24125

NAME OF AUTHOR/*NOM DE L'AUTEUR* Larry Frederick Saidman

TITLE OF THESIS/*TITRE DE LA THÈSE* Task Differences and Configural Effects in Human Judgments

UNIVERSITY/*UNIVERSITÉ* University of Alberta

DEGREE FOR WHICH THESIS WAS PRESENTED/ *GRADE POUR LEQUEL CETTE THÈSE FUT PRÉSENTÉE* Master of Arts

YEAR THIS DEGREE CONFERRED/*ANNÉE D'OBTENTION DE CE GRADE* 1975

NAME OF SUPERVISOR/*NOM DU DIRECTEUR DE THÈSE* Dirk L. Schaeffer, PhD.

DATED/*DATÉ* May 5/75 SIGNED/*SIGNÉ* Larry Saidman

PERMANENT ADDRESS/*RÉSIDENCE FIXE* R.R. 1, Spruce Grove, Alberta.

THE UNIVERSITY OF ALBERTA

TASK DIFFERENCES AND CONFIGURAL EFFECTS IN HUMAN JUDGMENTS

by

© LARRY F. SAIDMAN

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF ARTS

DEPARTMENT OF PSYCHOLOGY

EDMONTON, ALBERTA

SPRING, 1975

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled TASK DIFFERENCES AND CONFIGURAL EFFECTS IN HUMAN JUDGMENTS, submitted by LARRY F. SAIDMAN in partial fulfilment of the requirements for the degree of Master of Arts.

....
Supervisor

....

....

March 15, 1975

ABSTRACT

This study asks whether task differences play a part in the number and the magnitude of interactions or configural effects in judgments, using the Analysis of Variance model and the ω^2 statistic (Hays, 1963). It is hypothesized that intuitive judgments would produce a greater proportion of significant interaction effects in the judges' responses than would analytic judgments.

After a pilot study to determine what tasks were judged as most analytic, and what tasks were seen as most intuitive (selected tasks were success in university, and sociability, respectively), fifty introductory psychology students were asked to make judgments with regard to each of these tasks after being given information about hypothetical subjects' scores (high or low) on each of five cues. A completely crossed factorial design resulted in 32 different cue configurations for each judge; each was rated on a 9-point scale, and each repeated twice (in a random order), resulting in 64 stimulus configurations. An ANOVA was done for each subject for each task.

Results, based on average proportion of variance accounted for by each effect (ω^2), indicated that while the subjects in the analytic task condition showed significantly greater reliance on main effects than did those in the intuitive task condition, there were no significant differences in the number of interactions in the Js utilization of the cues. Judgments in the intuitive task condition were found to be less reliable.

A second study was conducted using a new set of cues which were

selected through another pilot study. Only 13 judges (volunteers) were used, and no significant differences were found. However, when the data for the two studies were combined, it was discovered that judges who based their judgments on four or more cues utilized interactions to a significantly greater degree than judges who based their judgments primarily on one cue. Such findings are discussed as giving support to the ANOVA technique as a measure of cognitive complexity.

A comparison of the data with 4 other studies was seen as giving support to task differences in J's use of configural components.

ACKNOWLEDGEMENT

I would like to express my appreciation to Dr. Dirk L. Schaeffer, from the Psychology Department of the University of Alberta, for his role as a friend and supervisor. Without his patience, inspiration, and occasional criticism, this thesis would not have been possible.

LARRY F. SAIDMAN

TABLE OF CONTENTS

Page

INTRODUCTION AND REVIEW 1

Clinical versus Actuarial Prediction 1

Human Judgment and Mathematical Models 7

Two Linear Models 7

i The Hammond Model 7

ii The Hoffman Model 13

Evidence for Linearity in Judgment 17

Studies Showing Nonlinearity in Judgments 25

Analytic and Intuitive Judgments 37

RESEARCH 48

Pilot Study 1 48

Method 48

Results 49

Discussion 51

Experiment 1 52

Method 52

Results 54

Discussion 60

Pilot Study 2 62

Method 62

Results and Discussion 63

Page

Experiment 2 64

Method 64

Results 64

Discussion of Both Experiments 66

Conclusion 75

***

APPENDIX. Sample of 1st three protocols, Exp. 1 77

REFERENCES 78

LIST OF TABLES

Table	Description	Page
1	Results of Pilot Study 1.	50
2	Mean Proportion of Variance Explained by Significant Main Effects and Interactions for Analytic and Intuitive Tasks: Experiment 1.	55
3	Number of Significant 2-cue, 3-cue, and 4-cue interactions for each of the two tasks.	57
4	Selected 2-cue Interactions: Comparisons Between the Two Tasks.	59
5	Mean Proportion of Variance Explained by Significant Main Effects and Interactions for Analytic and Intuitive Tasks: Experiment 2.	65
6	Comparison of the Results of the Two Present Experiments with those of Schaeffer and Jackson (1970), Hoffman, Slovic, and Rorer (1968), Schaeffer and Saidman (1971), and Slovic (1969).	71

LIST OF FIGURES

Figure		Page
1	Brunswik's Lens Model .	10

CLINICAL VERSUS ACTUARIAL PREDICTION

The roots of much of the work done on human judgment lies in the clinical versus actuarial prediction controversy, initiated by Lundberg (1941) and Allport (1937, 1965), and later developed and researched by Meehl (1954) and Sarbin, Taft, and Bailey (1960).

There have been two different modes of research aimed at investigating types of prediction or judgment; one deals with accuracy of predictions, while the other deals with hypothesized processes underlying the judgments themselves. This study concerns itself chiefly with the latter. The subject of interest in the present paper is the difference, if any, in the process used by judges in making actuarial and clinical predictions, or, in analytic and intuitive judgments. Rather than attempting to explain the exact nature of analytic or actuarial, and intuitive or clinical judgment processes, this study purports to see if any task differences affecting the judgment process can be uncovered.

A second concern of this paper involves the sensitivity of the Analysis of Variance approach (Hoffman, 1960; 1968) in explaining task differences in terms of the strength of the interaction components. Evidence to support the use of such a model is presented in a later section.

Lundberg (1941) has claimed that all prediction is truly actuarial, since it is a result of laws that are based upon previous experience--

even though these laws may be applied only to very specific cases. Such a statement is equivalent to saying that all prediction uses inductive reasoning--a point that few clinicians will dispute. Actuarial prediction, according to Meehl (1954) and Sarbin (1941) is given a somewhat narrower definition. To Meehl, an actuarial prediction is one that can be made quickly through the use of a statistician or clerk who has received no clinical training. In such a case, the laws would have to be fairly general ones, since it would take a clinician to realize the specific cues in which more specific laws would apply. Allport (1935, 1965) sees the main issues as being between two classes of science: nomothetic (that which deals with general laws), which corresponds to the actuarial or statistical mode of prediction, and ideographic (dealing with the structured pattern in a single individual), which is closer to the clinical type of prediction. According to Allport, the nomothetic method is related to explanation, and the ideographic method to understanding. Prediction based on general or dimensional information is called actuarial, and is accurate for many purposes; i.e., predicting the number of people that will meet a specified criterion (e.g., commit suicide). But it provides no basis for predicting the behavior of a specific individual. Only by understanding the individual's personality and present and future circumstances can we make such predictions.

Sarbin, Taft, and Bailey (1960) have developed Lundberg's earlier claim (i.e., that all prediction is actuarial), and have attempted to get rid of the distinction between nomothetic and ideographic, or actuarial and clinical prediction. Accordingly, clinical inference can be reduced to statistical inference. They state that:

> "No adequate study of clinical inference demonstrates a degree of validity which exceeds the validity of straightforward statistical or actuarial prediction . . . This lack exists in spite of the clinician's supposedly greater range of information and wider field of opportunity to integrate and evaluate the data concerning the object of the inference. The clinician has not been able to improve on actuarial prediction even though in principle he has a limitless amount of information on which to base his judgments." (p. 3)

Allport's (1965) answer to such a statement is that it only indicates that the clinician has to be better trained in understanding the individual's whole personality so that he can make better predictions. Also, he criticizes much of this research on the grounds that the clinicians are forced to make predictions about events that may not be relevant in the specific case. Hence, he criticizes the criterion used in such studies.

