necessitated that the individual be able to cope with stimuli or cues which have, through past experience, become probabilistically or equivocally associated with certain events, outcomes, or consequences.

Brunswik's conception of the process of perceptual achievement is best characterized by his formulation of the lens model. This model depicts a double convex lens in which "process details" emanating from an initial focus of distal stimuli give rise to a discernible pattern of proximal effects (or stimulus cues) on the various receptor surfaces of the organism. "Process details" from these cues in turn were said to converge to a terminal focus within the organism, thus mediating a perceptual response (Brunswik, 1952). Because the environment is "semierratic," the relations which exist between cues and objects possess only varying degrees of "probable applicability or validity." In accord with the probabilistic nature of cue-object relations, the organism faces the task of adjusting his utilization or subjective weighting of cues to best reflect their probabilistic relationships to an environmental referent (Brunswik, 1956). Figure 1 graphically depicts the elements of the lens model as well as the statistical relations (to be described in the following section) among the criterion, the stimulus cues, and the subject's response.

Within the lens model three important functional relationships have been identified. The first of these, ecological validity, reflects the probabilistic relationship

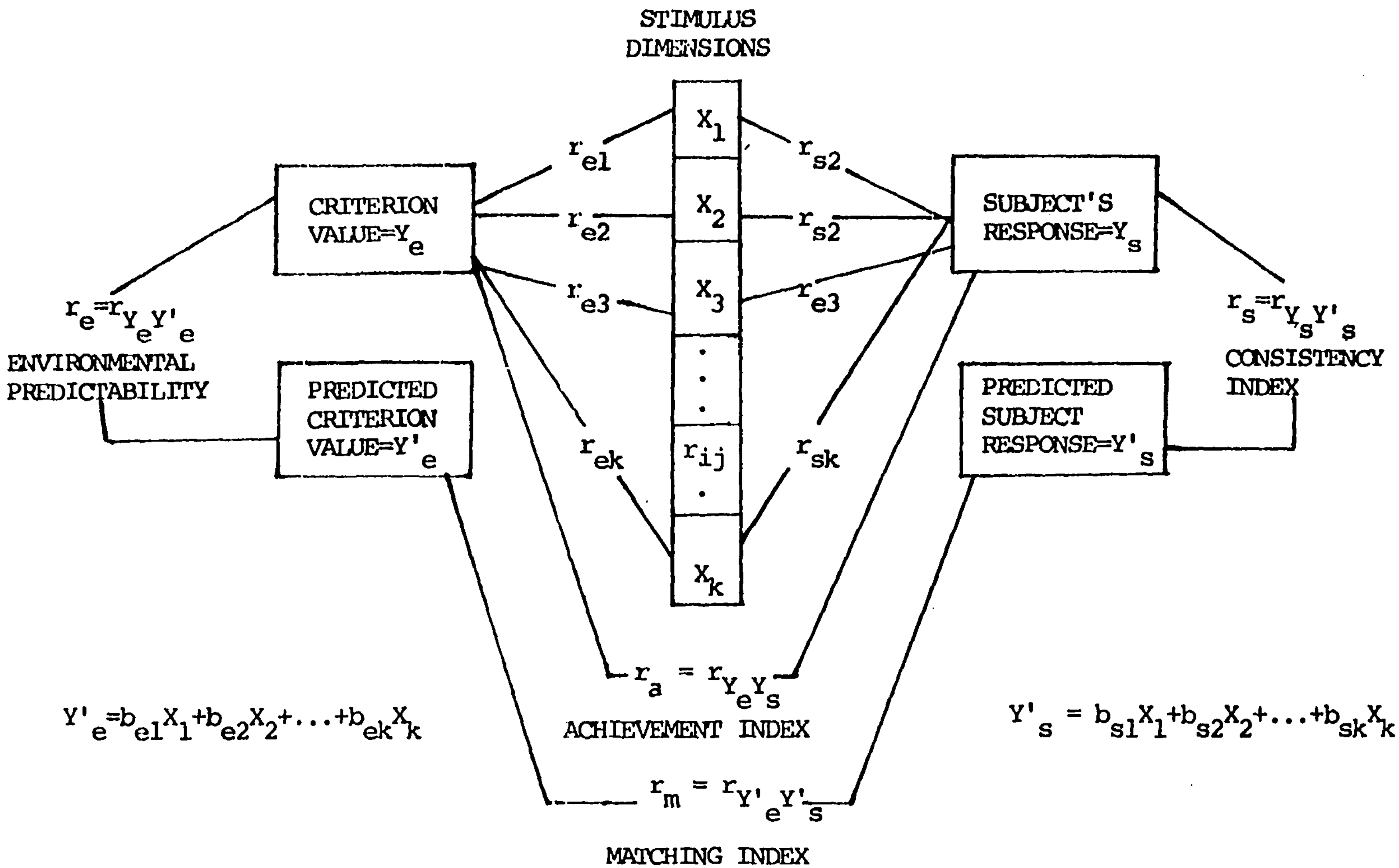

Figure 1. Diagram of the lens model showing relationships among cues, criterion, and subjects' responses (after Dudycha & Naylor, 1966).

existing between the distal variable (or criterion) and each of several proximal variables (or cues). The subject's relative weighting of these cues in accord with their probabilistic cue-object relationship is described as cue utilization. Lastly, the extent to which an individual's response approximates the value of the distal variable is taken as a measure of functional validity. In summary:

> Each over-all functional arc or achievement may be broken down into an extrasystemic and an intrasystemic constituent; these constituents have been called ecological validity and utilization, respectively. The general pattern of the mediational strategy of the perceptual system is predicated upon the limited ecological validity or trustworthiness of cues which we have observed in many contexts. . .The limitations in the dependability of single-cue variables force an uncertainty-geared probabilistic strategy upon perception. In order to improve the cognitive wager the perceptual system must accumulate and combine cues. (Brunswik, 1956, p. 140)

For Brunswik, the true aim of a proper psychology was to study the organism in its natural ecology. In order to do this he deemed it necessary to retreat from the confines of the classical form of systematic design. Instead, Brunswik maintained, the proper focus of psychology should be in "the employment of representative design to measure an individual's responses to representative samples of 'variate packages' from his environment which are otherwise left undisturbed" (Avant & Helson, 1973, p. 429). In other words, the proper study of an individual's adjustment to a probabilistic environment should entail a wide sampling over a natural ecological array of total stimulus situations

(Tolman, 1966).

Multiple Cue Probability Learning

Brunswik's original conception of the lens-model has since been elaborated and quantified in the study of human inference behavior. Drawing on Brunswik's notions that perception (or in the present case, inference behavior) relies on proper organismic adjustment to probabilistic relations between multiple proximal stimuli (or cues) and a distal variable (or criterion), recent investigators have refined the basic lens model and have derived statistical indices of performance. The format for multiple cue studies involves a learning task in which individuals are given one or several sources of information in the form of stimulus cues and are asked to make unitary predictions about the value of a criterion variable.

Within this basic framework there are several parameters of the multiple cue model which may be systematically studied. However, before describing these parameters, it is essential to discuss the quantitative indices formulated for the lens model. Brunswik's mathematical sophistication did not exceed the use of the correlation coefficient indexing the probabilistic relationships existing between proximal stimulus cues and distal criterion, between proximal stimulus cues and subject's perceptual response, and between distal criterion and subject's response (Brunswik, 1956).

These relationships have since been formalized within the context of multiple regression techniques (Hursch, Hammond, & Hursch, 1964; Tucker, 1964; Naylor & Schenck, 1964; Dudycha & Naylor, 1966; Castellan, 1973). All indices involve the three lens-model elements of : 1) the cue or stimulus dimensions ($X_1, X_2, \ldots, X_k$); 2) the criterion or distal variable (Ys); and 3) the subject's response or inference to the criterion variable (Ye).

The correlation existing between the criterion variable and the individual stimulus cues (r_{ei}) corresponds to Brunswik's concept of ecological validity. The index R^2e reflects the total amount of criterion variance accounted for by the combination of all probabilistic cues, or the level of system predictability.

Given many trial decisions, the least squares regression equations can be computed for both the subject

$$Y's = b_{s1}x_1 + b_{s2}x_2 + \ldots + b_{sk}x_k$$

and the environment

$$Y'e = b_{e1}x_1 + b_{e2}x_2 + \ldots + b_{ek}x_k$$

Dudycha and Naylor (1966) have described several performance indices which are derived by calculating all possible correlations among the observed and predicted criterion values and subject responses:

r_a = subject achievement, or the correlation between the subject's responses (Ys) and the true criterion values

(Ye);

r_s = subject consistency, or the correlation between the subject's responses (Ys) and their predicted responses (Y's);

r_m = subject matching, or the correlation between the subjects' predicted responses (Y's) and the predicted criterion values (Y'e).

The relationship existing between the components of the lens-model can be expressed in the equation:

$$r_a = r_e r_s r_m + C[(1-r^2_e)(1-r^2_s)]^{1/2}$$

Where C represents the correlation between the nonlinear variance in the environment and the nonlinear variance in the subject's estimates (Hursch, Hammond, and Hursch, 1964). When only a linear relationship exists between the criterion value and the cue value, this equation reduces to the form:

$$r_a = r_e r_s r_m.$$

Parameters of the Multiple Cue Model

Within the lens-model framework, the importance of several model parameters has been the subject of extensive investigation. The majority of studies have centered on the mathematical parameters inherent in the specification of the model. Included in this category are such parameters as cue-criterion correlations (r_{ei}), system predictability (R^2e), and cue intercorrelations (r_{ij}).

Two early studies focused on the effect of cue validity on subject response. Schenck and Naylor (1965) examined the effects of different ecological cue validities on subject response in a single cue environment. Their results indicated a direct linear relationship between cue validity and subject achievement. Extending the environment to include two cues, Dudycha and Naylor (1966) reported that cues with greater ecological validity yielded higher levels of achievement. More recent research (e.g., Naylor & Clark, 1968; Dudycha, Dumoff, & Dudycha, 1973; Brehmer, 1973a, 1973b) has supported this finding. However, as Hursch, Hammond, and Hursch (1964) and Brehmer (1973c) have pointed out, such a finding should be cautiously interpreted when system predictability is allowed to change as a direct function of cue validity (especially for the single cue environment). When cue validities and system predictability are allowed to covary systematically, their effects are completely confounded, thus yielding an equivocal interpretation with respect to subject achievement.

Results from studies by Uhl (1963) and Brehmer (1973c) indicated that higher levels of system predictability or task certainty result in improved cue consistency (Uhl, 1963) as well as achievement and matching (Brehmer, 1973c). However, since system predictability sets an upper limit on achievement, a higher level of achievement should be expected with high system predictability. In this context, Dudycha, Dudycha, and Schmitt (1974) have elaborated on the interpretative problems associated with the interrelations existing between cue intercorrelation (r_{ij}), cue validity (r_{ei}), and system predictability (R^2e).

More recently emphasis has been placed on the study of cue redundancy or cue intercorrelations. Naylor and Schenck (1968) manipulated cue intercorrelations and found that subject performance increased with higher levels of cue redundancy. Unfortunately, the effect of cue intercorrelation was confounded with cue validity in order to satisfy the requirement of constant system predictability. Knowles, Hammond, Stewert, and Summers, (1971) reported that negative cue intercorrelations served to impede learning (achievement) in contrast to positive or zero cue intercorrelation. In a more methodologically sound study, Schmitt and Dudycha (1973) held both cue validities and system predictability constant and examined the effect of positive and negative cue intercorrelations. The authors reported that differences in intercorrelations (positive versus negative) did not affect achievement. Their results with respect to consistency and matching were equivocal. In short, it appears that there is no strong evidence to indicate that subjects are able to successfully use redundant sources of information to any meaningful degree.

Nonmathematical Variables in Multiple Cue Probability Learning

Other investigators have, in contrast, focused their attention on certain nonmathematical variables which might have an influence on human inference behavior within the multiple cue framework. Among these variables are the number of cues in a prediction system, whether the cues and criterion values are accompanied by semantic labels, whether prediction takes place in a static versus dynamic environment, and the type of feedback given to subjects.

Number of Cues

The first of these variables concerns the number of cues or information sources available for use in the inference process. In general, it has been reported that increasing the number of cues in a profile past a certain point does not result in an increase in subject performance, but often in a reduction of decision performance (Hoffman & Blanchard, 1961; Oskamp, 1965; Einhorn, 1971; Conrad, 1973). However, studies employing only a single cue environment (e.g., Schenck & Naylor, 1965; Brehmer, 1973b; Dudycha, Dumoff, & Dudycha, 1973) are relatively sterile, in that *multiple* sources of information are normally utilized in real world decision making behavior. Accordingly, most multiple cue studies in recent years have employed an intermediate number of cues in the prediction system, usually from two to four orthogonal cues.

Cue and Criterion Labels

Very few of the reported multiple cue studies have employed semantic labels for the cues and criterion values and/or have systematically investigated the effects of such labels on human inference behavior. The most common approach has been to present cues and criterion values in the form of marks located on uncalibrated scales which are anchored at the extremes of the scales (e.g., Peterson, Hammond, & Summers, 1965; Hammond, Summers, & Deane, 1973). While being methodologically expedient, this method can be criticized on two accounts. First of all, this mode of presentation serves to confound information from graphic presentation with that from an interval scaled numeric display. It thus becomes difficult to interpret whether inferences are being made from graphic or numeric means or from a combination of the two sources. Furthermore, the external validity of this method of presentation can be called into question. Actual inferences are usually not made on the basis of this kind of display, in isolation of additional information (such as that provided by semantic labels). Other methods of presentation have utilized geometric cues and numeric criterion values (Todd & Hammond, 1965), binary digits as cues and criterion (e.g., Castellan, 1973b), both cues and criterion reprented on a circular dimension (Bolhuis-Bourma & Oostlander, 1972), and numeric cues and criterion values (e.g., Dudycha &

Naylor, 1966; Dudycha, Dumoff, & Dudycha, 1973). Nystedt and Magnusson (1973) and Conrad (1973) utilized cue labels, but did not systematically investigate the influence of these labels on predictive accuracy.

Two recent studies have focused on the effect of cue and criterion labels on multiple cue performance. Miller (1971) required subjects to predict the criterion of a final course examination from the cues of a) the result of a mid-term exam, b) the result of a mid-term essay, and c) the result of an examination coolness test. Miller manipulated whether or not the cues and criterion values were labeled, and further manipulated the attached labels to be either "congruent" or "incongruent" with the actual cue validities (i.e., whether or not the cue label made sense in light of the expected cue validity). Miller's results indicated that achievement was highest when congruence existed between cue validities and the accompanying labels. Moreover, subject achievement was lower in that condition in which no labels were employed whatsoever, but was higher than the condition that employed incongruent labeling.

In a related study, Muchinsky (1973) examined the influence of labels on subjects' ability to utilize suppressor variables in a multiple cue task. His results suggest that the use of labels provides subjects with a "psychologically meaningful context" in which they are better able to use the information presented to them in the prediction task, as opposed to a prediction task structured in an abstract context.

Shifting Cue Validities

Although most multiple cue probability learning studies have restricted their emphasis to the study of the human inference process in stationary tasks, a few studies have focused on the situation in which the individual is faced with a nonstationary, or dynamic environment, in other words, an environment in which the functional relationship between cues and criterion changes over time. In general, these studies have indicated that subjects are able to detect and "track" a shift in cue weights over time, but that adaptation to the new cue weights is considerably slower than adaptation to the initial set of weights.

One of the earliest of these studies was conducted by Peterson, Hammond, and Summers (1965). During 100 preshift trials subjects derived estimates of criterion values based on cue validities of .66, .33, and .00 in a three orthogonal cue environment. Following the shift, the cue validities of .66 and .33 were reversed while the cue with zero validity remained the same for a second set of 100 trials. The results indicated that, although the means of the response beta weights showed an appropriate rank ordering for both the preshift and the postshift trials, the preshift rank ordering was accomplished more quickly (20 trials) than was the postshift rank ordering (60 trials). The authors interpreted this finding to indicate that a change in the relative weighting of cues could be detected and responded to appropriately, but with great difficulty. Unfortunately, no data comparing subject achievement or consistency between the preshift and postshift trials were available.

A later study investigating changing cue validities was reported by Summers (1969). This investigator examined three different types of task shifts: change in the relevant cue (cue shift), change from positive to negative regression function (rule shift), and a change in both the validity of the relevant cue and the regression function relating the cue to the criterion (complete shift). Summers' results indicated that for postshift achievement, the initial decrement in performance was greatest for the rule and complete shift condition. The complete shift condition showed significantly slower adaptation to changes in the task properties than did the cue and rule shift conditions.

As in the Peterson, Hammond, and Summers (1965) study, the cue shift merely involved a reversal of the validities for two of the initial cues, and not a shift to a totally new set of cue validities. Subjects in these two studies were dealing with the same set of cue validities in both the preshift and postshift trials. It is thus difficult to determine on this basis whether the subjects learned to assign new cue weights or merely acquired an ability to reverse cue weights between two of the cues. These two experiments could have been improved if the postshift cue validities were completely different from those utilized in the preshift trials. However, it should be pointed out that the pattern of shifting cue validities in both of these experiments enabled the investigators to hold system predictability (R^2e) at a constant level for both the preshift and postshift trials.

A recent study by Dudycha, Dumoff, and Dudycha (1973) focused on the problem of shifting cue validities in a single-cue environment. Dudycha, Dumoff, and Dudycha examined two kinds of cue validity shifts and the interaction of these shifts with number of preshift trials and instructional set. Two orders of shift were employed, one in which the cue validity (and thus R^2e) shifted from high (r_{ei} = .895) to low (r_{ei} = .634), and the second in which the cue validity shifted from low (r_{ei} = .634) to high (r_{ei} = .895). Indexing performance by the measures of achievement (r_a) and consistency (r_s) these investigators found that performance in the postshift ecology was significantly poorer than that of the preshift ecology. Furthermore, both r_a and r_s attained higher levels when the shift involved a change from a high cue validity (and thus system predictability) to a low cue validity, than when the shift involved a change from a low cue validity to a high cue validity. The results of this study were interpreted in light of the AB-BA transfer paradigm; that is, when changing from a task of high difficulty to one of low difficulty (that is, from high to low cue validities) positive transfer should be evidenced. However, when the change involves a shift from a task of low difficulty to one of high difficulty, (that is, from low to high cue validities) negative transfer should be evidenced.

Although methodologically sound, the Dudycha, Dumoff, and Dudycha (1973) study can be criticized on two important facets. First, the study was restricted to a single cue environment; thus a change in cue validity necessarily produced a corresponding change in system predictability. In this case, the cue shift was in accuality also a shift in system predictability. Secondly, the ecology employed in this experiment was quite sterile, involving a single cue. Subjects' task was limited to tracking a shift in a single piece of information. Such an analogous situation is rarely, if ever, to be found in "real-life" inference situations. No discrimination among available cues was necessary.

In addition to the foregoing criticisms of the shifting cue validity studies reviewed above, a further comment is in order. All three experiments employed conditions in which the proper weighting of available cues changed over time. These studies provided no basis with which to compare their results to a control condition in which the cue validities did not change, but rather remained the same throughout the learning trials.

Brehmer (1973c) has reported a related study in which subjects were initially trained individually to depend upon either a linear or a nonlinear cue in a two-cue environment. Subjects were then assigned to groups of two in an interpersonal learning (IPL) task. Subjects' response data (r_a, r_s, r_m,) were analyzed with respect to the degree of change required from each subject's initial cognitive system (i.e., initial dependence on either a linear or a nonlinear cue). When subjects were required to drastically adjust their initial cognitive systems (maximum change), they were less successful in making correct criterion estimates and in matching the ecological weighting system, than when little change or no change was required.

Feedback

With respect to the effects of feedback on multiple cue learning, research has centered on two major modes of feedback. The first of these, known as "outcome feedback", allows the subject to compare his response with a specified criterion response. On the other hand, "lens-model feedback" allows the subject access to information concerning the lens-model parameters; more specifically, feedback is given in the form of ecological cue validities (r_{ei}), cue intercorrelations (r_{ij}), cue utilization coefficients (r_{si}), or some combination thereof.

The research to date has granted overwhelming support to the superiority of lens-model feedback (Newton, 1965; Todd & Hammond, 1965; Hammond, Summers, & Deane, 1973). The rationale for this superiority is that, since multiple cue learning involves the learning of probabilistic relations between cues and criterion, outcome feedback provides some amount of erroneous information. This may result in a decrease in subject consistency and ability to match cue weights (Hammond, Summers, & Deane, 1973). It has been suggested, therefore, that "other forms of feedback--more suited to the nature of the task and to the nature of human cognition--need to be developed and introduced into studies of human learning" (Hammond, Summers, & Deane, 1973, p. 34).

An alternative form of feedback which has received relatively little attention is that involved when more than one individual contributes information to the inference task. As Hammond (1972) and Brehmer (1973a) have pointed out, human learning often takes place in a social context in which individuals may obtain information about the task both from the task itself and from other persons who have had previous experience with the task. Following this "triple-system lens model paradigm" developed by Hammond (1972), subjects who have been individually trained to have specific inference policies are brought together in a new inference task. In such a task, all subjects observe the same cues (e.g., Brehmer, 1973a) or specified subsets of cues (e.g., Young, 1973) and make individual estimates of criterion values. After interpersonal interaction, a joint estimate is made.

Experiments utilizing this paradigm have shown that interpersonal learning is generally faster than individual learning of the same task (Brehmer, 1971; Earle, 1973), but that subjects' performance generally lacks consistency. On the other hand, in interpersonal learning, as compared to individual learning, subject performance with a single cue task is not significantly better than with a multiple cue task. A possible explanation for this observed effect is that subjects alter their strategy or policy in order to reduce conflict with their partner's existing strategy (Brehmer, 1973a).

Studies undertaken in the interpersonal learning framework typically involve bringing together subjects who have previously acquired separate inference strategies. A related approach not investigated in the literature is that of allowing previously inexperienced individuals to learn a multiple cue task together in a social context. Thus individual participants would learn from each other independently of previous experience with a particular inference policy. Typically, such a procedure would be followed in workshops designed to improve decision making ability of individual participants. The question of group versus individual decision making has been addressed elsewhere (Davis, 1969; 1973), but not in the context of multiple cue probability learning as proposed in the present study. Such a system of group feedback needs to be investigated systematically within the multiple cue model.

Purpose of the Study

The present study sought to enrich the basic theoretical framework of multiple cue probability learning as a normative model of real-life decision making behavior. According to such a normative model, decision making performance should be solely a function of the mathematical relations existing between the stimulus cues and criterion (Brunswik, 1952). Studies undertaken within the multiple cue framework have since attempted to take into account variables other than the mathematical relationships within the model which might affect human inference behavior. The present study was an additional attempt to further test and expand normative decision theory and to increase its generalization to real-life decision making situations.

Accordingly, the purpose of this study was to extend the basic multiple cue normative decision making model to include a "sampling of stimulus-situations" (Brunswik, 1952) which might accompany real-world decision making behavior. This was attempted by systematically manipulating several "ecologically valid" independent variables and examining their effect on subject performance. At the same time, other related variables (identified in past research) were held constant at levels which are consistent with a real-life decision making situation. It was reasoned that such a strategy would increase the external validity of the study and thus allow for more accurate generalizations to real-life decision making situations.

The task followed the basic format of multiple cue probability learning studies. Subjects observed cue values and made criterion predictions over a number of trials. Both the cues and the criterion values were assigned semantic labels to provide subjects with a more "meaningful psychological environment" (Muchinsky, 1973). In this case, the subjects, as undergraduate students, predicted final exam scores from three sources of information. Task predictability (R^2e) was established at a level high enough so as to allow subjects a fairly high degree of predictive accuracy (and thus hopefully high task motivation), but low enough to minimize the possibility of ceiling effects over learning trials. Since previous research had failed to demonstrate consistently the superiority of redundant information, all cues were orthogonal.

The choice of the independent variables was dictated by consideration of previous multiple cue research and of a real-life decision making situation. Individuals whose jobs require them to make decisions on the basis of different sources of information (e.g., credit managers, stockbrokers, weather forecasters, college admissions personnel) usually derive prediction strategies after consultation with others in similar positions. By gaining feedback from relevant others, subjects may be able to develop and refine their decision strategies.

Furthermore, these individuals must learn to acquire a flexible strategy for combining sources of information to derive a criterion estimate. That is, they must learn to become sensitive to changes in cue-criterion relations over time. In addition, they must be equally prepared to deal with sources of information whose relation to a criterion might be widely discrepant or of relatively equal relevance or importance in predicting the criterion variable.

The present study investigated the effects of group discussion as a form of feedback, subject performance in a dynamic versus static environment, and subject performance within different orders of cue shift. In addition, subject performance was measured over blocks of trials to assess learning. Subject performance was measured by the standard multiple cue indices of achievement, consistency, and matching.

With regard to these independent variables, several specific predictions were made based on the normative theory and on previous multiple cue research. First, according to the normative theory, given a constant level of system predictability, subject performance should be independent of the cue validity distribution. However, it was hypothesized that subjects will utilize more efficiently a set of _different_ cue validities (i.e., $r_{e1} = .76$, $r_{e2} = .40$, $r_{e3} = .10$) than a set of _same_ cue validities (i.e., $r_{e1} = .50$, $r_{e2} = .50$, $r_{e3} = .50$). That is, subjects should be able to learn to utilize "good" cue validities and to ignore "poor" ones as opposed to cues that possess equal predictive validity (Dudycha & Naylor, 1966).

With respect to subjects' ability to track a shift in cue validities, normative theory would suggest that the order of shift (i.e., a shift from either Same to Different or from Different to Same) should not affect the level of subject performance. However, previous research by Dudycha and Naylor (1966) suggests that adding a cue of greater predictive validity should increase subject performance whereas adding a cue with less predictive validity should result in a decrease of subject performance. Since the cue-criterion configuration of (r_{ei} = .76, .40, and .10) reflects cues of different (greater and lesser) predictive ability than the cue-criterion configuration employing cues of equal predictive ability (r_{ei} = .50, .50, and .50), the results from the Dudycha and Naylor study would indirectly suggest that subject performance should be superior in that condition in which cue validities shift from Same to Different rather from Different to Same.

Those subjects who learn the multiple cue task within the context of group discussion should demonstrate greater consistency (r_s) in their judgments and should be able to learn the task more quickly than will individuals learning the task on an individual basis (Hammond, 1972; Brehmer, 1973a). In addition, subjects in the group discussion feedback condition should be able to detect and track a change in shifting cue validities more efficiently than will subjects in the individual learning condition. This prediction was based on the results of studies concerning the effects of feedback on multiple cue learning as well as research concerning individual versus group decision making (Davis, 1969; 1973).

METHOD

Multiple Cue Inference Task

The inference task for this experiment followed the basic design of previous multiple cue probability learning studies. Subjects were asked to respond to 150 trials of cue-criterion pairings in which they predicted a criterion from a set of three cues. Both the cues and criterion values were meaningfully labeled in this experiment. The meaningful labels, similar to those used by Miller (1971), were chosen so as to provide the subjects (college students) with a "meaningful psychological environment" (Muchinsky, 1973), and to be representative of the ecological validities found in such an environment. The actual labels chosen for the cue validities were: a) score on a mid-term exam, b) score on a term paper, and c) a lab score, all taken from the same course. Accordingly, the criterion value was labeled as the score of a comprehensive final examination taken in the same course.

Task Properties

The multiple cue task used in this study was constructed such that the criterion was linearly related to each of the cues. The strength of these relationships, as represented by the cue validities, varied according to the experimental conditions. In some of the conditions the cue validities were .76, .40, and .10, while in other conditions, the cue validities were .50, .50, and .50. For both sets of cue validities, cues were orthogonal and resulted in a system of equal task predictability (R^2e) = .75).

Both cues and criterion were generated as two digit numbers, normally distributed with means of 50 and standard deviations of 10. These numerical values were generated by means of the correlated score generation program developed by Wherry, Naylor, Wherry, and Fallis (1965). The values were generated according to the cue validity, cue intercorrelation, and system predictability parameters specified above.

The factor structure used as input to the stimuli generation program is included in Appendix A. The stimuli were generated so as to derive a close approximation (+.05) to the theoretically specified relationships between cues and criterion. During the generation phase, only the three blocks of cue-criterion values that most nearly approximated the specified correlative values for each condition were chosen for the Same and Different cue validities. An additional requirement for selection of blocks was that the variance among the three cue values for any one trial was maximum. This requirement was imposed in order to maximize the sensitivity of the subject performance indices to differences in cue configuration (i.e., Same versus Different cue validities). Appendix A gives the theoretical and empirical correlations between cues and criterion for each block of 25 trials.

Group Discussion Conditions

In order to evaluate the effect of group discussion as a form of feedback in a multiple cue task, half of the subjects were assigned to a group discussion condition while the other half were assigned to a no group discussion condition. All subjects in the group discussion condition were tested in groups of three individuals each. Those subjects assigned to the no group discussion condition were tested in groups whose size varied from two to six. For the subjects assigned to the group discussion condition, the three subjects in the experimental session were considered as constituting a group. They were instructed that periodically during the experiment they would be given the opportunity to exchange ideas with the members of their group concerning techniques or strategies utilized in performing the task, using the group as a "sounding board" to assess the quality of their own individual prediction strategies. However, they were told that their actual performance in the prediction task would be entirely an individual effort. The group discussion thus took place during the interval between each of the six trial blocks, or, more specifically, after every 25 trials. The discussion period lasted approximately two minutes.

For the subjects not assigned to the group discussion condition, the individuals in an experimental session performed the multiple cue task for the 150 trials, but did not interact with each other during the rest periods between trial blocks. Instead, they were instructed to evaluate individually their own strategies in the prediction task. Specific task instructions for both the group discusion and no group discussion subjects are given in Appendix B. Thus all subjects, regardless of group discussion assignment, learned the multiple cue task under conditions of distributed, rather than massed practice.

Shift Conditions

To determine subjects' responses to changing or shifting cue validities, two shift conditions (shift versus no shift) were incorporated into the design of the experiment. Also, two levels of order of shift were included within each of the two shift conditions. This resulted in four shift x order conditions completely crossed with two levels of group discussion. For the first of these conditions, subjects made inferences using the set of three different cue correlations (.76, .40, and .10) for the first 75 trials. For the last set of 75 trials, judgments were made on the basis of a set of identical or Same cue correlations (.50, .50, and .50). Thus subjects shifted from using a set of Different cue correlations to using a set of Same cue correlations (D-S). Likewise for the second shift condition, the cue correlations shifted from Same to Different (S-D). The no shift groups may be regarded as control groups with which to compare the performance of the shift groups. Thus there was a _potential_ shift in both the shift and no shift conditions while there was an _actual_ shift only in the shift condition.

Those subjects assigned to the no shift condition made all their criterion estimates either on the basis of the set of Different cue correlations (D-D) or on the basis of a set of Same cue correlations (S-S) for the entire 150 trials. In brief, the shift orders (D-S, S-D) were included within the shift condition, while the shift orders of (D-D, S-S) were included within the no shift condition. The complete experimental design is presented in Table 1.