Sarbin, Taft, and Bailey (1960) explain all prediction through a process of logical inference, where premises are constructed and are followed by conclusions which are related to the proper use of inferential forms. They also claim that statistical and so-called clinical inference is based on probabilities, even though all-or-none inferences are sometimes used in making a particular decision.

Much of the reason underlying the development of this dispute has been that clinicians have been unable to report the datum or configuration of data which led to a particular diagnosis. For Hammond (1955) it is this inability of the clinician to communicate the basis of his decisions that is the starting point for the analysis of the clinical method; noncommunicability directly reflects clinical behavior. The tasks that Sarbin et al (1960), Hammond (1955), Hoffman (1960), Meehl (1950, 1954, 1960), and others have set for themselves is to explain the clinician's utilization of data in making predictions. Based on Brunswik's "lens model" (1955), devised to explain various phenomena in the area of perception, Hammond (1955) has used a multiple-regression analysis that attempts to approximate the judgment process used by each clinician in making judgments across subjects.

Such a multiple-regression model seems to offer an alternative to the Sarbin, Taft, and Bailey attempt to equate clinical prediction with predictions using a machine that utilizes nothing but frequency distributions. The Hammond model is described briefly in the next section.

Goldberg (1968, 1970), using the Hoffman (1968) Analysis of Variance (ANOVA) development of Hammond's multiple-regression technique transferred the inference process of each of the clinicians he studied into a mathematical model. Since he found little discrepancy between the clinician's actual judgments and those judgments using the judge's best fitting linear equation, Goldberg claims that a linear mathematical model can be developed for each judge that will accurately describe the judge's inference process. Using such a model, Goldberg postulates, will improve the accuracy of each clinician, because it would not be subject to the human frailties (e.g., forgetting to appropriately weight a cue) that a clinician would. In this way, the clinician's judgmental unreliability is separated from his judgmental strategy. The purpose of humans in the decision making process is only to discover or identify new cues which will improve predictive accuracy, and to construct new systematic procedures for combining predictors in increasingly optimal ways (Goldberg, 1970, p. 423). Such a procedure, of course, is still at odds with Allport's ideographic science, since it does not predict the behavior of an individual client; only of the individual therapists in making predictions about several clients. Also, it does not take into account any nonlinear or configural modes that the clinician might use for combining data; although it does have a nonlinear component that acknowledges their existence. Goldberg, however, claims that it is much more important to eliminate the random error component in human judges than to capture valid nonlinear variance in the judge's decision process.

Meehl (1950) feels that the configurality issue is an important one. He gives a hypothetical illustration of a configural analysis that would prove important in a task of clinical judgment from a psychological test. There are two dichotomous items on a test; schizophrenics and normals are compared, and it is found that there is no difference between normals and schizophrenics in their tendency to answer the items true or false. But, it is found that normals answer the two items in the same way, but schizophrenics always give opposite answers to the same items. If the answers to the items were considered one at a time, they would have no predictive power; but considering them configurally changes this. Meehl proposes the construction of personality tests which make use of configural relationships, since they would be more subtle tests than those which are strictly linear.

Meehl (1959) also feels that the presumed ability of the clinician to react on the basis of higher configural relations is one of the factors which might favor clinical intuition over traditional statistical methods of combining data. Most of the studies seeking such relationships have employed a linear model, and have used tasks which require the judge to make a judgment through combining various test protocols. The majority of these studies claim that the linear model is adequate at pulling out a large proportion of the judgmental variance; others have shown higher residuals, some of which may represent nonlinear or configural components. The following section will describe some of these studies and look for reasons for the discrepancies among their results. Also, it is proposed that the nature of the task to be predicted is an important variable affecting judges' utilization of configural effects. This hypothesis is later examined.

HUMAN JUDGMENT AND MATHEMATICAL MODELS

Two Linear Models for Describing Human Judgment

Judgment, suggests Allen Newell (1968), cannot be explained through any kind of mathematical formula. It . . . "fills the gap in rational calculation. If the calculation could do it all, then no judgment is required" (p. 4). Such a statement is probably valid regarding any human activity. Since there is never perfect within-subject consistency, or perfect reliability among subjects, and no mathematical formula can account for human error, it probably is impossible to explain or predict human judgment with perfect accuracy through a mathematical formula. In fact, because of the vast individual differences among different judgment patterns in different people, it would seem impossible to find a general equation that will even approximate the judgmental processes used by "people in general" to judge or predict "things in general" or "events in general" or "other people's behaviors in general". A "somewhat less impossible" task would be to explain the judgment pattern of a single judge across all situations through the use of some kind of mathematical equation.

An excellent summary of some of the research on the use of multiple regression models in judgment research can be found in Slovic and Liechtenstein (1971).

i The Hammond Model

Researchers such as Hammond and Summers (1965), Hoffman (1968), and Goldberg (1968, 1969, 1971) have attempted to find some kind of mathematical model to explain the judgmental pattern of a single judge. They have concluded that a weighted linear equation offers a fair estimate of the judgment process of all judges for all events.

The foundation of these studies of human judgment is Brunswik's lens model, which is based on his theory of probabilistic functionalism and representative design. Although Brunswik primarily concerned himself with perception, his interest in cue utilization resulted in a model which is of great relevance to the study of several areas of human judgment. Brunswik (1955) states that all functional psychology is inherently probabilistic and demands a "representative" research design of its own. The organism must adjust himself to a semi-erratic environment, and there is never a perfect relationship between the actual environment (distal stimulus) and the cues. The ecological trait-cue relationships are always probabilistic. While it is often important to determine the nature of such relationships, Brunswik was primarily interested in identifying the probabilistic relationships existing between proximal stimuli (cues) and the observer's judgments. Such a relationship between stimulus cues and judgments is known as cue utilization (Brunswik, 1955, 1956).

Cue utilization can best be illustrated by Hammond's adaptation of Brunswik's lens model analogy, an outgrowth of Representative Design (Hammond, 1955). The Representative Design approach is summarized very concisely by Cohen (1973) as being based on the conviction that between proximal cues and judgments, and also proximal cues and distal attributes,

> ". . . relationships of interchangeable functional utility (vicarious functioning) may often exist: the same distal variable--say, 'being in love' may be expressed in differing ways (equifinality) and may, despite different proximal manifestations, for example, writing letters, holding hands, blushing, reducing ones contact with others, lead to the same judgment (equipotentiality) --being in love. This interchangeable functional utility is, according to Hammond (1955) one of the most significant reasons for the inability of diagnosticians to identify which cues or items of information are of greatest importance in leading them to their judgments" (p. 10).

The lens model proposes that "achievement" or "accuracy" depends on the ability of the organism to make its cognitive system become an adequate model of the task system so that this system produces the same output as the task system. The model is illustrated in Figure 1.

Studies using the Hammond lens model have been conducted by Hammond, Hursch, and Todd (1964), who developed the simple linear regression model which claims to approximate judgmental responses using a least squares best-fitting hyperplane. From this model comes the "C" component, which supposedly measures nonlinearity in cue utilization. However, Schaeffer, in a translator's footnote (Cohen, 1973) points out that it should actually be considered a "nonlinear or nonadditive" component, since nonlinearity often refers to a curvilinear relationship between two objects. As well, C combines additive and multiplicative (configural) nonlinearity. The higher the multiple correlation between all cues and the subject's judgment, the greater the likelihood that the judge has processed the cues in a linear, additive manner. However, as Hoffman <u>et al</u> (1968) have pointed out, while "C" may indicate valid, nonlinear cue utilization, it does not indicate the form of the nonlinearity; it confounds configural effects with various other

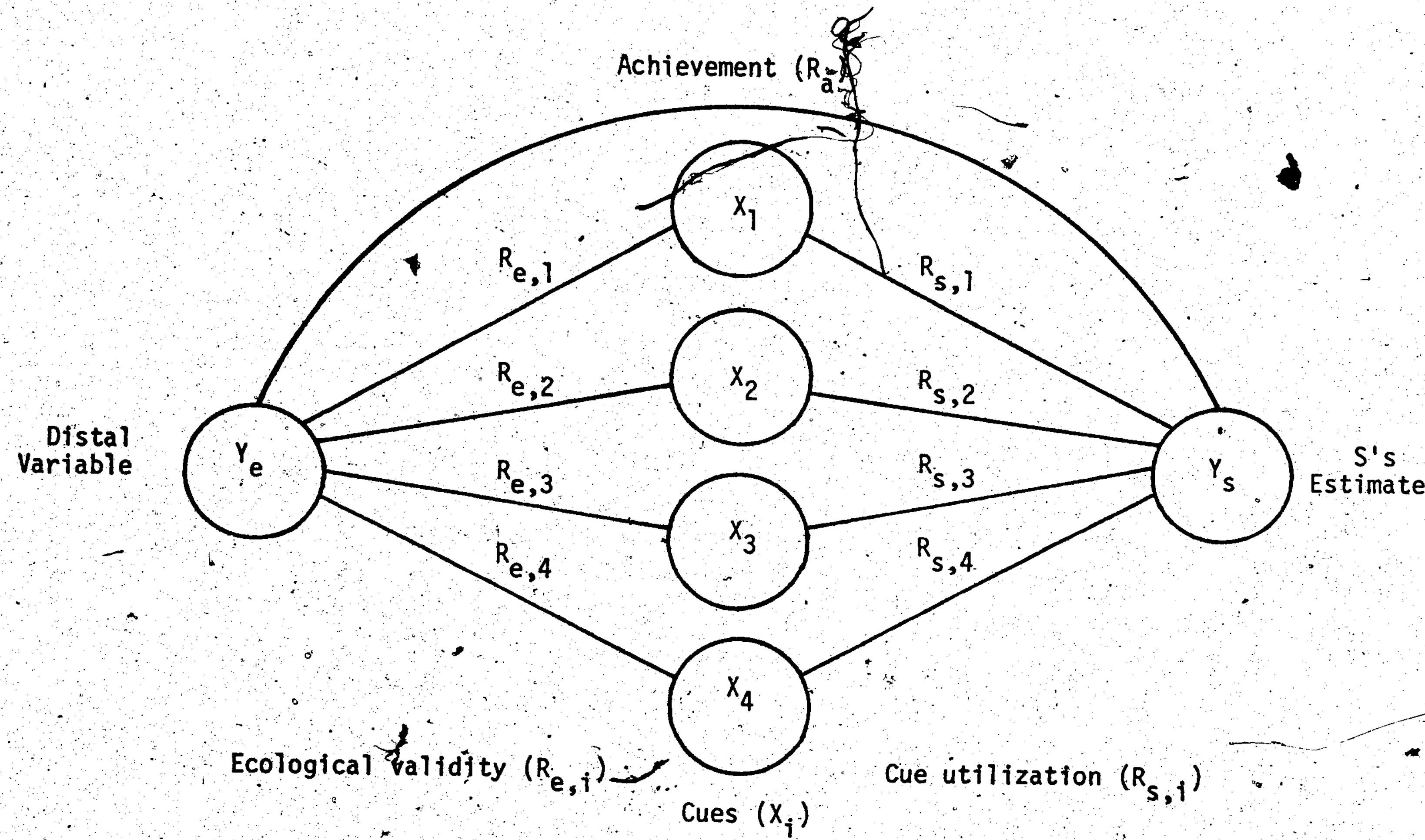

Figure 1. Brunswik's Lens Model

sources of nonlinearity. Information on the development and use of the linear regression model can be found in Hursch, Hammond, and Todd (1964) and Hammond, Hursch, and Todd (1964).

Some of the research using the multiple regression model has concerned itself with both sides of the lens model; i.e., the relationship between the subjects' utilization of cues ($R_{s,i}$) and the ecological validity ($R_{e,i}$). Since the interests of the present study are only in the judgment processes, with the criterion not being considered, only the right half of the lens model shown in Figure 1 is relevant.

A common criticism of the linear model concerns interpretations of high multiple correlation coefficients as indicating that the subject thinks in a linear manner. Many of the researchers who either use the regression model or the related Analysis of Variance approach for investigating judgment (e.g., Hoffman, 1960, 1968; Goldberg, 1968; Yntema and Torgerson, 1961; Green, 1968) emphasize the fact that the linear model only describes the data; it does not pretend to explain how the judge is thinking. Research, such as the current study, often seeks to find situations where the model can discover nonlinear components, while other research simply attempts to illustrate that the linear model may adequately describe the data. But the latter type does not claim to prove whether the subject thinks linearly or nonlinearly.

However, this is not to say that a linear model is not capable of at least partially describing the complexity of the individual judges' thought processes for various tasks; Hoffman (1968) has developed an ANOVA model, described in the next section, that, by examining main effects and interactions in an ANOVA design, may be capable of such a task. But first of all, some kind of evidence that a linear model is capable of distinguishing among different tasks must be shown (e.g., that some tasks and the judgments of some subjects or judges are better described or can be better predicted through the use of main effects in the ANOVA design than can others).

Dawes (1974) postulates that the formal models, such as the linear model, that have been developed to describe Ss behavior in processing information are primarily models of the task. He investigated performance on two very different tasks; one where Ss were to predict actual first-year graduate grade point averages utilizing ten cues, and another where they were to predict the order that three metal balls of different weights, placed at different distances from the center, will roll off a rotating disc as the speed of rotation is increased. In the first task, judges know that higher scores on certain cues (e.g., the GRE) are related to higher first-year grades in graduate school--and they use this information in the correct direction. In the second task, Ss must engage in "formal thought", which can be represented by a "truth table" constructed through using a combinatorial model, based on the relationship between distance from the center and the weight of the balls. But Dawes emphasizes that . . . "to propose that the model is one of the subjects would imply that the subjects engage in 'formal thought' consistently. And they don't". (p. 9).

Also supporting Dawes' hypothesis is a study by Dawes and Carrigan (1974) which demonstrated that the correlation between a subject's linear model and the criterion (in predicting grade-point averages) was no different from the correlation between an arbitrary weighting of the standardized variables (except for sign) and the criterion. Dawes (1974) claims that this data suggests that the validity of this model was simply a result of the fact that the model was linear, rather than a result of its relationship to the particular behavior of the subjects.

ii The Hoffman Model

Studies at Oregon Research Institute (ORI) have sought, through careful control of the given cues, what proportion of the variance of human judgments can be explained by a linear combination of the cues. The main instigator of this research is Paul Hoffman. A good summary of the research design and of some of the findings can be found in Hoffman (1968).

The general procedure of such studies involves quantified multivariate information which is presented to the judges, who are required to designate a category or value along a judgmental scale or dimension. Some of the early studies at ORI consisted of plotted values on nine predictor variables (highschool rating, status scale, percent self-support, English Effectiveness, responsibility, mother's education, study habits, emotional anxiety, and credit hours attempted) from which judges were asked to judge the intelligence of a person who had been rated at various positions on each of these cue dimensions. Or, they would be asked to judge sociability from values of eight scales of the Edwards Personal Preference Schedule (EPPS) (deference, exhibitionism, affiliation, succorance, dominance, abasement, change, and heterosexuality). A regression equation was then computed for each subject or judge.