Subjects

The subjects for this experiment were 96 male undergraduate students enrolled in introductory and lower division psychology classes at Iowa State University. Subjects were given course credit toward their final grade as a minor academic inducement for participation. Assignment of subjects to the experimental conditions was based on the order in which they signed up for participation in the experiment. The order in which the experimental conditions were run was decided on a random basis.

Stimuli and Apparatus

All sets of cues and criterion were typed onto a role of teletype paper arranged such that, for each trial, the three

Table 1

Experimental Design

	Preshift Segment	Postshift Segment
	Blk 1 Blk 2 Blk 3	Blk 1 Blk 2 Blk 3
Trial	1------------------75	76-----------------150
Group Discussion Shift, Order 1	Different	Same
Group Discussion Shift, Order 2	Same	Different
Group Discussion No Shift, Order 1	Different	Different
Group Discussion No Shift, Order 2	Same	Same
No Group Discussion Shift, Order 1	Different	Same
No Group Discussion Shift, Order 2	Same	Different
No Group Discussion No Shift, Order 1	Different	Different
No Group Discussion No Shift, Order 2	Same	Same

Note: Different: Cue 1 (Mid-Term Exam) $r_{e1} = .76$ $R^2_e = .76$
Cue 2 (Mid-Term Essay) $r_{e2} = .40$
Cue 3 (Mid-Term Lab) $r_{e3} = .10$

Same: Cue 1 (Mid-Term Exam) $r_{e1} = .50$ $R^2_e = .76$
Cue 2 (Mid-Term Essay) $r_{e2} = .50$
Cue 3 (Mid-Term Lab) $r_{e3} = .50$

cues and the corresponding criterion value were located on the same row. This teletype paper was loaded onto a device which was specially equipped to fit under an opaque projector.

This device was fitted with a special template which allowed the experimenter to project one row of three cues and the corresponding criterion value on a screen at the same time. For each trial, the cues were presented first, followed by the presentation of both the cues and the criterion. Each set of cue-criterion values were presented sequentially by rolling the teletype paper on the special device. The cue and criterion values for blocks of Same and Different cue validities are given in Appendix A.

For all experimental conditions, the cues were labeled "MID-TERM EXAM SCORE", "MID-TERM ESSAY SCORE", and "MID-TERM LAB SCORE." Accordingly, the criterion value was labeled "FINAL EXAM SCORE." These labels were mounted directly onto the template.

Procedure

Each experimental session lasted approximately one hour and a half and consisted of from two to six subjects. Before the beginning of the experimental task, all subjects were instructed that the purpose of the experiment was to study how individuals use information from several variables, or cues, in making predictions about another variable, called the criterion. Subjects were told that their task in the experiment was to observe the values of the three cues and, on the basis of these values, to estimate the value of the criterion. More specifically, subjects were informed that their task would involve making predictions of the criterion (final exam score) by using the cues of a mid-term exam score, a mid-term essay score, and a mid-term lab score, all taken from the same course. This course was one taught at another university.

Subjects were informed that they were to predict a two-digit number, representing the final exam score, on the basis of three two-digit numbers representing scores on a mid-term exam, a term paper, and a lab score respectively. Because the cue and criterion values were generated with means of 50 and standard deviations of 10, special instructions regarding the meaning of the numbers with respect to the cue and criterion labels were necessary. Accordingly, subjects were told that a slightly different grading system was being used for this course than was normally encountered in college courses. A score of 50 represented an average score, and the scores could range from 10 to 90. Subjects were also instructed that the final exam, being comprehensive, was to reflect the same skills, knowledge, etc. manifested in the term paper, mid-term test, and lab score. Subjects were told that they were to make their estimates of the criterion on the basis of the value of the cues shown on a particular trial as well as on the basis of their cumulative experience with cue-criterion pairings from previous trials. The criterion value displayed was to be regarded as the best possible estimate for a particular trial, and that, although perfect prediction was impossible because of the nature of the task, the accuracy of their estimates should improve over trials.

Those subjects assigned to the group discussion condition were told that, periodically during the course of the experiment, they would be given rest periods during which they would be allowed to discuss the task among themselves. During this time they would be encouraged to discuss different strategies that might be useful in making predictions about the criterion. It was stressed, however, that all predictions would be made on an individual basis once the experimental trials had resumed.

On the other hand, those subjects assigned to the no group discussion condition were informed that, periodically during the course of the experimental task, they would be given rest periods during which they should contemplate or evaluate the strategies they had been employing to predict the criterion.

To establish the rationale for the shift manipulation, subjects were told that, in the course of the experiment, it was possible that they may be dealing with cues and a criterion taken from another course and instructor. They were informed that the relationship between the cues and criterion might not necessarily remain the same for the new course, since different instructors might not agree on the relative importance which they think should be placed on mid-term exam scores, essay scores, and lab scores in determining the final exam score. Therefore subjects should be aware that the relationships existing between cues and criterion may change over the course of the experimental task. After the first 75 trials, all subjects were informed that the cue scores were taken from another course. Therefore the relative importance of the cue scores in predicting the criterion may not be the same as before. These instructions established the rationale for a potential shift in cue validities.

Following these instructions, five practice trials were given in order to familiarize subjects with the task and to answer any questions. Each experimental trial lasted approximately fifteen seconds--ten seconds in which subjects viewed the cues and recorded a response followed by five seconds in which the actual criterion value was paired with the corresponding cue values. For each experimental trial, subjects observed the cue values, recorded their prediction of the criterion value (Y's) and then observed the actual criterion value (Ye) (outcome feedback), in order to compare the actual value with their predicted value. Each subject recorded his estimate of the criterion value on an answer sheet with 150 numbered spaces provided for his responses (Appendix C). After the experiment terminated, subjects were administered a post-experimental questionnaire. A copy of this questionnaire is included in Appendix D.

Experimental Design

The design for this study incorporated two levels of group discussion feedback (group discussion versus no group discussion), two levels of cue shift (shift versus no shift) and two orders of cue shift (Same to Different versus Different to Same within the shift condition, and Same to Same versus Different to Different within the no shift condition). Learning trials were grouped into six blocks of 25 trials each. The first three blocks (trials 1-75) were considered as preshift trials. The second three blocks (trials 76-150) were considered as postshift trials. Thus the two pre-post segments constituted one of the within factors. The three blocks within the preshift segment constituted the second within factor. In summary, the between subject factors consisted of two levels of group discussion, two levels of shift, and two levels of order of shift. The within factors consisted of two levels of pre-post segments and three levels of blocks.

RESULTS

Subjects' responses (Ys) were divided into three preshift and three postshift blocks of 25 trials (hence referred to as the pre-post segments factor) for the purpose of data analysis. For each of these trial blocks the three multiple cue indices of achievement (r_a), consistency (r_s), and matching (r_m) were calculated and transformed into Fisher Zr values. These Zr values were used as data in three 2 (levels of group discussion) x 2 (levels of shift) x 2 (levels of order of shift) x 2 (pre-post segments) x 3 (blocks) factorial analyses of variance with repeated measures over levels of the last two factors.

A separate data set was generated by dividing the twelve subjects in all eight experimental conditions into four groups of three subjects each. For all three indices Fisher Zr values were summed over the three subjects within a group and these sums were used as data for three additional 2x2x2x2x3 factorial analyses of variance with repeated measures over levels of the last two factors. This data set was generated because the group discussion manipulation may have resulted in nonindependence of subject responses, and thus necessitated the employment of an additional mean square error term which was used in the post hoc comparisons.

The results of the statistical analyses are presented as they pertain to subject achievement, consistency, and matching. Several main effects and interactions were contained in higher-order interactions and therefore were not selected for graphical representation. For all analyses, all factors with the exception of subjects were treated as fixed. The analyses of variance presented in Tables 2, 3, and 4, are those performed on the data from individual subjects rather than from groups. The mean square error for both the individual and group analyses are included in the analysis of variance tables. To avoid interpretative difficulty, the reader is instructed that the first mean square error of each pair is the appropriate mean square error for the analysis of variance presented in each table. On the other hand, the second mean square error of each pair is the mean square error for the analysis of variance on the group data. This mean square error was used in the appropriate post hoc tests in order to provide a conservative test of differences between treatment means.

The results of the analysis of variance of achievement data are presented in Table 2. This analysis indicated that no between-subject main effects reached statistical significance ($p > .10$). Only the within-subject main effect of blocks within pre-post segments was significant (F 2,176 = 10.214, $p < .01$).

Table 2

Summary of the Analysis of Variance on the Fisher Z Scores for Subject Achievement over Blocks of Trials

Source of Variance	df	MS	F
Between Subjects			
Group Discussion (GD)	1	.048	.387
Shift (S)	1	.001	.008
Order of Shift (O)	1	.065	.542
GD x S	1	.025	.202
GD x O	1	.336	2.709
S x O	1	.973	7.846**
GD x S x O	1	.018	.145
Subjects/Groups (S/G)[1]	88	.124	
Subjects/Groups (S/G)[2]	24	.143	
Within Subjects			
Pre/Post Segments (P)	1	.012	.214
G x P	1	.041	.732
S x P	1	.268	5.107*
O x P	1	.000	.000
G x S x P	1	.000	.000
G x O x P	1	.011	.196
S x O x P	1	.039	.696
G x S x O x P	1	.059	1.053
P x S/G[1]	88	.056	
P x S/G[2]	24	.043	

$^{**}p<.01$

$^{*}p<.05$

[1]Mean square error for the analysis of variance on individual data. This was used as the mean square error for the analysis of variance presented in this table.

[2]Mean square error for the analysis of variance on group data. This was used as the mean square error for the post hoc comparisons.

Table 2 (continued)

Source of Variance	df	MS	F
Blocks (B)	2	.286	10.214**
G x B	2	.031	1.107
S x B	2	.034	1.214
O x B	2	.024	.857
G x S x B	2	.101	3.607*
G x O x B	2	.012	.428
S x O x B	2	.045	1.607
G x S x O x B	2	.040	1.429
B x S/G[1]	176	.028	
B x S/G[2]	48	.026	
P x G	2	.007	.246
G x P x B	2	.009	.310
S x P x B	2	.016	.568
O x P x B	2	.072	2.551
G x S x P x B	2	.047	1.665
G x O x P x B	2	.018	.623
S x O x P x B	2	.099	3.514*
G x S x O x P x B	2	.024	.850
Error[1]	176	.028	
Error[2]	48	.025	
Total	575		

However several of the higher-order interactions were significant sources of variation. The between-subject interaction of shift x order (F 1,88 = 7.846, $p<.01$) reached statistical significance. The interaction between shift x pre-post segments was also significant (F 1,88 = 5.107, $p<.05$) as were the interactions between group discussion x shift x blocks (F 2,176 = 3.607, $p<.05$) and between shift x order x pre-post segments x blocks (F 2,176 = 3.514, $p<.05$). The group discussion x shift x blocks interaction is represented in Figure 2 while the shift x order x pre-post segment x blocks interaction is plotted in Figure 3.

The results of the analysis of variance of subject consistency data are presented in Table 3. Consistency varied significantly across blocks within pre-post segments (F 2,176 = 31.264, $p<.01$). The two-way interaction between order x pre-post segments (F 1,88 = 6.325, $p<.05$) also reached statistical significance.

The three-way interaction between shift x order x pre-post segments was also significant (F 1,88 = 5.707, $p<.05$) and is presented in Figure 4. Sheffe's test for nonpairwise comparisons (Kirk, 1968) indicated that, under the shift condition, subject consistency was significantly greater when predictions were made from Different cue validities than from Same cue validities (F 1,24 = 164.39, $p<.01$).

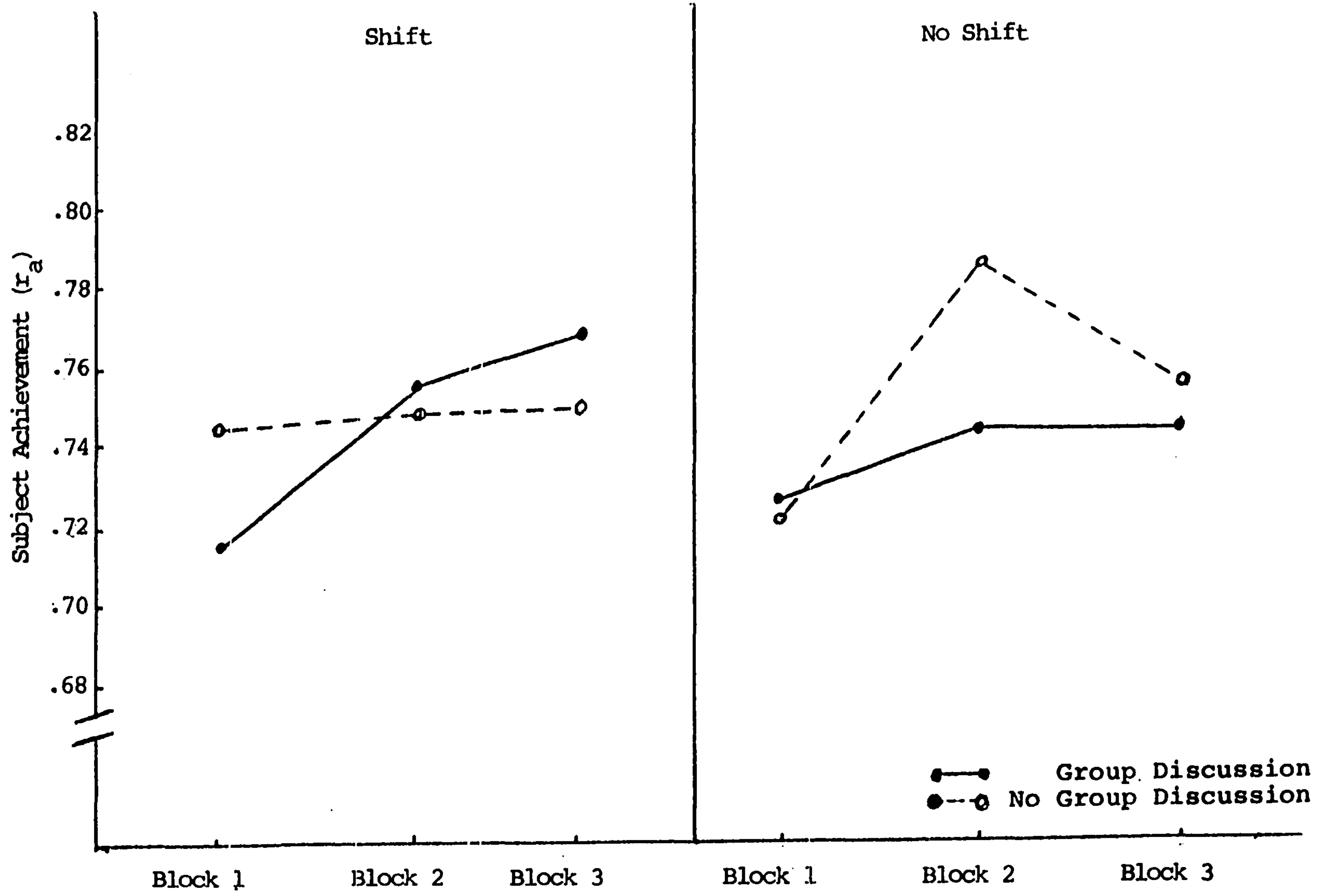

Figure 2. Significant interaction between Group Discussion x Shift x Blocks for subject achievement (r_a).