According to Hoffman (1968), this regression equation describes the judgment process, approximating the judge's weighting of each of the predictors (cues).

> "The square of the multiple correlation (R^2) is a measure of the precision by which a linear combination of the variables weighted by parameters that have been iniquely estimated for each judge, can account for the variance of his own judgments" (Hoffman, 1968, p. 55).

Such a model is useful for characterizing the relative importance placed upon objective data in judgmental studies.

Hoffman (1968) criticizes Hammond's linear model on the grounds that the Rsi term is not a good index of cue utilization because the cues are not usually independent. Hence, Hoffman (1960, 1968) and Hoffman, Slovic, and Rorer (1968) have developed a judgment task which insures the independence of the cues; namely, the Analysis of Variance (ANOVA) technique, in which (a) the cues are categorical treatment variables, and (b) a completely crossed design is used, hence deviating a great deal from the representative design concept. Such a model has the potential for describing both the linear and nonlinear aspects of the judgment process, and identifies particular sources of nonlinearity through utilizing configural terms.

This model, while still emphasizing linear relationships, does contain interactive terms. Rather than using beta weights, as does the Hammond model, Hoffman converts them into relative weights (see Hoffman, 1968). His most commonly used statistic is the ω^2 statistic, as described by Hays (1963). ω^2

> ". . . provides an estimate of the proportion of the total variation in a person's judgments that can be predicted

> from a knowledge of the particular level of a given cue or pattern of cues" (Hoffman, 1968, p. 68);

it measures the relative importance of individual or patterned use of cues relative to the importance of other cues. Through this technique, it can be determined not only which cues are most relevant to the judge, but whether or not he is using these cues in some kind of configural relationship, and to what degree he is using any cue either configurally or linearly. A main effect for, say, cue X_1 would imply that if the other cues were held constant, the judges response would vary systematically with X_1. A significant interaction between X_1 and X_2 would imply that the effects of variation of cue X_1 is partly dependent on cue X_2. Thus, "the inclusion of interaction terms in a model takes account of the possibility that for a particular judge, the interpretation of one item of information may be contingent upon a second" (Hoffman, 1960, p. 122).

Using this technique, there is no need for the cues to be correlated in nature, as long as all possible combinations of cues are given to the judges, constituting a completely crossed design with all cues orthogonal. However, as Hoffman <u>et al</u> (1968) point out, it should still be explained to the subject that he has been presented with a <u>selected</u> array of cues, and therefore to expect a high proportion of unusual cases. Otherwise, the task may strike some judges as not being particularly believable.

Anderson (1969) has mentioned a few considerations that should be brought to the attention of anyone using an ANOVA model in the study of judgmental processes. When significant interactions occur in an

ANOVA, it may sometimes be the result of judges discounting information in order to resolve inconsistencies in their judgments, rather than reinterpreting the information, suggesting a more complex process. Anderson and Jacobson (1965) claim results from personality impression tasks have supported this contention. Such an interpretation might constitute evidence against using configural or interaction components as evidence for greater complexity of judgments in such a case. Other difficulties involved interactions being a result of floor and ceiling effects, response preferences, anchor effects, and related factors. However, studies done using monotonic transformations, claims Anderson, support the conclusion that response scales are generally valid, providing moderate experimental care is used, when ordinary rating methods are employed.

Thus, interactions discovered through the ANOVA model are generally valid indicators that some kind of configural process is being used by the judges, even though they are generally conservative estimates of the actual strength of the interaction. However, in order to determine whether or not a configural effect is meaningful, the nature of the specific interactions should be explored. An interaction would have a strong likelihood of being meaningful if the components involved in the interaction also produce strong main effects.

One contentious issue regarding the interpretation of configural effects in the ANOVA has been expounded by Slovic (1969). Slovic contends that since configural effects are qualitative rather than quantitative, one cannot talk about "degrees" of configurality; cues are either combined configurally or additively. If this were the case, the

Hoffman (1968) technique of comparing average ω^2 for interactions would be meaningless. Slovic also contends that the presence of interactions obscures the meaning of the main effects as well.

Despite these criticisms, it seems intuitively logical that an ω^2 of .15 would suggest a stronger concentration on the configural components than would an ω^2 of .05. Hence, the present study employs the Hoffman model, but also looks at the number of significant interactions (disregarding magnitude) and the number of judges utilizing cues in a way that produces at least one significant interaction.

Evidence for Linearity in Judgment

Although Hoffman (1968) refuses to totally accept the linearity principle (i.e., that judges can generally combine data only in a linear manner), he claims that most of the evidence favors it. One of the problems in analyzing the literature in this area is that it is not always entirely clear whether the linearity principle refers to the fact that judges are most accurate when utilizing linear rather than nonlinear relations, or whether it refers to the fact that judges either do or don't use nonlinear data, disregarding the matter of accuracy. Another question, of course, is whether or not judges can be taught to utilize nonlinear relationships. Since the present study deals with cue utilization, and is primarily interested in the process by which judges combine cues, the matter of accuracy will be treated as only a subsidiary problem. Because of the unreliability of most of the criterion measures (e.g., the criterion of judgments of MMPI protocols is most frequently the diagnosis made by psychiatrists), many studies concluding that configural or nonlinear usage of cues does not improve accuracy must be taken with at least one grain of salt.

If the linear model were accepted, it would . . .

> "imply that individuals do not alter their mode of 'weighting' the dimensions of information regardless of the pattern or configuration of values inherent in the object being judged" (Hoffman, 1968, p. 59).

Some of the evidence Hoffman (1968) uses to support the principle of linearity comes from his studies at ORI, where the multiple correlations for judges judging sociability and intelligence from different profiles of predictor variable ranged from .80 to .90, as did the test-retest reliabilities of the same judges (i.e., 64% to 81% of the judgmental variance could be attributed to linear combinations of the profiles, while 19% to 36% could be attributed to unreliability). "These studies have consistently shown that virtually no improvement in predictive accuracy can be expected beyond that achieved by a linear model" (p. 60).

The classic study supporting the linearity theory of judgment is reported by Hoffman et al (1968). Gastroenterologists were asked to judge hypothetical stomach ulcer cases as being benign or malignant, from seven cues that would normally be available on a stomach X-ray. It was found that:

> "roughly 90% of the judge's reliable variance of response could be predicted by a simple formula combining only individual symptoms in an additive fashion and completely ignoring interactions" (pp. 343-344).

The largest of 57 possible interactions, for the most configural judge, accounted for only 3% of the variance of his responses. Also, there were very low inter-judge intercorrelations (median correlation was

.38). Test-retest correlations ranged from .60 to .92, suggesting a fair amount of intra-judge reliability. Before the experiment, several judges had given evidence suggesting that such a task was generally done using a configural approach.

Part of the discussion section of Hoffman's study, however, stressed that one could not completely discount the importance of interaction effects, even when the contributions seem so small. In some cases such interactions could enhance diagnostic accuracy.

Goldberg (1968) cites the above study plus two others as support for the ability of the linear model to describe most clinical judgment tasks. Summarizing the literature, he concludes that even for judgment tasks which have been specifically selected to show the clinician at his best, the clinician's validity never goes beyond the level of validity of judgments which use a simple actuarial formula of the form: $x=b_1x_1 + b_2x_2 + ...+ b_kx_k$, where $\underline{z}$= the vector of judgmental responses, $x_k...x_k$= the values of the matrix of $\underline{k}$ cues by $\underline{N}$ targets presented to the judge, and b_1 and b_2 are constants representing the weight of each cue in the judgmental model.

The judge produces the $\underline{Z}$ values from knowledge of the $\underline{x}$ values for each of the $\underline{N}$ targets. A linear regression equation can determine the $\underline{b}$ values or regression weights, and the accuracy of the model can be determined by finding the correlation coefficient (R_a) between the calculated regression weights ($\underline{b}$) and each judgmental response (Goldberg, 1968, p. 486). Goldberg also used the ω^2 statistic in order to ascertain the importance of individual and configural usage of cues relative to the importance of other cues.