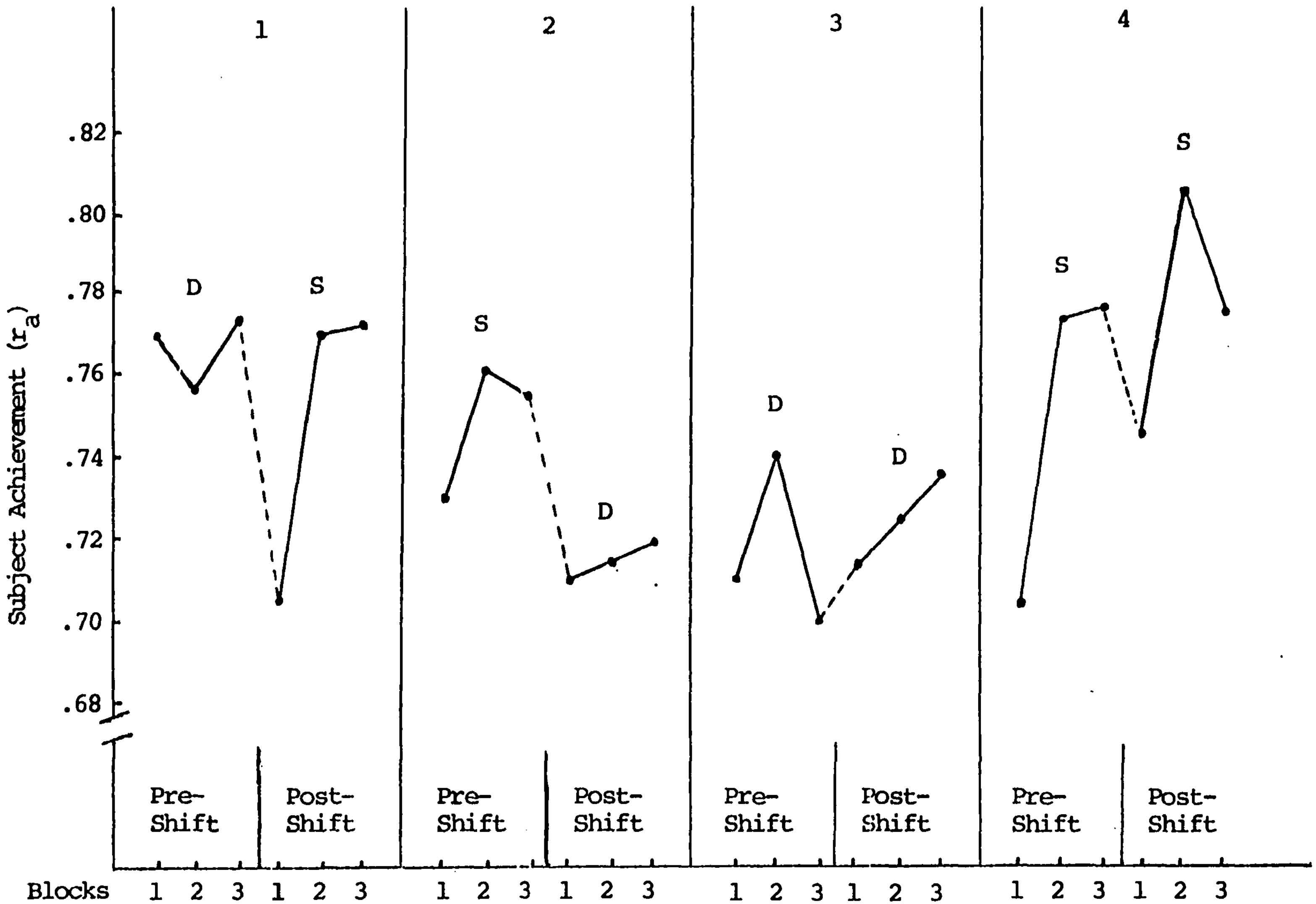

Figure 3. Significant Shift x Pre-Post Segments x Blocks interaction for subject achievement (r_a). (D = Different cue validities, S = Same cue validities)

Table 3

Summary of the Analysis of Variance on the Fisher Z Scores for Subject Consistency over Blocks of Trials

Source of Variance	df	MS	F
<u>Between Subjects</u>			
Group Discussion (GD)	1	.179	.505
Shift (S)	1	.077	.223
Order of Shift (O)	1	1.091	3.083
GD x S	1	.116	.327
GD x O	1	1.313	3.709
S x O	1	.001	.003
GD x S x O	1	.000	.000
Subjects/Groups (S/G)[1]	88	.354	
Subjects/Groups (S/G)[2]	24	.328	
<u>Within Subjects</u>			
Pre/Post Segments (P)	1	.000	.000
G x P	1	.080	.650
S x P	1	.014	.114
O x P	1	.778	6.325*
G x S x P	1	.031	.252
G x O x P	1	.008	.065
S x O x P	1	.702	5.707*
G x S x O x P	1	.234	1.902
P x S/G[1]	88	.123	
P x S/G[2]	24	.089	

*$p < .05$

[1]Mean square error for the analysis of variance on individual data. This was used as the mean square error for the analysis of variance presented in this table.

[2]Mean square error for the analysis of variance on group data. This was used as the mean square error for the post hoc comparisons.

Table 3 (continued)

Source of Variance	df	MS	F
Blocks (B)	2	2.376	31.264**
G x B	2	.031	.408
S x B	2	.000	.000
O x B	2	.151	1.987
G x S x B	2	.170	2.237
G x O x B	2	.251	3.302*
S x O x B	2	.514	6.763**
G x S x O x B	2	.052	.684
B x S/G[1]	176	.076	
B x S/G[2]	48	.062	
P x G	2	.033	.478
G x P x B	2	.013	.194
S x P x B	2	.001	.013
O x P x B	2	.359	5.191**
G x S x P x B	2	.065	.941
G x O x P x B	2	.148	2.142
S x O x P x B	2	.192	2.778
G x S x O x P x B	2	.055	.789
Error[2]	176	.069	
Error[1]	48	.058	
Total	575		

**$p < .01$

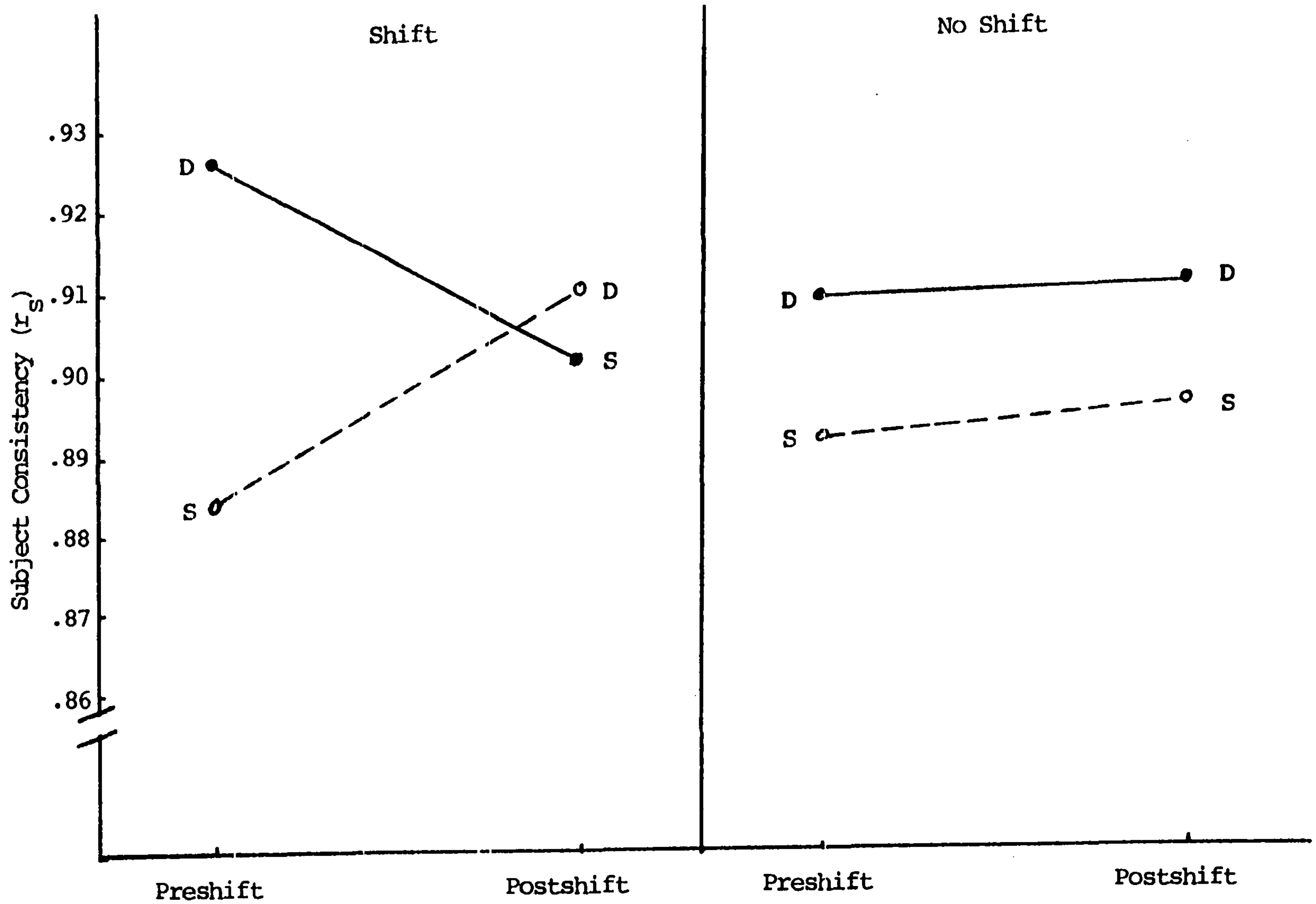

Figure 4. Significant Shift x Order x Pre-post Segments interaction for subject consistency (r_s).

The within-subject variable of trial blocks entered into three three-way interactions. These were: 1) group discussion x order x blocks ($\underline{F}$ 2,176 = 3.302, $\underline{p}<.05$), presented in Figure 5; 2) shift x order x blocks ($\underline{F}$ 2,176 = 6.763, $\underline{p}<(.01)$, presented in Figure 6; and 3) pre-post segments x order x blocks ($\underline{F}$ 2,176 = 5.191, $\underline{p}<.01$), presented in Figure 7. These significant interactions indicated that order of shift interacted with group discussion, shift and pre-post segments over trial blocks. These interactions primarily reflect fluctuation of consistency across blocks and yielded no clear interpretation with regard to the hypotheses of interest.

The results of the analysis of variance of matching data are presented in Table 4. Matching significantly varied over levels of order of shift ($\underline{F}$ 1,88 = 6.052, $\underline{p}<.05$) and over levels of trial blocks ($\underline{F}$ 2, 176 = 4.547, $\underline{p}<.05$).

The two-way interaction between shift x order was significant ($\underline{F}$ 1,88 =17.735, $\underline{p}<.01$) as was the two-way interaction between shift x pre-post segments ($\underline{F}$ 1,88 = 5.522, $\underline{p}<.05$). Comparable to the consistency data, the three-way interaction between shift x order x pre-post segments was statistically significant ($\underline{F}$ 1,88 = 7.793, $\underline{p}<.01$) and is graphically represented in Figure 8.

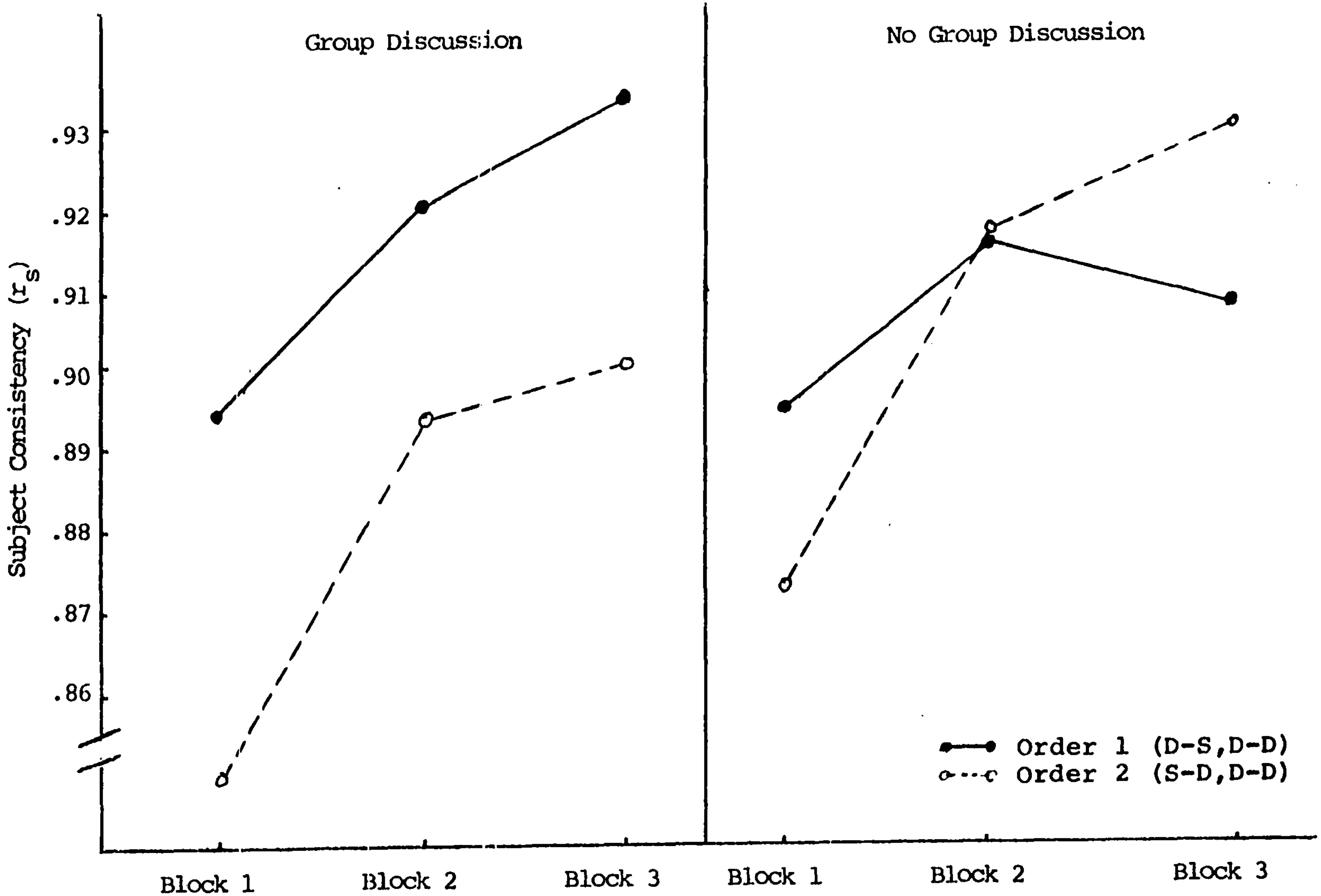

Figure 5. Significant interaction between Group Discussion x Order x Blocks for subject consistency (r_s).

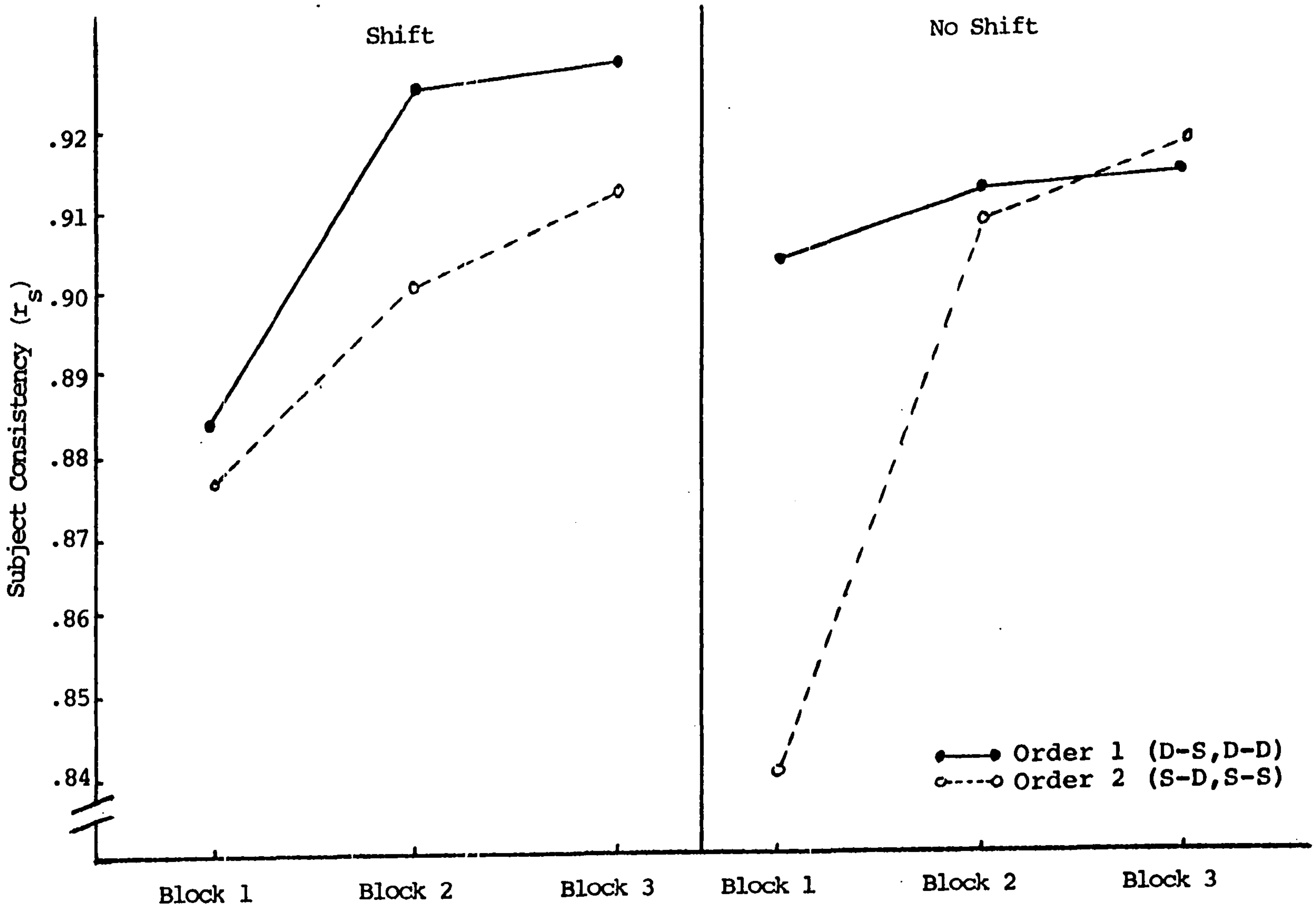

Figure 6. Significant interaction between Shift x Order x Blocks for subject consistency (r_s).

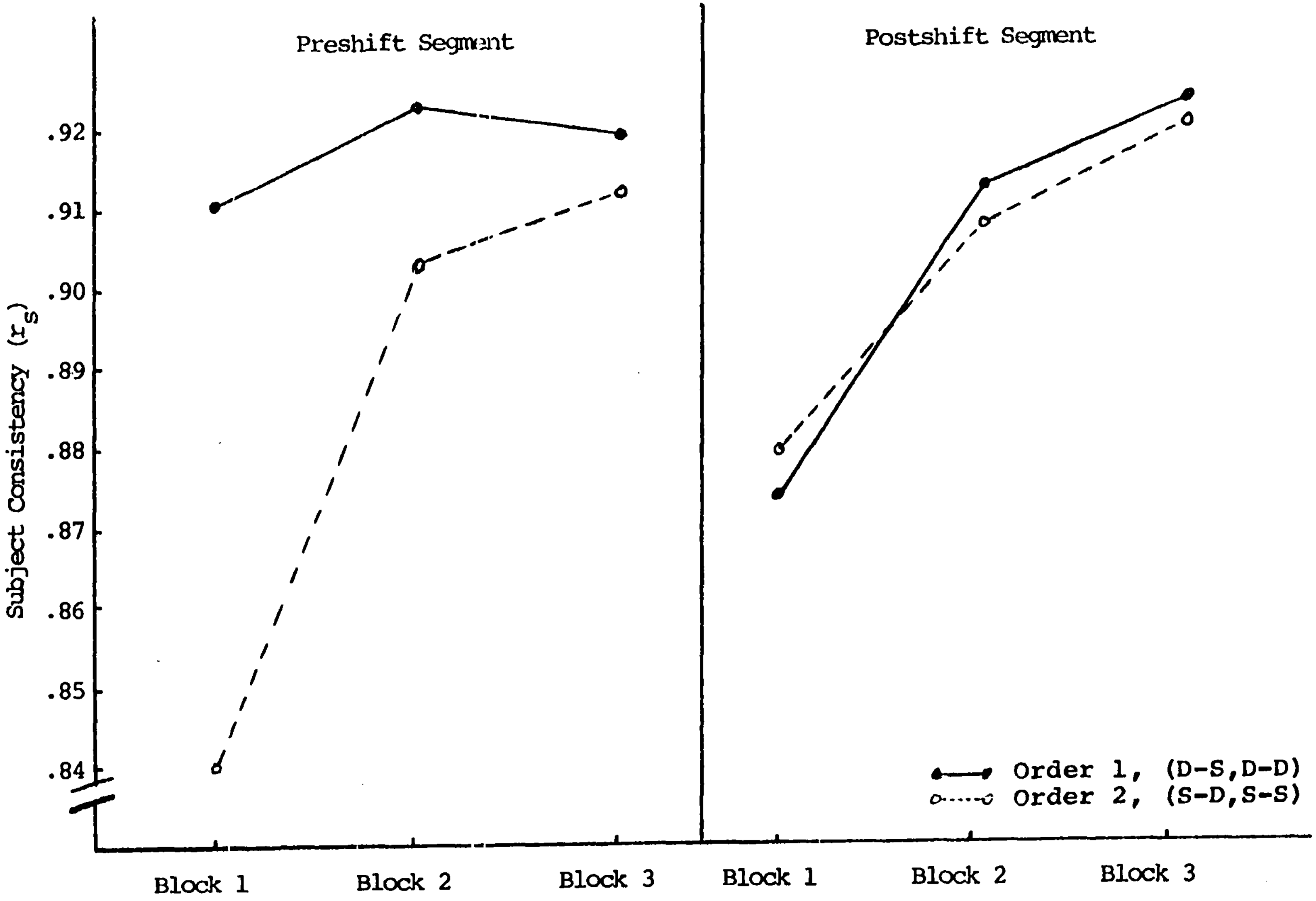

Figure 7. Significant Order x Pre-Post Segments x Blocks interaction for subject consistency (r_s).

Table 4

Summary of the Analysis of Variance on the Fisher Z Scores for Subject Matching over Blocks of Trials

Source of Variance	df	MS	F
<u>Between</u> <u>Subjects</u>			
Group Discussion (GD)	1	1.286	1.278
Shift (S)	1	.009	.009
Order of Shift (O)	1	6.089	6.052*
GD x S	1	2.709	2.693
GD x O	1	.569	.566
S x O	1	17.842	17.735**
GD x S x O	1	.023	.023
Subjects/Groups (S/G)[1]	88	1.006	
Subjects/Groups (S/G)[2]	24	1.419	
<u>Within</u> <u>Subjects</u>			
Pre/Post Segments (P)	1	2.014	2.889
G x P	1	.223	.319
S x P	1	3.849	5.522*
O x P	1	1.170	1.679
G x S x P	1	.017	.024
G x O x P	1	.000	.000
S x O x P	1	5.432	7.793**
G x S x O x P	1	1.136	1.630
P x S/G[1]	88	.697	
P x S/G[2]	24	.739	

$^{**}p<.01$

$^{*}p<.05$

[1]Mean square error for the analysis of variance on individual data. This was used as the mean square error for the analysis of variance presented in this table.

[2]Mean square error for the analysis of variance on group data. This was used as the mean square error for the post hoc comparisons.

Table 4 (continued)

Source of Variance	df	MS	F
Blocks (B)	2	1.587	4.547*
G x B	2	.733	2.112
S x B	2	.224	.646
O x B	2	.182	.524
G x S x B	2	.290	.836
G x O x B	2	.198	.571
S x O x B	2	.638	1.838
G x S x O x B	2	.112	.322
B x S/G[1]	176	.347	
B x S/G[2]	48	.428	
P x G	2	.005	.015
G x P x B	2	.256	.701
S x P x B	2	.374	1.024
O x P x B	2	1.853	5.065**
G x S x P x B	2	.727	1.986
G x O x P x B	2	.387	1.058
S x O x P x B	2	.023	.064
G x S x O x P x B	2	.259	.709
Error[1]	176	.366	
Error[2]	48	.322	
Total	575		

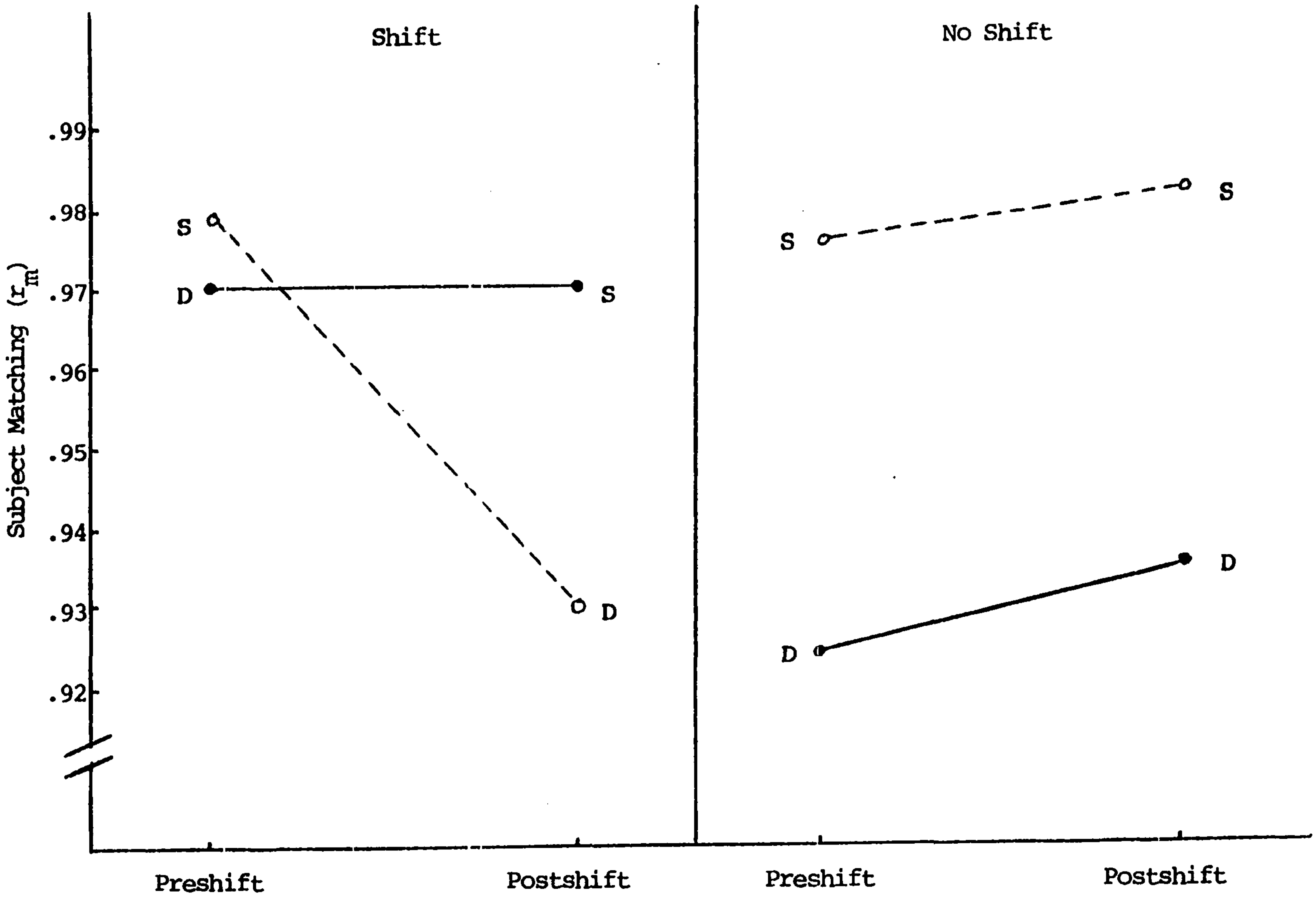

Figure 8. Significant Shift x Order x Pre-Post Segments interaction for subject matching (r_m).

Sheffe's test for nonpairwise comparisons indicated that, for the no shift condition, matching was significantly better when predictions were made from Same cue validities rather than from Different cue validities (F 1,24 = 15.76, $p<.05$).

Furthermore, Newman-Keuls tests for pairwise comparisons (Kirk, 1968) yielded two significant differences. First, matching was significantly greater for predictions based on Same versus Different cue validities when the cue validities shifted from Same in preshift trials to Different in postshift trials. Secondly, matching was significantly better in the no shift condition when predictions in postshift trials were made from same versus different cue validities ($p<.05$).

The only other higher-order interaction for matching to reach statistical significance was the three-way interaction between order x pre-post segments x blocks (F 2,176 = 5.065, $p<.01$). This interaction is plotted in Figure 9. As with the consistency data, this interaction primarily reflected fluctuation of matching performance over blocks and yielded no clear interpretation with regard to the hypotheses of interest. The reader is referred to the discussion for detailed explication of these results.