Goldberg (1968) cites three studies to support the ability of the linear model to explain human judgment. The first is the Hoffman, Slovic, and Rorer study of gastroenterologists reported earlier. The second study involved psychiatrists making a decision as to whether or not to grant temporary leave to a psychiatric patient (Rorer, Hoffman, Dickman, and Slovic, 1967) using six supposedly relevant variables (e.g., does the patient have a drinking problem?--yes or no). On the average, less than 2% of the variance of these judgments was associated with the largest interaction term. Again, there was a lack of inter-judge agreement.

The third study reported was Wiggins and Hoffman's (1968b) analysis of Paul Meehl's (1959) data comparing clinicians' judgments with five statistical methods of identifying psychotic MMPI profiles--a task which Meehl had claimed should be highly configural. The results indicated that for most of the judges, the linear model represented most of the variance in clinicians' judgments. However, for some judges, one of the nonlinear models provided a slightly better representation of their judgments than did the linear model.

Goldberg (1968) also cites two studies by Slovic (1966, 1968) which support configural cue utilization. Thus, Goldberg concludes . . .

> "that judges can process information in a configural fashion, but that the general linear model is powerful enough to reproduce most of these judgments with very small error" (p. 491).

Hammond, Hursch, and Todd (1964) are some of the most outspoken proponents of the linear model. They examined Grebstein's (1963) data concerning the judgment of IQ from Rorshach data, using the correlations between residual variances ('C') for clinicians of varying experience. They asked the question as to how much room for improvement in judging IQ's from Rorshachs actually exists through utilizing configural methods. Their conclusion was "not very much". While the correlations between Rorshach psychograms was not particularly high (.50 for the naive clinicians; .68 for the sophisticated clinicians), they found that the particular task did not provide the clinician with opportunities to use any special properties he might have (i.e., ability to use configural judgments) that might distinguish him from the multiple regression equation. For example, one particular sophisticated clinician, who had an achievement of .68, could have increased to only .84 if the correlation between the residual of his response system and the residual aspects of the Rorshach IQ system had been perfect (+1). On the other hand, if the correlation of residuals ('C' component) was -1, his achievement would have been reduced only to .37. A zero correlation of residual would have resulted in an achievement of .61. Thus, according to those particular researchers, even if the judge did develop a substantial value of 'C', his achievement was very close to the maximum he could have possibly received. Hence, they conclude that there would be little value in teaching clinicians to utilize configural relations, since the particular clinician mentioned could only have increased his performance from .68 to .84. Such a statement is made despite the fact that using no configural judgments, the clinician's judgments account for 37% of the variance (R_a=.61), while perfect utilization of configural components plus the same usage of linear components would account for over 70% of the variance of his judgments; a difference of 33%.

Similar findings were reported by Peterson, Hammond, and Summers (1965). The subjects' response system was compared to the optimal response system as defined by a linear multiple-regression equation. The correlation between response and criterion was .73, while the optimal value was .83.

Another study supporting the linear model, reported by Hammond et al (1964) was conducted by Newton (1965). Ninety-nine college sophomores estimated freshman grade averages from four cues: IQ, high school rank, a college board score, and a personality rating by their high school principal. The judgments turned out to be highly linear, with negligible values of 'C'.

Naylor and Wherry (1965) rated fictitious soldiers on 23 behavioral scales, and 50 airforce officers were required to judge them on suitability for the airforce. Correlations between cues and subjects ranged from .75 to .99, with the great majority above .90.

Rudolf Cohen (1973), in a very meticulous study involving the judgment of personality on the basis of photographs and handwriting samples, found that the greatest proportion of reliable variance of the judgments could be explained by a linear, additive model of the different graphological and physiognomic characteristics.

Support for the linear model also comes from some (but not all, as will be mentioned later in the description of Einhorn's research) of the studies which try out alternative models to the linear model. Re-analyzing Meehl's (1959) data, where 28 clinicians diagnosed psychotic and neurotic patients from the MMPI, Goldberg (1969) states that . . .

> "neither clinical experts, moderated regression analyses, profile typologies, the perception algorithm, density estimation procedures, Bayesian techniques, nor sequential analysis--when cross-validated--have been able to improve on a simple, linear, function" (p. 523).

Once again, Goldberg (1969) does not use this as evidence that clinicians in general think linearly; he concedes that such a task may not be the right task for testing the clinician's ability to utilize complex configural relationships.

Berndt Brehmer, at the University of Umea in Sweden, has recently become known for his research in this area. While many of his studies have been more supportive of the existence of configural or nonlinear relationships in the judgment process (e.g., 1969, 1971a, 1971c), others (1971b, 1973) tended to support the linear model. In a study (1971b) in which changes in policy (usage of information determining the judge's opinion) were investigated as a result of interpersonal interaction, the main finding was that the subjects changed their policies to adapt to the task and to make them more similar to the policy of the other person in the pair. A sub-finding, important for the present discussion, was that when one subject in each pair was trained to rely on the nonlinear cue and the other on the linear cue, the performance was higher in the linear condition, primarily because of the fact that subjects trained to make linear judgments were more consistent. Another study by Brehmer (1973) found that in a task where judges were to make decisions from bar graphs, and nonlinear cues were used in conjunction with linear ones, the linear cues were used most effectively. Summers and Hammond (1966) had previously shown, however, that judges can learn to make inferences from nonlinear as well as linear task relations, but that the nonlinear deductions are more difficult to learn.

If what Goldberg (1968) says is true, that is, that "the accuracy of a linear model was almost always at approximately the same level as the reliability of the judgments themselves" (p. 488), then why don't clinicians simply admit that they can't judge configurally, and use a linear regression model for making all judgments? Goldberg (1968) postulates three possible reasons for the discrepancy between what clinicians say they do and what the linear model says they do.

(1) Humans behave like data processors but believe they are more complex.

(2) Human judges do behave configurally, but the power of the linear model is so great that it obscures the configural processes.

(3) Human judges behave linearly in most judgmental tasks, but for some kind of tasks they use more complex processes.

Since there have been various studies, many of which will be described in the next section, that have shown some variance that can be explained as configural and/or nonlinear effects, there would seem to be evidence against Goldberg's first hypothesis. Hypothesis 2 has been supported by Yntema and Torgerson (1961), Rorer (1967), Bert Green (1968), and Norman Anderson (1962); they have shown that the linear multiple-regression model and the related analysis of variance model are extraordinarily conservative tests for picking out more complex interactions.

Hypothesis three is the main one that will be investigated in this thesis. It seems likely that there are some tasks that are so obviously "configural" that even the conservative ANOVA design should be able to pull them out. Before such a search for configural tasks can be made, it is important to survey some of the literature in order to determine first of all, whether there are any tasks that are judged configurally by most judges, and if so, if there is any difference between tasks which result in the utilization of a high degree of configurality and those tasks which do not.

Studies Showing Nonlinearity in Judgments

While this paper's main concentration is on configural effects, this section will deal with many kinds of data that suggest nonlinear judgment processes.

Regarding utilization of nonlinear relationships to a criterion, Summers and Hammond (1966) discovered that while linear relationships were easier to handle than nonlinear ones (in this case they were sine curve relationships), judges could be trained to utilize both types when they were instructed that both types were necessary for perfect accuracy, were given illustrations and were told which cue was linear and which one was nonlinear.

An earlier study by Hammond and Summers (1965) produced similar findings. This paper also presented statistical data suggesting the classic study by Grebstein (1963)--where clinicians were asked to predict IQ from Rorshach protocols--did not permit the use of nonlinear relations. The task itself was so linear (i.e., a linear combination of the cues were all that was needed to make valid predictions), that clinicians could hardly have improved their performance by attempting to discover and utilize nonlinear relationships in the task.