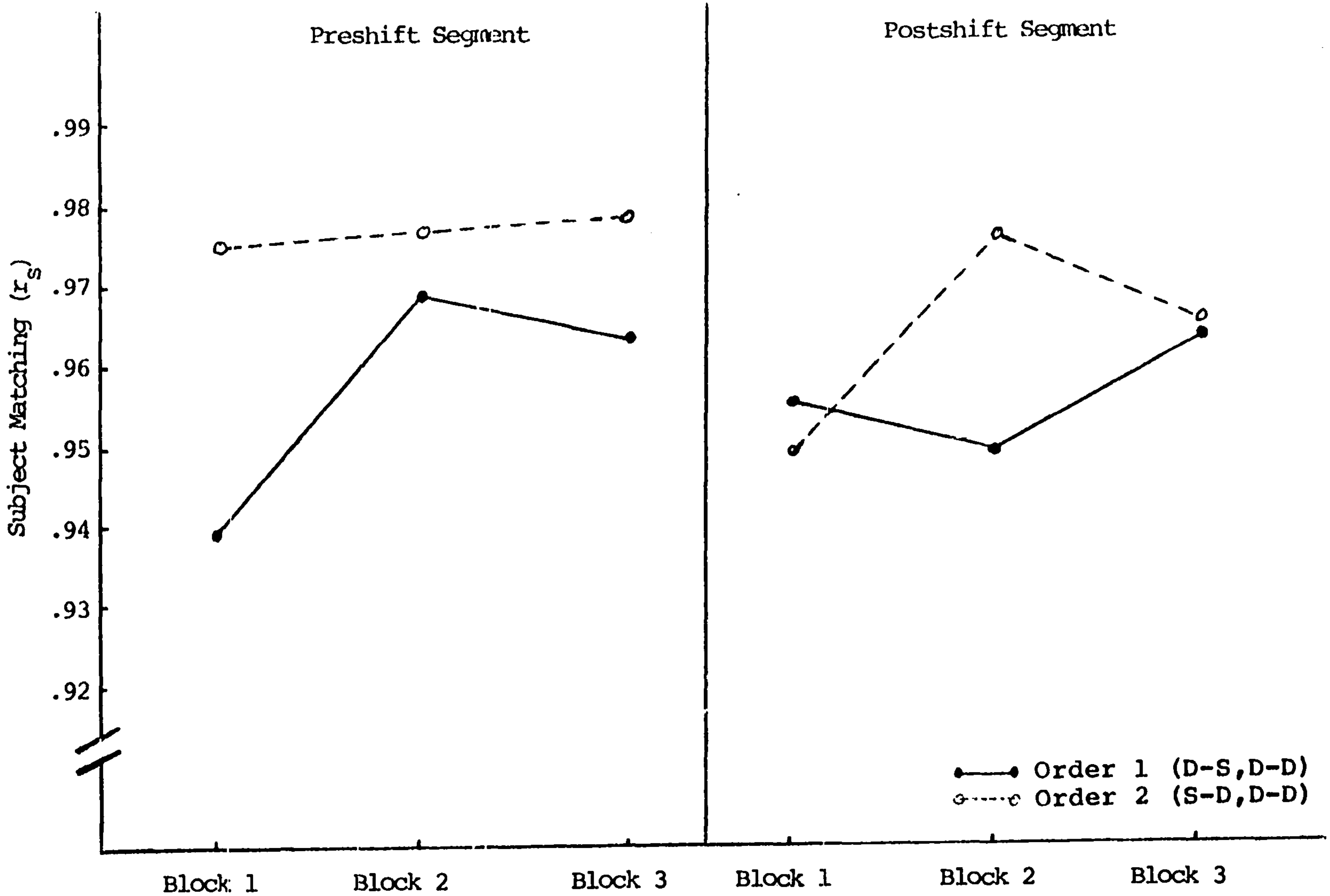

Figure 9. Significant interaction between Order x Pre-Post Segments x Blocks for subject matching (r_m).

Table 5 represents a summary of subjects responses to the post-experimental questionnaire items concerning how they thought they weighted the cues in the preshift and postshift segments (Appendix D). Subjects responded to these items by assigning weights ranging from "1" (weighted the cue very little) to "99" (weighted this cue very much).

Table 5

Average Weights Assigned to Cues by Subjects for all Experimental Conditions

Pre-Shift Segment Exam	Pre-Shift Segment Essay	Pre-Shift Segment Lab	Post-Shift Segment Exam	Post-Shift Segment Essay	Post-Shift Segment Lab
Group Discussion					
Different			Same		
71.67	49.25	23.83	42.52	54.63	58.00
Same			Different		
56.91	52.17	61.50	66.50	63.58	40.17
Different			Different		
72.65	49.33	29.92	62.58	57.00	41.44
Same			Same		
57.92	55.00	33.00	50.00	38.33	58.33
No Group Discussion					
Different			Same		
64.91	46.66	30.42	52.41	45.00	52.50
Same			Different		
61.00	43.50	50.00	62.00	46.00	50.00
Different			Different		
67.18	39.09	27.82	67.55	42.73	33.90
Same			Same		
41.50	50.90	55.40	46.80	48.00	56.10
Averaged Over Group Discussion/No Group Discussion					
Different			Same		
68.29	47.96	27.13	47.47	49.82	55.26
Same			Different		
58.96	47.84	55.75	64.25	54.79	45.09
Different			Different		
69.97	44.21	28.87	65.07	49.87	37.67
Same			Same		
49.71	52.95	44.20	48.40	43.17	57.22

DISCUSSION

Because of the nature of the questions addressed in the present study the results of subject achievement, consistency, and matching will be discussed as they relate specifically to the hypotheses of interest. These hypotheses concern evaluating inference behavior with Different versus Same cue configurations (given equal system predictability), with stationary versus shifting ecological cue validities, and with feedback given to subjects in the form of group discussion versus no such feedback given.

Two of the hypotheses tested in this study can be best addressed by examination of the four-way interaction between shift x order x pre-post segments x blocks for the achievement data (Figure 3). These hypotheses concern subject achievement with configurations of Same versus Different cue validities and with configurations of cues whose validities change over the course of time.

The four-way interaction is divided into four separate panels, these panels representing the function relating subject achievement over blocks of trials to the between-subject conditions. Variation of data across the pannels represents between-subject variation while variation of data within each panel represents within-subject variation. Because the between-subject variation is large relative to the within-subject variation (see Table 2), the reader is instructed that the functions can be best interpreted across trials within a condition (panel) rather than between conditions.

By dividing the plotted interaction into preshift and postshift segments it becomes apparent that the function relating achievement to between-group conditions over blocks differs markedly for the preshift and postshift segments. In the preshift trials, no overall difference in achievement between Same versus Different cue validities is evidenced. Theoretically, when identical cue validities are used in preshift trials (Same-Different; Same-Same) the plotted functions relating achievement to the experimental conditions over trial blocks should overlap. Likewise, the same expectation is true when Different cue validities are used in preshift trials (Different-Same; Different-Different).

However, the plotted interaction indicates that, although this expectation is upheld for Same cue validities for preshift trials (panels 2 and 4), the functions are markedly discrepant when predictions are made from Different cue validities in preshift trials (panels 1 and 3). Since both the Different-Same and Different-Different groups were treated identically in preshift trials, the discrepancy in these functions can only be attributed to sampling variation. Moreover, since the error mean square for between-subject comparisons was large relative to the error mean square for within-subject comparisons (see Table 2), achievement performance for the four between-subject groups (Different-Same; Same-Different; Different-Different; Same-Same) should be considered as approximately equal.

For the purpose of evaluating the hypotheses addressed in this study, the reader's attention is directed to the function relating achievement to the between-subject conditions over blocks for the postshift segment. This function is represented in the right half of panels 1 through 4. As in previous studies in which cue validities shifted in the course of the task, (e.g., Peterson, Hammond, & Summers, 1965; Summers, 1969; Dudycha, Dumoff, & Dudycha, 1973), there is a marked drop in achievement for the initial postshift trials when an actual (as opposed to potential) shift took place (Different-Same: panel 1; Same-Different: panel 2). This drop in achievement for the shift conditions is to be compared to the no shift conditions (Different-Different: panel 3; Same-Same: panel 4) in which little or no decrease in achievement was evidenced. This difference alone probably accounted for the interaction attaining its level of statistical significance ($p<.05$).

Examination of the postshift achievement data for the shift x order x pre-post segments x blocks interaction is most informative with regard to the hypotheses of interest. The first hypothesis concerned the superiority of subject achievement for a configuration of Different cue validities versus a configuration of Same cue validities. This hypothesis was clearly not supported. In fact, inspection of the postshift data suggest the opposite to be true.

Comparison of the achievement data for the no shift conditions of Different-Different and Same-Same indicates that, when subjects made predictions based on Same cue validities throughout the task (panel 4), there was evidenced a marked learning function over both preshift and postshift segments of the task. On the other hand, when Different cue validities were used throughout the task (panel 3) preshift achievement performance was highly variable while postshift performance showed little improvement over trials; there was no overall difference in achievement between preshift and postshift segments.

Unfortunately, the combined preshift and postshift data do not permit an unequivocal comparison between achievement with Different and Same cue configurations. However, inspection of the postshift achievement data suggests a clearer interpretation. Regardless of whether an actual (as opposed to potential) shift in cue validities occurred in the task, subject postshift achievement performance was somewhat better when predictions were made from a configuration of Same versus Different cue validities. However, since the difference in achievement between the postshift Same versus postshift Different cue validities over all three postshift trial blocks did not reach statistical significance as indexed by the conservative Sheffe test, this result must be cautiously interpreted.

The second hypothesis concerned the order of shift when an actual shift in cue validities occurred (i.e., Different-Same; Same-Different). It was hypothesized that achievement performance would be better when cue validities shifted from Same to Different (panel 2) than when the validities shifted from Different to Same (panel 1). Again, the data do not confirm the hypothesis. For both the Different-Same and Same-Different conditions, both preshift achievement performance and initial postshift performance were equivalent. However, after the shift in cue validities had occurred, achievement in the Different-Same condition (in which Same cues were used in postshift trials) increased over trials relative to the Same-Different condition (in which Different cues were used in the postshift trials). Again this must be cautiously interpreted as the conservative Sheffe test comparing the postshift data for these two groups was not statistically significant ($p > .10$). However, since for this particular interaction, an a posteriori test, such as the Scheffe test, does not make comparisons of postshift achievement data conditional on preshift performance, it is suggested that direct interpretation of the functions depicted in the plotted interaction is more appropriate for addressing the questions posed in the present study.

With regard to the postshift superiority of the Different-Same versus Same-Different conditions, further insight may be gained by examination of the shift x order x pre-post segments interactions for both the consistency and matching data. These interactions are presented in Figures 4 and 8 respectively. Examination of the function relating consistency and matching to the shift conditions of Different-Same and Same-Different over pre-post segments reveals a seemingly paradoxical result. When the direction of shift is from Same in preshift trials to Different in postshift trials, consistency increases while matching decreases. However, when the shift is in the direction of Different-Same, consistency decreases while matching remains constant.

As Deane, Hammond, and Summers (1972) have pointed out, a high level of matching may be accompanied by a low level of consistency, as well as the converse. The implication of this relationship for the results of the present study seem clear. When subjects shifted to a configuration of Different cue validities after previous exposure to a configuration of Same cue validities, they utilized an inappropriate cue utilization strategy, but did so consistently. On the other hand, when subjects shifted to a configuration of Same cue validities after previous exposure to a configuration of different cue validities, their consistency decreased, reflecting a search for the correct cue utilization scheme over trials.

In summary it appears that when there is either a potential or actual shift in cue validities in an inference task, achievement is better when subjects make predictions in postshift trials based on a configuration of equally valid cues rather than on a configuration of differently valid cues. (The reader is reminded that, although cue validities changed over trials, system predictability remained constant). This finding is contrasted with that reported by Todd and Hammond (1965) in which achievement was significantly better in a task with three differently valid cues (r_{ei} = .77, .52, and .27) than in a task with three equally valid cues (r_{ei} = .50, .50, and .50).

However, there are several critical differences between the Todd and Hammond study and the present study which make direct comparison of results difficult. First of all, Todd and Hammond used geometric cues rather than numeric cues with semantic labels. Secondly, their study involved no potential or actual shift in cue validities in which subjects might alter their cue utilization strategies. And most important, the tolerance limits within blocks for the empirical cue-criterion correlations and cue intercorrelations were higher (i.e., ± .10 to ± .20 than in the present study (i.e., ± .05. Thus it is possible in the Todd and Hammond (1965) study that block to block variation in empirical cue validities may have resulted in a marked deviation from the theoretical set of Different of Same cue validities. This possibility is especially critical when evaluating achievement performane when predictions were made from cues of supposedly "equal" validity.

Furthermore, as Dudycha, Dudycha, and Schmitt (1974) have demonstrated, when cue intercorrelation is left to vary, there is a potential confounding of effects with both system predictability and cue validities. It is also important to note that, in the present study, blocks of cue values for both the Same and Different cue validity configurations were selected so as to maximize the variance of the cue values within each trial. Thus the present study was more sensitive to differences between strategies emphasizing different versus equal weighting strategies.

The question still remains as to why achievement performance in the postshift trials was better for Same versus Different cue validities regardless of whether a potential or actual shift in cue validities took place. That is, postshift achievement performance was better for Same cue validities which were preceded in the preshift trials either by Same cue validities (in which only a potential shift was involved) or by Different cue validities (in which an actual shift in cue validities occurred).

The most parsimonious explanation would be the following. Given that subjects in a decision task are aware of a potential shift, they gravitate toward a strategy of equal weighting of information after a potential shift. Thus regardless of the cue configuration which preceded the potential shift, they adopt a cue utilization strategy in which all information is treated as equally important. This strategy is maintained until a strategy involving different weighting of cues appears to be more appropriate.

Examination of the post-experimental questionnaire data suggest that such might be the case. Subjects were asked to assign numbers to the three cues according to how they thought they weighted the cues to predict the criterion for both the preshift trials and the postshift trials. These numbers ranged from 1-99, with "1" indicating a very low weight and "99" indicating a very high weight. Table 5 presents a summary of these results.

It is apparent that differently valid cues occurring after the shift are weighted more nearly equally than differently valid cues occurring before the shift. This tendency toward equal weighting is most apparent for differently valid cues in the postshift segment preceded by differently valid cues in the preshift segment.

The remaining hypotheses addressed in this study concerned the effectiveness of group discussion as a possible form of feedback in a multiple cue probability learning task. There was no support for the hypothesis that subjects who learned the task with group discussion feedback should demonstrate greater consistency than those who learned on an individual basis. The main effect of group discussion for the consistency index did not reach statistical significance ($p > .10$). Also not supported was the hypothesis that subjects in the group discussion condition should learn the task faster than those in the no group discussion condition. Neither the group discussion x blocks interaction nor the group discussion x pre-post segments x blocks interaction were significant for the achievement data ($p > .10$).

However the significant group discussion x shift x blocks interaction for achievement (Figure 2) did give strong support for the hypothesis that group discussion feedback should enable subjects to detect and track a shift in cue validities over trials relative to no such feedback. Inspection of the function relating achievement to group discussion/no discussion over blocks suggests that group discussion feedback facilitates multiple cue learning in which cue validities shift over trials, relative to no group discussion feedback. On the other hand, when no actual shift in cue validities occurred, group discussion feedback was much less facilitative over trials than no group discussion.

These data suggest that a primary function of group discussion feedback may be to sensitize the members of a group to the possibility of a shift in cue validities. When a shift actually occurred, this form of feedback was facilitative in helping individuals in the group detect and properly adjust to the change.

However, feedback from other group members also has its drawbacks. The data from the present study suggest that group discussion feedback may give erroneous information regarding the possibility of a shift in cue validities, sensitizing the group to perceive a shift when in fact none existed.

Reference to the literature on the effect of different forms of feedback in multiple cue learning is appropriate. Castellan (1974), in commenting on the effect of different types of feedback on multiple cue learning stated that "feedback other than outcome feedback is facilitative to the judgment process if the feedback is appropriate or informative with respect to the true nature of the relation between the cues and criterion or distal variable" (p. 45).

Although Castellan was referring to feedback concerning task properties and subject utilization strategies (e.g., Hammond, Summers, & Deane, 1973) an analogy can be drawn to the present study. Group discussion feedback was effective in sensitizing subjects to a particular property of the task in this experiment: that a shift in cue validities occurred. However, it appears that this form of feedback was not facilitative when a potential but not actual shift in cue validities occurred. Group discussion feedback may indeed be a double-edged sword; its use in similar inference tasks should be regulated accordingly.

CONCLUSIONS

This study attempted to extend the basic multiple cue normative decision-making model by evaluating the effects on multiple cue inference behavior of several "ecologically valid" variables. Thus the effects of group discussion feedback and shifting versus stationary cue validities were evaluated in a meaningful environment. This environment was characterized by moderately high system predictability and three orthogonal stimulus cues with meaningful semantic labels.