Cue consistency is another aspect of the task which seems to have some effect on the use of nonlinear relationships. It seems intuitively logical that a judge would more likely use a configural approach to solving a problem if there were some kind of cognitive inconsistency suggested by the cues. For example, if a judge knew that a person was intelligent, sociable, and well-adjusted, he would gain a favorable impression of him with little need to process the cues in anything but a simple, additive, manner. But if he later found out that the person had been convicted of a serious crime, he might need to reprocess some of these cues, deviating from the model of "intelligent + sociable + well-adjusted - criminal". It is more likely that he will re-define at least one of the characteristics and come up with a new interpretation (assuming he is not a simple-minded judge). Several studies on cue consistency have been done. Slovic (1966) had judges make judgments of intelligence on the basis of English Effectiveness (EE) scores, and High School Ratings (HSR). When the profiles were consistent (i.e., differences between EE and HSR were small), the judgments were dependent on both cues. When the profiles were inconsistent, the judges seemed to rely on only one of the cues, suggesting a primitive, but configural utilization of cues.

Brehmer (1971c) did not find this tendency for the individual to give up the use of one of the cues for inconsistent cue configurations, but did find lower multiple correlations for cues and responses for inconsistent combinations than for consistent ones, suggesting that the linear model describes the judges' utilization of cues less well for the inconsistent combinations than for the consistent ones.

Other research suggests great individual differences in the utilization of inconsistent cues. The results attained by Wiggins and Hoffman (1968a) resulted in the suggestion that judges sometimes resolve inconsistent cues by weighting one or the other cue, and sometimes by utilizing entirely new cues for judgment--likely a result of a configural combination of the given cues. This hypothesis goes along with the frequently reported study by Gollin (1954). In this experiment, motion pictures were shown with a woman behaving in ways suggesting certain diverse and discrete character qualities. Two behavioral themes were used: promiscuity and kindness. Gollin found that subjects dealt with them in one of three different ways:

1. Some dealt with both behavioral themes and attempted to relate the presence of these diverse behaviors in one person (related impressions).

2. Both behavioral terms were used, but no attempt was made to relate them (aggregated impressions).

3. Persons were characterized in terms of one behavioral theme--immoral or nice (simplified impressions).

In Gollin's study, 18 judges used related impressions, 23 used aggregated impressions, and 38 used simplified impressions. In a multiple-cue probability learning study, the related impressions method would be expressed as a configural utilization of cues, the aggregated impressions would constitute linear judgments, while the simplified impressions method would be illustrated by the usage of only one cue, and would preclude the possibility of the ANOVA model picking out any interaction effects. However, such a method does possess a certain simplistic type of configurality, since the use of the one cue results from the judge seeing the inconsistent configuration. This type of configurality is not a part of complex judgments, however, and is not the type of configurality that the present study is interested in. However, if in an orthogonal, completely crossed ANOVA design, a judge used two cues when the cues were consistent but ignored one of them when they were inconsistent, such a strategy could likely result in a larger configural component. In such a case, it could be debated as to whether the configuration truly indicates a more complex thinking process than a linear strategy which utilizes all the cues.

So far it has been shown that cue utilization may be dependent on individual differences, instructional set, and cue consistency. The two Hammond and Summers studies cited earlier (1965, 1966) and studies by Hursch, Hammond, and Hursch (1964) have suggested that the relationship between the cues and the criterion are also important. However, Brehmer (1971d) in a study where cue-criterion relationships were varied, obtained results that were not consistent with the hypothesis that judges match their utilization of the cues in a multiple-cue probability learning task to the correlations between the cues and the criterion, suggesting that the relationship between the cues and the criterion is only one variable influencing the method in which judges utilize cues.

Another finding of the Brehmer (1971d) study was that judges' utilization of cues did not match the beta weights of the task, as determined by the multiple regression equation, suggesting that they could not be explained adequately by a linear model. Also, the judges were less consistent when the cues were inconsistent than when the cues were consistent, and they were also quite sensitive to changes in the task, suggesting they were processing the cues with some care. Another study by Brehmer (1971e) showed configural relationships between two tasks.

Results such as these have contributed to Brehmer's (1971c) usage o[illegible] thinking as a method of defining cognitive complexity. J[illegible] have to be relatively cognitively complex to process incon[illegible]ues in a manner that would utilize nonlinear and configural rela[illegible]ps. Also suggested is that certain tasks invoke a more comple[illegible]dgment or cue processing strategy than do others.

A[illegible]er by Schumer, Cohen, and Schwoon (1968) plus various papers by And[illegible]on (1962, 1965, 1968, 1972) demonstrate how linear models can be best used to explain configural data. Schumer et al (1968) compared three linear models: Anderson's averaging model (later revised to become a weighted averaging model), the congruity model of Osgood and Tannenbaum (similar to the weighted averaging model, only with the weighting specified), and a simple additive or summation model. The first two models predict that the judgment of combinations of cues will be so[illegible]here between the judgments of each individual item, while the summat[illegible]model implies that the judgment of two items will be equal to the sum of each item judged separately. Most of the tests failed to support the summation model. It was found, also, that a few subjects judged the combination of two cues in the opposite direction from their judgment of each of the two cues separately. None of the three linear models could explain such data. Most of the judgments revealed interactions which are incompatible with any linear model which postulates weighting of the individual items in a manner that is independent of the given combinations.

Anderson (1962, 1965, 1968, 1972) has done a great deal of research emphasizing the importance of configural effects in judgment processes; and he has investigated them through a linear model. The averaging rule of stimulus combination states that adding a moderate stimulus to an extreme stimulus decreases the polarity of the response. A straight additive model would predict the response to be polarized. Anderson (1965) illustrates the principle with this example:

> "Suppose you consider 'painstaking' to be a moderately desirable trait, and well-spoken to be highly desirable. Would you then like a 'painstaking, well-spoken' person more than a 'well-spoken' person?" (p. 394).

According to any kind of simple additive model, you would. However, Anderson's averaging model postulates you would not. Though the model used to investigate it is linear, the averaging process itself, claims Anderson (1972), is configural, . . . "because the effective weight of each stimulus depends on the whole set" (p. 93). Evidence for the averaging process has been frequently reported in the psychological literature. When Anderson (1968) followed a set of three high-value adjectives with three mildly favorable ones, the response was significantly lower than when the three high-value adjectives were presented alone. Lampel and Anderson (1968) found interactions between the judgments of photographs and adjectives by forty female subjects judging males on the criterion of desirability. The nature of the interactions was a changing of weights; for unattractive males, the adjectives made little difference, while for attractive males, the adjectives became more important. A study by Shanteau and Anderson (1969), where combinations of drinks and sandwiches were judged, a linear subtractive model was found to explain the data well. However, one quarter of the cases showed interactions. Sidowski and Anderson (1967) investigated the attractiveness of four occupations (doctor, lawyer, accountant, and teacher) for four cities in the Unites States. A significant teacher-city interaction, concentrated at the most unfavorable city, was found. In a study investigating the behavior of diagnosticians judging "disturbed behavior", Anderson (1972) found that stimulus items that had the most extreme values were weighted much heavier than were the ones with more moderate values. Once again, this supports a type of configurality in which the cues are weighted differently depending on their given values. The multiple-regression equation would be unable to account for such a dependency between cue value and the subjects weighting of the cues; i.e., it does not account for the subject changing the beta weightings for each stimulus array.

Few successful attempts have been made to explain nonlinear effects through models other than the linear model. Einhorn (1970, 1971), however, was able to find a task in which a nonlinear model was more successful at predicting the judges' responses than was a linear model. The task involved each judge choosing a job within their own occupations, with information given about each job. Each variable was presented in the form of a 7-point scale, and judges were asked to rank-order the 15 jobs given, then to rate the importance of each of the attributes in making their decision. Einhorn (1971) found that a conjunctive model (a model which assumes that a person must have a certain minimum ability on all attributes) yielded higher correlations with the subjects responses than did the linear model or the disjunctive model (where a person is judged on his best ability, regardless of his other attributes). A sign test showed that the conjunctive model had higher correlations than the linear model for 32 out of the 37 cases. Einhorn admits that such a task was conducive to a conjunctive solution. He states:

> "Choosing a job may have certain minimum standards on the number of attributes, less than which one is unwilling to accept. Furthermore, the cost of a false positive in this solution would be expected to be high; i.e., choosing a job that turns out poorly would involve a high cost to the decision maker" (p. 14).

Einhorn (1971) also had judges (faculty members or last year graduate students in psychology) choose among applicants to graduate school in psychology. Here, he found that there was no systematic use of either the linear or the nonlinear models in this task, supporting the view that the model best explaining a judge's judgment strategy is <u>partially dependent on the nature of the task</u>. Further support is given by Goldberg (1971), who compared Einhorn's conjunctive and disjunctive models with linear, logarithmic, and exponential models, using, as the judgmental task, the differentiating of neurotic from psychotic patients on the MMPI (Meehl's (1959) data). He found that the linear model provided the best representation of the judgments made by all clinicians; only the logarithmic model provided the linear model with any real competition. An earlier study by Meehl (1960), however, found the linear model to be a poorer predictor of psychotic tendency from MMPI scores, than was the pooled judgment of 21 clinicians. However, Meehl admitted that the amount of deviation from the linear model in these pooled judgments was not very great. Goldberg (1970) himself found that using squared regression weights of the judges on judgmental responses on 11 MMPI scores showed a fair amount of nonlinear contributions to the accuracy of the judgments. However, only in one case (out of 29 judges) was there any sizeable discrepancy function favoring the judge over his model.

Goldberg (1971) stresses that the ability of the linear model to describe judgments does not imply that the judges actually processed the cues in a linear fashion. Much of the variance was clearly nonlinear, as was shown by the residuals after the linearly predictable variance from the 11 MMPI scores had been partialled out. The power of the linear model would not negate the possibility that many of the cues could have been processed configurally. Hoffman (1968) points out that models can only describe and not explain the data. The "paramorphic" problem is one of the reasons that Hoffman refuses to accept the view that most people are unable to progress beyond linear utilization of cues in making judgments. The paramorphic problem involves the fact that:

"a Two or more models of judgment may be algebraic equivalents of one another, yet suggest radically different underlying processes.

b .Two or more models of judgment may be algebraically different, yet be equally predictive given fallible data. Thus, there are sets of data for which the function $Z' = A_1X + A_2Y$ and the function $Z'' = B_1X + B_2Y$ lead to exactly the same residual variance . . .

Hence, the achievement of a high level of predictive accuracy for a linear model does not negate the possibility of configural relationships; but it does place the additional burden on the experimenter to find a more adequate test, a different experiment, or a special type of data structure that would be more likely to reveal a degree of superiority for his hypothesized configurations" (pp. 62-63).

The most important reason for not accepting the linearity principle is simply because most skilled clinicians reject it. They report that their interpretation of a given dimension of a patient's behavior (in a test, interview, or social situation) is conditional upon the values of other dimensions. Hence, they claim a configural strategy. It is possible that a task that cannot be explained through a linear model may be one that employs a configural strategy; some studies that have revealed configural strategies are discussed below.

Kleinmüntz (1968) found a relationship in the judgments by clinical psychologists and neurologists of MMPI profiles that could best be called configural. He had them think aloud into a tape recorder, and from these reports, he constructed a computer program. The result was a complex sequential representation of verbal reports that took in the whole configuration.

An unpublished study by Martin Jr. in 1957 (Hoffman, 1960) reported interactions and nonlinearity (e.g., variables are most important when their values are high) to have been demonstrated in descriptions by clinicians as to how they made assessments from eight Edwards Personal Preference Schedule (EPPS) variables. Wiggins and Hoffman (1968b) also found evidence for configural judgments in clinicians judging neurotic or psychotic profiles from MMPI's. They compared three models: a linear model, a quadratic model, and a sign model (a linear combination of 70 clinical signs as described by Goldberg [1965]). The sixteen (out of 29) judges who predicted best from the quadratic or sign models had a significant, relatively large, nonzero correlation for at least one of the configural terms.

Looking at the ANOVA studies, Rorer and Slovic (1966) found evidence for configural cue utilization, but claimed that the configural component was not helpful in improving predictive accuracy. One of the most frequently cited studies is Slovic's (1969) analysis of stockbroker's decision processes. Two young brokers were used as subjects, with one of the brokers (broker A) selecting 11 variables to be used. The agreement between the two brokers was quite poor ($r=.32$). A separate analysis of variance was performed on each broker's response. Broker A revealed significant interactions (one of the interactions being the fourth strongest effect). Broker B showed seven significant main effects and five two-way interactions. Using the ω^2 technique described earlier, broker A was found to have 72% of his variance predictable from five main effects, and another 7% of the variance predictable from six significant two-way interactions. Broker B showed 80% of his responses could be predicted from seven significant main effects, and 5% from five two-way interactions.

Slovic re-analyzed the data using the magnitude of effect index, based upon the influence of a factor upon the mean judgments; a gauge that Slovic felt was more meaningful for assessing relative importance of configural effects. He found that configurality accounted for 27% of the total effects on broker A, and 19% on broker B. But even this index, claims Slovic, is a conservative estimate of configurality. It can be argued that:

> "Whenever the interaction between two factors is significant, these factors were being used configurally, and the variance accounted for by both their main effects and their interactions should be counted as configural variance" (Slovic, 1969, p. 269).

This method would boost the configural variance for broker A to 36%, and for broker B to 85%. Further evidence for configural effects was that two of the interactions were common to both brokers, and were also of the same form for both brokers.

Two unpublished papers from the University of Alberta have also shown configural relationships. The first one was on the topic of creativity (Schaeffer and Jackson, 1970), and involved eleven judges rating the creativity of 128 hypothetical individuals, using seven attributes arranged in 128 profiles representing all possible combinations of two levels (high and low) of each trait. The percentage of the variance accounted for by interactions ranged from 1% to 43% with the median being 7.5%. The number of interactions ranged from 1 to 63 (median=12.5) out of a possible 113. The authors claim that this indicates that judgments of creativity are fairly complex (i.e., number of interactions is used as an index of cognitive complexity).

The other paper is by Schaeffer and Saidman (1971), where significant interactions were found when subjects were asked to state their preference for various short musical compositions, with different characteristics of the composition varied (e.g., melody was lyrical or nonlyrical, harmony was simple or complex). The magnitude of the interactions was similar to that of the Schaeffer and Jackson study. A hypothesized melody and harmony interaction turned out to be the third most important effect.

Both the tasks in these two studies seemed to have been of a nature such that complex, configural judgments were expected. The next section will examine two different kinds of judgmental tasks; analytic tasks, which are hypothesized to be judged in a linear manner, and intuitive tasks in which more complex interactions are expected.

ANALYTIC AND INTUITIVE JUDGMENTS

For purposes of the present study, the primary distinction between types of judgments will be between scientific analysis and what is called clinical intuition. It is hypothesized that both these types of judgment can be investigated scientifically.

For a clinician to make an intuitive judgment, he needs cues. And these cues must be processed in some way. However, while the analytic judge first finds out what the cues are, and after a certain amount of experimentation, determines a system by which he will process these cues, the intuitive judge is frequently unaware of the cues he is using. Johnson (1955) has made the claim that all judgments of complex stimuli may be determined by stimulus properties which the judges are unable to identify. One frequently cited study that supports this viewpoint was conducted by McKeachie (1952). Six girls were interviewed six times by males, three times with and three times without lipstick. Without lipstick they were rated as more conscientious and less interested in boys than with lipstick. The important finding, however, was that none of the judges were aware that the presence of lipstick affected their judgments.