The major hypotheses regarding achievement performance with Different versus Same cue validities and with direction of shift of cue validities were not supported. Instead, trends were in the opposite direction of those hypothesized. No conclusions could be drawn from the data concerning the relative superiority of configurations of Different and Same cue validities. However, the data suggested that achievement performance after a potential or actual shift in cue validities was best when predictions were made on the basis of equally weighted cue validities, regardless of the configuration of cues which preceded the potential or actual shift.

Furthermore, group discussion feedback proved to have facilitative effects only when cue validities shifted. When there was no shift in cue validities, achievement performance was better when no group discussion feedback was available.

The present study was characterized by several methodological improvements over previous multiple cue studies. For instance, the effects of shifting cue validities were evaluated with multiple instead of single cues with system predictability held constant. Also, when there was a shift in cue validities, the shift involved a change to a configuration of new validities instead of a simple reversal of existing cue validities. In addition, to better evaluate the effects of the shift, control conditions were included in which the configuration of cue validities remained constant throughout the experiment.

Furthermore the cue values were generated so as to maximize the sensitivity of the task to differences between strategies emphasizing different versus equal weighting emphasis. However, many of these improvements over past methodologies made comparisons of the data in the present study to that reported in previous literature difficult.

Given the results of this study, it is suggested that future research be directed first of all toward evaluation of different cue validity configurations. The present study would have been improved by additional control conditions which would have permitted an unequivocal comparison of equally valid versus differently valid cues both with and without an anticipated shift. Secondly, it is suggested that the potential utility of group discussion as a viable form of feedback in a decision-making context be identified.

The major theoretical focus of the present study was the evaluation of certain variables thought to be characteristic of real-life decision-making behavior. It is hoped that future research in the area will also address itself to the study of those variables which have a direct bearing on real-life decision situations.

REFERENCES

Avant, L. L. & Helson, H. H. Theories of Perception. In B. Wolman (Ed.), Handbook of general psychology. Engelwood Cliffs: Prentice-Hall, 1973.

Bolhius-Bourma, A. and Oostlander, A. M. Multiple cue probability learning: an experiment with circular dimensions. Organizational Behavior and Human Performance, 1972, 7, 383-396.

Brehmer, B. Effects of communication and feedback on cognitive conflict. Scandinavian Journal of Psychology, 1971, 12, 205-216.

Brehmer, B. The effect of cue intercorrelation on interpersonal learning of probabilistic inference tasks. Umea Psychological Reports. No. 73, 1973. (a)

Brehmer, B. Note on the relation between single-cue probability learning and multiple-cue probability learning. Organizational Behavior and Human Performance, 1973, 9, 246-252. (b)

Brehmer, B. Effects of task predictability and cue validity on interpersonal learning of inference tasks involving both linear and nonlinear relations. Organizational Behavior and Human Performance, 1973, 10, 24-46. (c)

Brunswik, E. The conceptual framework of psychology. In R. Carnap and C. Morris (Eds.), International encyclopedia of unified science. Chicago: University of Chicago Press, 1952.

Brunswik, E. Perception and the representative design of psychological experiments. University of California Press: Berkeley, 1956.

Castellan, N. J. Comments on the "Lens Model" equation and the analysis of multiple cue judgment tasks. Psychometrika, 1973, 38, 87-100. (a)

Castellan, N. J. Multiple-cue probability learning with irrelevant cues. Organizational Behavior and Human Performance, 1973, 9, 16-29. (b)

Castellan, N. J. The effect of different types of feedback in multiple-cue probability learning. Organizational Behavior and Human Performance, 1974, 11, 44-66.

Conrad, K. A. Environmental factors in cognitive complexity: performance differences on multiple-cue tasks. Unpublished doctoral dissertation. Iowa State University, 1973.

Davis, J. H. Group performance. New York: Addison-Wesley, 1969.

Davis, J. H. Group decision and social interaction: A theory of social decision schemes. Psychological Review, 1973, 80, 97-125.

Deane, D. H., Hammond, K. R., & Summers, D. A. Acquistion of knowledge in complex inference tasks. Journal of Experimental Psychology, 1972, 92, 20-26.

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.

Dudycha, A. L., Dumoff, M. G., and Dudycha, L. W. Choice behavior in dynamic environments. Organizational Behavior and Human Performance, 1973, 9, 328-338.

Dudycha, A. L., Dudycha, L. W., and Schmitt, N. W. Cue redundancy: some overlooked analytical relationships in MCPL. Organizational Behavior and Human Performance, 1974, 11, 222-234.

Earle, T. C. Interpersonal learning. In L. Rappoport and D Summers (Eds.), Human Judgment and social interaction. New York: Holt, Rinehart, and Winston, 1973.

Einhorn, H. J. Use of nonlinear, noncompensatory models as a function of task and amount of information. Organizational Behavior and Human Performance, 1971, 6, 1-27.

Hammond, K. R. Inductive knowing. In J. Royce and W. Rozeboom (Eds.), The psychology of knowing. New York: Gordon and Breach, 1972.

Hammond, K. R., Summers, D. A., and Deane, D. H. Negative effects of outcome-feedback in multiple-cue probability learning. Organizational Behavior and Human Performance, 1973, 9, 30-34.

Hoffman, P. J. and Blanchard, W. A. A study of the effects of varying amounts of predictor information on judgment. Oregon Research Institute Bulletin, 1961.

Hursch, C., Hammond, K. R., and Hursch, J. L. Some methodological considerations in multiple cue probability studies. Psychological Review, 1964, 71, 42-60.

Kirk, R. E. Experimental design: procedures for the behavioral sciences. Belmont, Cal.: Brooks/Cole, 1968.

Knowles, B. A., Hammond, K. R., Stewert, T. R., and Summers, D, A. Positive and negative cue redundancy in multiple cue probability tasks. Journal of Experimental Psychology, 1971, 90, 157-159.

Miller, P. McM. Do labels mislead? A multiple cue study within the framework of Brunswik's probabilistic functionalism. Organizational Behavior and Human Performance, 1971, 6, 480-500.

Muchinsky, P. M. The influence of a suppressor variable on multiple cue probability learning. Unpublished Doctoral Dissertation. Purdue University, 1973.

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

Naylor, J. C. and Schenck, A. A revised summary of the multiple cue model and its related performance indices. Research Paper No. 2. Ohio State University, 1964.

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

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

Nystedt, L. and Magnusson, D. Cue relevance and feedback in a clinical prediction task. Organizational Behavior and Human Performance, 1973, 9, 100-109.

Oskamp, S. Overconfidence in case-study judgments. *Journal of Consulting Psychology*, 1965, *29*, 261-265.

Peterson, C. R., Hammond, K. R. and Summers, D. A. Multiple probability learning with shifting weights of cues. *American Journal of Psychology*, 1965, *78*, 660-663.

Schenck, A., and Naylor, J. C. Some data concerning performance indices based upon the multiple regression model when applied in a standard multiple cue situation. Paper presented at meetings of the Midwestern Psychological Association, Chicago, 1965.

Schmitt, N. W. and Dudycha, A. L. Negative and positive cue redundancy in multiple cue probability learning. Paper presented at Midwestern Psychological Association Convention. Chicago, Ill. May, 1973.

Summers, D. A. Adaptation to change in multiple probability tasks. *American Journal of Psychology*, 1969, *82*, 235-240.

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

Tolman, E. C. Eulogy. In K. R. Hammond (Ed.), *The psychology of Egon Brunswik*. New York: Holt, Rinehart, and Winston, Inc., 1966.

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

Uhl, C. Learning interval concepts: I. Effects of differences in stimulus weights. *Journal of Experimental Psychology*, 1963, *66*, 264-273.

Wherry, R. J., Sr., Naylor, J. C., Wherry, R. J., Jr., and Fallis, R. F. Generating multiple samples of multivariate data with arbitrary population parameters. *Psychometrika*, 1965, *30*, 303-313.

Young, D. L. Team performance as a function of task structure and work structure. Unpublished doctoral dissertation. Iowa State University, 1973.

ACKNOWLEDGMENTS

I would like to thank everyone who has assisted me in the completion of this dissertation and in my years of graduate study leading to it. This includes many faculty, fellow graduate students, and friends whose names are too numerous to mention in such a small space.

However, special thanks are due to several individuals. First of all, I would like to extend the warmest feeling of appreciation that I can to my major professor, Dr. Paul M. Muchinsky, whose patience, support, and good advice made the seemingly impossible task of a dissertation a reality for me. Likewise, a special word of thanks is extended to Dr. Leroy Wolins, my co-chairman, for his expert advice on the analysis and interpretation of the data. Both of these individuals contributed enormously in their own unique way to the completion of this work.

I would also like to thank Dr. Lloyd L. Avant, Dr. Donald H. Schuster, and Dr. Harold W. Davey for serving as members of my doctoral committee and for their careful reading of the dissertation and helpful comments. Thanks are also due to Dr. Jan P. Wijting and Dr. Terry L. Dickinson, who, as former members of my masters and doctoral committees, aided greatly in my graduate education.

Finally, I would like to thank two of the most wonderful people I know--my mother and father. Through their love, understanding, encouragement, and support I have been able to complete this endeavor.

APPENDIX A.

Factor Structure, Theoretical and Empirical Correlations and Cue and Criterion Values for Blocks of Same and Different Cue Validities

Table A-1

Block one of Same Cue Validities

	Factor Structure					Variable Correlations[1]			
	I	II	III	IV		Y	Cue 1	Cue 2	Cue 3
1	.999	.000	.000	.000	Y	1.000	.488	.543	.549
2	.500	.500	.707	.000	1	.499	1.000	-.026	-.016
3	.500	-.500	.000	.707	2	.499	.000	1.000	-.031
4	.500	.500	-.704	.000	3	.499	.000	.000	1.000

Cue and Criterion Values

Trial	Mid-Term Exam Score	Mid-Term Essay Score	Mid-Term Lab Score	Final Exam Score
001	65	60	71	71
002	68	40	47	49
003	48	66	68	65
004	54	44	61	49
005	42	48	46	39
006	42	47	57	50
007	47	59	54	50
008	37	48	58	51
009	60	58	54	54
010	60	56	38	49
011	68	69	46	65
012	42	46	40	39
013	37	62	38	47
014	67	52	59	61
015	49	45	46	45
016	56	54	50	54
017	54	38	65	45
018	60	41	41	47
019	54	57	17	35
020	54	27	50	38
021	23	58	54	43
022	43	34	55	42
023	52	44	46	49
024	29	56	51	43
025	53	49	38	51

[1]Theoretical cue-criterion correlations are located in the lower triangular matrix. Empirical cue-criterion correlations are located in the upper triangular matrix.

Table A-2

Block Two of Same Cue Validities

	Factor Structure				Variable Correlations[1]				
	I	II	III	IV		Y	Cue 1	Cue 2	Cue 3
1	.999	.000	.000	.000	Y	1.000	.462	.515	.532
2	.500	.500	.707	.000	1	.499	1.000	-.025	-.030
3	.500	-.500	.000	.707	2	.499	.000	1.000	-.043
4	.500	.500	-.707	.000	3	.499	.000	.000	1.000

Cue and Criterion Values

Trial	Mid-Term Exam Score	Mid-Term Essay Score	Mid-Term Lab Score	Final Exam Score
026	63	37	41	47
027	43	46	59	52
028	41	51	47	53
029	63	70	55	72
030	41	44	64	51
031	53	62	38	53
032	50	32	55	50
033	52	65	46	55
034	60	58	67	73
035	37	49	57	49
036	56	36	66	53
037	47	56	54	53
038	39	56	46	45
039	54	43	46	41
040	67	44	58	59
041	49	59	48	53
042	60	49	50	58
043	60	40	50	47
044	43	62	59	54
045	50	46	57	44
046	51	60	52	56
047	54	60	57	59
048	43	40	54	43
049	45	57	75	68
050	38	49	43	30

[1]Theoretical cue-criterion correlations are located in the lower triangular matrix. Empirical cue-criterion correlations are located in the upper triangular matrix.

Table A-3

Block Three of Same Cue Validities

	Factor Structure					Variable Correlations[1]			
	I	II	III	IV		Y	Cue 1	Cue 2	Cue 3
1	.999	.000	.000	.000	Y	1.000	.515	.516	.522
3	.500	.500	.707	.000	1	.499	1.000	.005	.010
3	.500	.500	.000	.707	2	.499	.000	1.000	-.012
4	.500	.500	.707	.000	3	.499	.000	.000	1.000

Cue and Criterion Values

Trial	Mid-Term Exam Score	Mid-Term Essay Score	Mid-Term Lab Score	Final Exam Score
051	57	57	59	70
052	67	51	32	56
053	42	52	55	50
054	37	67	31	42
055	49	52	43	45
056	66	44	52	57
057	53	54	45	62
058	54	41	51	48
059	63	55	41	54
060	62	48	37	53
061	62	48	61	72
062	38	37	57	45
063	58	68	60	68
064	43	68	38	52
065	42	41	57	43
066	37	44	46	44
067	44	31	35	30
068	39	52	54	56
069	39	56	54	48
070	57	59	49	55
071	55	37	38	41
072	53	76	58	70
073	56	39	50	51
074	55	44	72	57
075	52	55	42	41

[1]Theoretical cue-criterion correlations are located in the lower triangular matrix. Empirical cue-criterion correlations are located in the upper triangular matrix.

Table A-4

Block One of Different Cue Validities

	Factor Structure					Variable Correlations[1]			
	I	II	III	IV	Y	Cue 1	Cue 2	Cue 3	
1	.999	.000	.000	.000	Y	1.000	.787	.420	.114
2	.760	-.400	.000	.512	1	.759	1.000	-.019	-.045
3	.400	.760	-.400	.320	2	.399	.000	1.000	-.024
4	.100	.190	.461	.861	3	.099	.000	.000	1.000

Cue and Criterion Values

Trial	Mid-Term Exam Score	Mid-Term Essay Score	Mid-Term Lab Score	Final Exam Score
001	57	37	37	46
002	56	50	40	48
003	61	30	41	51
004	43	72	48	53
005	52	49	55	44
006	57	39	57	51
007	64	50	65	65
008	56	48	47	60
009	56	58	31	55
010	34	47	64	37
011	50	61	42	55
012	49	51	64	52
013	56	63	68	59
014	31	46	61	31
015	63	48	74	61
016	49	37	59	48
017	68	61	47	66
018	56	40	58	51
019	51	63	56	64
020	54	50	41	43
021	33	53	38	35
022	46	32	59	41
023	54	41	42	52
024	55	56	59	54
025	62	56	51	67

[1]Theoretical cue-criterion correlations are located in the lower triangular matrix. Empirical cue-criterion correlations are located in the upper triangular matrix.

Table A-5

Block Two of Different Cue Validities

	Factor Structure					Variable Correlations[1]			
	I	II	III	IV		Y	Cue 1	Cue 2	Cue 3
1	.999	.000	.000	.000	Y	1.000	.778	.444	.135
2	.760	.000	.000	.512	1	.759	1.000	.030	.049
3	.400	.000	-.400	.320	2	.399	.000	1.000	-.019
4	.100	.000	.461	.861	3	.099	.000	.000	1.000

Cue and Criterion Values

Trial	Mid-Term Exam Score	Mid-Term Essay Score	Mid-Term Lab Score	Final Exam Score
026	38	31	53	36
027	52	50	45	60
028	46	56	38	54
029	41	49	31	38
030	64	42	50	57
031	68	44	28	61
032	55	55	40	59
033	44	42	48	39
034	44	42	62	49
035	49	34	34	32
036	54	40	54	49
037	41	56	46	48
038	55	35	55	41
039	54	60	41	57
040	69	40	53	60
041	53	49	65	53
042	43	47	50	52
043	60	46	53	55
044	38	49	58	37
045	39	49	45	46
046	44	39	54	43
047	52	39	46	48
048	38	60	54	35
049	52	45	54	52
050	73	63	66	78

[1]Theoretical cue-criterion correlations are located in the lower triangular matrix. Empirical cue-criterion correlations are located in the upper triangular matrix.

Table A-6

Block Three of Different Cue Validities

	Factor Structure					Variable Correlations[1]			
	I	II	III	IV		Y	Cue 1	Cue 2	Cue 3
1	.999	.000	.000	.000	Y	1.000	.808	.375	.098
2	.760	.000	.000	.512	1	.759	1.000	.019	.019
3	.400	.000	-.406	.320	2	.399	.000	1.000	.032
4	.100	.000	.461	.861	3	.099	.000	.000	1.000

Cue and Criterion Values

Trial	Mid-Term Exam Score	Mid-Term Essay Score	Mid-Term Lab Score	Final Exam Score
051	37	65	54	42
052	59	69	53	59
053	53	48	56	49
054	53	60	66	55
055	30	49	48	36
056	62	56	44	58
057	56	59	42	54
058	58	51	56	62
059	44	49	50	47
060	50	31	52	41
061	36	69	43	48
062	51	46	48	55
063	41	65	47	44
064	52	56	66	60
065	48	64	54	49
066	38	39	48	35
067	50	35	61	48
068	41	35	46	32
069	44	63	60	52
070	54	43	39	54
071	52	35	63	46
072	54	41	45	58
073	48	53	56	54
074	37	43	39	42
075	28	47	66	36

[1]Theoretical cue-criterion correlations are located in the lower triangular matrix. Empirical cue-criterion correlations are located in the upper triangular matrix.

APPENDIX B.

Experimental Instructions

Instructions for Group Discussion Condition

In this experiment I am interested in studying how people make predictions about certain events on the basis of several sources of information. Many individuals are in occupations in which they must make predictions about some state of affairs (called the "criterion") based on certain information (called "cues") which is somehow related to the criterion.

For example, a credit manager's job requires him to use information from such cues as a person's average monthly debt and number of creditors to predict that person's credit risk. As another example, the weather forecaster's job requires that he predict what the weather will be on the basis of such cues as temperature and barometric pressure. However, in both of these examples perfect prediction is seldom obtained. Thus two credit applicants may have the same amount of debt and the same number of creditors, but be different credit risks. Likewise, on different days the same temperature and barometric pressure may result in different weather conditions. Nonetheless, both the credit manager and the weatherman become skilled at coming very close to the best possible prediction based on the information with which they have to work.

In this experiment you will be asked to predict students' scores on a final examination from a course (the criterion) based on information from three sources (the cues): 1) a mid-term exam score; 2) a mid-term essay score; and 3) a mid-term lab score. The final exam for this course is comprehensive; that is, it reflects the same skills and knowledge as contained in the mid-term exam, the mid-term essay, and the mid-term lab.

Both the cues and the criterion will be represented as two digit numbers. It is important to note that these scores are taken from a course taught at another college which uses a slightly different grading system than is used at Iowa State University. That is, both the cues and the criterion can range from 10 to 90. A score of "50" represents an "average" score. Scores above 50 are considered to represent above average scores; scores below 50 are considered to represent below average scores.

The procedure to be followed in the experiment is this:

1) You will be shown on the screen three two-digit numbers for each trial. These numbers represent the cues of: 1) a mid-term exam score; 2) a mid-term essay score; and 3) a mid-term lab score.

TRIAL	MID-TERM EXAM SCORE	MID-TERM ESSAY SCORE	MID-TERM LAB SCORE	FINAL EXAM SCORE
1	52	45	64	

2) On the basis of these cue values you will make a prediction of the final exam score. You will write down your prediction on the response sheet in the space corresponding to the appropriate trial number.

3) After you have recorded your prediction of the final exam score, you will be shown both the cues for that trial and the best possible prediction of the final exam score based on the cues for that trial.

TRIAL	MID-TERM EXAM SCORE	MID-TERM ESSAY SCORE	MID-TERM LAB SCORE	FINAL EXAM SCORE
1	52	45	64	48

4) We will then begin the next trial. This procedure will be repeated until we have completed a total of 150 trials.

Thus your prediction of the final exam score (the criterion) will be a positive, two-digit number between 10 and 90.

On each trial, consisting of three cues and a criterion, you should make your predictions on the basis of the cue values for a particular trial as well as on the basis of your past experience with cue-criterion pairings from previous trials. On each trial the criterion value given is to be regarded as the best possible prediction or estimate for that trial. Because of the nature of the task, perfect prediction is seldom achieved. However, as the experiment proceeds, you should become more accurate in your predictions. You are not expected to be correct all the time, but try to do your best.

At different times during the experiment you will be given brief rest periods. During these periods you should talk things over with the other individuals in the group and discuss different strategies or ways of using the cue values to make a prediction of the criterion. However, you must make all your predictions on an individual basis once the experiment has resumed, and not in consultation with the other members of the group.

During the last 75 trials of the experiment you will be making predictions of a final exam score for a different course than in the first 75 trials. As before the cues will be: 1) a mid-term exam score; 2) a mid-term essay score; and 3) a mid-term lab score. As before the criterion will be the final exam score from the same course. The scores will range from 10 to 90; a score of 50 still represents an average score.

However, the relationship between the cues and the criterion may not be the same as in the first part of the experiment. This is due to the fact that different instructors may disagree on how inportant a mid-term exam score, a mid-term essay score, and a mid-term lab score are in contributing to the score of a final comprehensive examination. It is possible that the cue values should be used differently to predict the criterion for the second course, but not necessarily so.

Before we begin the experiment, you will receive five practice trials. For each of these practice trials you are to observe the cue values, make a predicton, write the prediction on the answer sheet, and observe the best possible prediction of the criterion value based in the cues for that trial.

Instructions for No Group Discussion Condition

In this experiment I am interested in studying how people make predictions about certain events on the basis of several sources of information. Many individuals are in occupations in which they must make predictions about some state of affairs (called the "criterion") based on certain information (called "cues") which is somehow related to the criterion.

For example, a credit manager's job requires him to use information from such cues as a person's average monthly debt and number of creditors to predict that person's credit risk. As another example, the weather forecaster's job requires that he predict what the weather will be on the basis of such cues as temperature and barometric pressure. However, in both of these examples perfect prediction is seldom obtained. Thus two credit applicants may have the same amount of debt and the same number of creditors, but be different credit risks. Likewise, on different days the same temperature and barometric pressure may result in different weather conditions. Nonetheless, both the credit manager and the weatherman become skilled at coming very close to the best possible prediction based on the information with which they have to work.

In this experiment you will be asked to predict students' scores on a final examination from a course (the criterion) based on information from three sources (the cues): 1) a mid-term exam score; 2) a mid-term essay score; and 3) a mid-term lab score. The final exam for this course is comprehensive; that is, it reflects the same skills and knowledge as contained in the mid-term exam, the mid-term essay, and the mid-term lab.

Both the cues and the criterion will be represented as two digit numbers. It is important to note that these scores are taken from a course taught at another college which uses a slightly different grading system than is used at Iowa State University. That is, both the cues and the criterion can range from 10 to 90. A score of "50" represents an "average" score. Scores above 50 are considered to represent above average scores; scores below 50 are considered to represent below average scores.

The procedure to be followed in the experiment is this:

1) You will be shown on the screen three two-digit numbers for each trial. These numbers represent the cues of: 1) a mid-term exam score; 2) a mid-term essay score; and 3) a mid-term lab score.

TRIAL	MID-TERM EXAM SCORE	MID-TERM ESSAY SCORE	MID-TERM LAB SCORE	FINAL EXAM SCORE
1	52	45	64	

2) On the basis of these cue values you will make a prediction of the final exam score. You will write down your prediction on the response sheet in the space corresponding to the appropriate trial number.

3) After you have recorded your prediction of the final exam score, you will be shown both the cues for that trial and the best possible prediction of the final exam score based on the cues for that trial.

TRIAL	MID-TERM EXAM SCORE	MID-TERM ESSAY SCORE	MID-TERM LAB SCORE	FINAL EXAM SCORE
1	52	45	64	48

4) We will then begin the next trial. This procedure will be repeated until we have completed a total of 150 trials.

Thus your prediction of the final exam score (the criterion) will be a positive, two-digit number between 10 and 90.

On each trial, consisting of three cues and a criterion, you should make your predictions on the basis of the cue values for a particular trial as well as on the basis of your past experience with cue-criterion pairings from previous trials. On each trial the criterion value given is to be regarded as the best possible prediction or estimate for that trial. Because of the nature of the task, perfect prediction is seldom achieved. However, as the experiment proceeds, you should become more accurate in your predictions. You are not expected to be correct all the time, but try to do your best.

At different times during the experiment you will be given brief rest periods. During these periods you should think over your individual strategy or way of using the cue values to make a prediction of the criterion. It is very important that you do not discuss strategies with the other individuals in the room.

During the last 75 trials of the experiment you will be making predictions of a final exam score for a different course than in the first 75 trials. As before the cues will be: 1) a mid-term exam score; 2) a mid-term essay score; and 3) a mid-term lab score. As before the criterion will be the final exam score from the same course. The scores will range from 10 to 90; a score of 50 still represents an average score.

However, the relationship between the cues and the criterion may not be the same as in the first part of the experiment. This is due to the fact that different instructors may disagree on how inportant a mid-term exam score, a mid-term essay score, and a mid-term lab score are in contributing to the score of a final comprehensive examination. It is possible that the cue values should be used differently to predict the criterion for the second course, but not necessarily so.

Before we begin the experiment, you will receive five practice trials. For each of these practice trials you are to observe the cue values, make a predicton, write the prediction on the answer sheet, and observe the best possible prediction of the criterion value based in the cues for that trial.

Debriefing

This experiment investigated how individuals make decisions in a meaningful context. The predictions you were asked to make are very similar to those that might be made in a real setting. The hypotheses examined in this experiment are concerned with the ways people make decisions or predictions based on different kinds of numeric information. No other "secret" or "hidden" hypotheses were examined.

Many other students will be serving as subjects for this experiment. Therefore, I ask you not to discuss the experiment with other students, since this may bias their feelings about the experiment. That way I can be certain that these subjects are responding honestly to the experiment and not to any preconceived notions about the experiment.

This experiment is part of my doctoral dissertation. I thank you for your time in participating in the experiment. At this time I cannot be more specific about the purpose of the experiment. However, after I have collected and analyzed all the data I will be glad to send you a summary of the results of the study. If you would like this information, please give me your name and address.

Again, thank you for your participation.

APPENDIX C.

Response Sheet

Response Sheet

1. _____	26. _____	51. _____	76. _____	101. _____	126. _____
2. _____	27. _____	52. _____	77. _____	102. _____	127. _____
3. _____	28. _____	53. _____	78. _____	103. _____	128. _____
4. _____	29. _____	54. _____	79. _____	104. _____	129. _____
5. _____	30. _____	55. _____	80. _____	105. _____	130. _____
6. _____	31. _____	56. _____	81. _____	106. _____	131. _____
7. _____	32. _____	57. _____	82. _____	107. _____	132. _____
8. _____	33. _____	58. _____	83. _____	108. _____	133. _____
9. _____	34. _____	59. _____	84. _____	109. _____	134. _____
10. _____	35. _____	60. _____	85. _____	110. _____	135. _____
11. _____	36. _____	61. _____	86. _____	111. _____	136. _____
12. _____	37. _____	62. _____	87. _____	112. _____	137. _____
13. _____	38. _____	63. _____	88. _____	113. _____	138. _____
14. _____	39. _____	64. _____	89. _____	114. _____	139. _____
15. _____	40. _____	65. _____	90. _____	115. _____	140. _____
16. _____	41. _____	66. _____	91. _____	116. _____	141. _____
17. _____	42. _____	67. _____	92. _____	117. _____	142. _____
18. _____	43. _____	68. _____	93. _____	118. _____	143. _____
19. _____	44. _____	69. _____	94. _____	119. _____	144. _____
20. _____	45. _____	70. _____	95. _____	120. _____	145. _____
21. _____	46. _____	71. _____	96. _____	121. _____	146. _____
22. _____	47. _____	72. _____	97. _____	122. _____	147. _____
23. _____	48. _____	73. _____	98. _____	123. _____	148. _____
24. _____	49. _____	74. _____	99. _____	124. _____	149. _____
25. _____	50. _____	75. _____	100. _____	125. _____	150. _____

APPENDIX D.

Post-Experimental Questionnaire

1. Please list in order (as best you can) the strategy or strategies that you used in making predictions of the criterion score from the three cue scores.

____

____

____

____

2. Did you find that the opportunity to talk things over or discuss strategies of prediction with the other individuals in the group helped you to improve the accuracy of your predictions of the criterion scores? (circle one)

YES NO DON'T KNOW

3. Did you feel that the relationship between the cues and the criterion in the second half of the experiment was different than in the first half of the experiment? (circle one)

YES NO DON'T KNOW

4. If you answered YES to number 3, did the opportunity to talk things over or discuss strategies of prediction with the other individuals in the group help you to decide that the cue-criterion relationships had changed? (circle one)

YES NO DON'T KNOW

5. The task consisted of 150 trials. If you began to lose interest in the task before it was finished, at about what trial did you begin to lose interest?

Trial_____

6. In general, I found this task to be (circle one):

a) very easy

b) moderately easy

c) neither easy nor difficult

d) moderately difficult

e) very difficult

7. Of the following courses, from what type of course do you think the cues were taken for the first half of the experiment (that is, the first 75 trials)? (circle one)

a) Agriculture b) Education c) Engineering

d) Home Economics e) Natural Science f) Social Science

8. Of the following courses, from what type of course do you think the cues were taken for the last half of the experiment (that is, the last 75 trials). (circle one)

a) Agriculture b) Education c) Engineering

d) Home Economics e) Natural Science d) Social Science

9. Please indicate how you weighted the cues to predict the criterion, or how important you thought each cue was in predicting the criterion, for the first part of this experiment (the first 75 trials). Do this by rating each cue from "1" (I weighted it very little) to "99" (I weighted it very much).

_____Mid-term Exam Score _____Mid-term Essay Score _____Mid-term Lab Score

10. Please indicate how you weighted the cues to predict the criterion, or how important you thought each cue was in predicting the criterion, for the second part of this experiment (the last 75 trials). Do this by rating each cue from "1" (I weighted it very little) to "99" (I weighted it very much).

_____Mid-term Exam Score _____Mid-term Essay Score _____Mid-term Lab Score

11. I am interested in how you think the three cues used in this experiment should be weighted for different types of college courses. In other words, how important is each cue for predicting the criterion. Please indicate how you think these cues <u>should</u> be weighted to predict the final exam score for the courses listed below. Do this by rating each cue from "1" (should be weighted very little) to "99" (should be weighted very much).

a) Agriculture: _____Mid-term Exam Score _____Mid-term Essay Score _____Mid-term Lab Score

b) Education: _____Mid-term Exam Score _____Mid-term Essay Score _____Mid-term Lab Score

c) Engineering: _____Mid-term Exam Score _____Mid-term Essay Score _____Mid-term Lab Score

d) Home Economics _____Mid-term Exam Score _____Mid-term Essay Score _____Mid-term Lab Score

e) Natural Science _____Mid-term Exam Score _____Mid-term Essay Score _____Mid-term Lab Score

f) Social Science _____Mid-term Exam Score _____Mid-term Essay Score _____Mid-term Lab Score