In some cases it may be not only that intuitive judges are unaware of the cues, but it is this unawareness that enables them to be such competent intuitive thinkers. A book by Hanbury Hankin (1928) provides evidence to support this seemingly far-fetched theory. Most of the evidence is anecdotal rather than scientific; he refers to stories of doctors who are known for the accuracy of their diagnosis but are unable to explain how they made them, people with abnormal calculating power, the ability of the motorist to subconsciously estimate the speeds of other cars before proceeding across an intersection, and of the intuitive understanding people have of music. Hankin recites the story of Zerah Colburn, a nine year old boy who could do rapid calculation and had a remarkable power of "factorizing" (e.g., if given the number 171,395, he would say it was equal to 5x34279, 7x24415, 59x2905, and 413x415). At first he had no idea how he was able to do this, but four years after the power first appeared, Zerah Colburn was able to acquire a partial knowledge of how he performed these calculations. However,

> ". . . as soon as he had acquired his general view of the subject, his power of rapid calculation left him finally and completely. It was only so long as his data were not known, or not clearly known, to his consciousness, that they were fully available for the use of his subconscious mind" (p. 76).

Such a phenomenon of losing an intuitive or instinctual ability when one starts to analyze it is probably a fairly common conception, as illustrated by the popular fable of the centipede who had no trouble walking until he started concerning himself with what each of his 100 legs was doing. This whole issue regarding the effects of identifying the cues in intuitive thinking requires empirical investigation. But first it is necessary to get some idea of what intuitive thinking is.

What may be operating in intuition is a process that is primarily synthetic rather than analytic. Instead of needing to break down the parts and then put them back together, a person with a great intuitive ability may have the ability to immediately see the whole configuration. For instance, one faculty necessary for great calculating ability, according to Hankin (1928) is the ability to rapidly recognize numbers; e.g., to know at a glance the number of a flock of sheep or a handful of peas thrown on the ground. Hankin uses the example of the abnormal calculating ability of G. P. Bidder to illustrate this. Bidder suggested that his ability in calculating may have been partly due to his becoming familiar with numbers by playing with pebbles or peas before he knew the meaning of symbols. "If he heard the number 64, he did not at once think of the symbols six and four. He thought of eight rows of eight pebbles each" (p. 67). His power was a natural instinct to him; he was unable to explain it.

The equating of intuition with synthetic judgment seems to have received some support from various philosophers. Ewing (1941), in an annual philosophical lecture published by the British Academy, suggests that intuition involves the finding of a connection between two ideas--

a connection that cannot be proved, but is either seen or not seen. He states:

> "We could not start at all in any reasoning without assuming that we immediately perceive a connexion between certain premises and their conclusions. To argue at all, we must see the connexion between the propositions which constitute the different stages of argument not by mediate reasoning, but intuitively . . . . Such immediate reasoning would only be another name for what is commonly called 'intuition'" (p. 9).

Bunge (1962), in a book which strongly opposes the use of so-called intuitive thinking as a replacement for scientific thinking, talks about the power of synthesis of global vision--" . . . the ability to synthesize heterogeneous elements to combine formerly scattered items into a unified or 'harmonious' whole; i.e., a conceptual system" (p. 86). Of course, such a conceptual system can be analyzed scientifically. However, it is only when the nature of the system is not known that we refer to such judgments as intuitive. Hence, it would seem unlikely that a logical deductive process or even a consciously inductive process would be used in making intuitive judgments. In any deductive process, the initial premise is the only aspect that might employ intuition. The method of getting to this intuitively formed initial premise would most likely be through an unconscious inductive process.

A very important difference between intuitive and analytic reasoning is that the criteria are often different. For example, in making a narrow prediction about an aspect of a person's future behavior, it is likely that an analytic process would most often be used. In wanting to know how likely a person was to succeed in some occupation, one would want to know his success in other occupations, how motivated he was, and other related information in order to make a judgment through induction. Also, one might want to know something about his attitudes in order to make various logical deductive inferences (e.g., this person is aggressive, and aggression is what is needed to succeed in life; or, the person is aggressive, and aggression will make it hard for him to get along with his fellow workers, thus hindering his chances of success). Intuitive reasoning might be used, however, in judging a whole person; i.e., in asking a question such as "how much do I like this person?" In terms of making predictions, it is felt that an intuitive prediction would involve a vaguer concept than would an analytic prediction.

In distinguishing between intuitive and analytic judgment tasks, the presentation of cues presents a problem. For making intuitive judgments, judges usually do not have to be presented with specific cues; if the "right" cues aren't there, no intuitive judgment will come about; if they are there, then an "instant judgment" may take place. Analytic judgments, on the other hand, require cues that are specifically relevant to the judgment to be made. If the cues aren't good ones, then the judgment is liable to be inaccurate. It is fairly easy to ask judges to make analytic judgments, given a defined set of cues. With intuitive judgments, however, the judges may be unable to make the judgment because the cues or the tasks don't seem right. In such a case, judges may revert to an analytical process. Hence, presenting a task that is considered to be intuitive does not necessarily mean that judges will respond to it intuitively.

Using some of the concepts mentioned above, definitions of intuitive judgments and analytic judgments have been formulated. These definitions have been inspired by Bunge (1962), although he would not necessarily agree with them.

Intuitive judgments are usually made quickly and require a minimal amount of detailed information. They may be based on metaphorical thinking and/or they may entail the ability to synthesize heterogeneous (and sometimes disparate) elements into a whole. They are generally considered 'common sense' judgments.

Analytic judgments are usually made after much deliberation, and utilize an analytical thinking process; i.e., break down the components of the decision and weigh the evidence or consequences of the decision (or possible consequences). They are logically deduced or consciously induced, and a fair amount of information is generally needed before an analytic judgment is made.

The above distinction between analytic and intuitive judgments implies three distinct (though likely related) concepts of intuition. Firstly, intuition is seen as a fast, careless judgment where very few cues are attended to. This aspect of intuition incorporates the cliche of the "layman's intuition". Secondly, an intuitive judgment is seen as a judgment where cues are processed unconsciously, but complexly. In such a case an experienced analytic judge who no longer needs to consciously process the cues can be considered to be making his judgments intuitively. The third concept of intuition treats it as a global judgment, as opposed to an inferential judgment. It incorporates the notion that the whole is greater than the sum of its parts.

All three distinctions, however, are consistent with the hypothesis th[illegible]ntu[illegible]gments employ configural strategies. Fast, careless ju[illegible]ts [illegible]sregard cues when they are inconsistent (possibly leading [illegible] mor[illegible]teractions revealed by the ANOVA), unconscious processing [illegible]ay not be conscious because of the complexity of the Js utiliz[illegible] of them, and the global or "synthetic" aspect of the definition [illegible]ously suggests a combination of cues that goes beyond a linear [illegible]ess.

Taf[illegible]955) in his paper on analytic versus nonanalytic judgments, used defi[illegible]ons similar to the above, stressing the inferential vs. global distinction. He writes:

> "I[illegible]alytic judgments, the judge (J) is required to concep-t[illegible]nd often to quantify specific characteristics of t[illegible]ject (S) in terms of a given frame of reference. T[illegible] mainly involves the process of inference, typical per-formances of J being rating traits, writing personality descriptions, and predicting the percentage of a group making a certain response. In nonanalytic judgments, J responds in a global fashion, as in matching the persons with personality descriptions and in making predictions of behavior. An empathic process is usually involved in non-analytic judgments" (p. 1).

Interesting to note is Taft's consideration of predicting behavior (of an individual) as a nonanalytical process, suggesting that most clinicians would make intuitive rather than analytic judgments. The present author considers the prediction of the more specific behavior to be more analytic than intuitive. However, the cues that would be used are also important. Nonanalytic judgments, according to Taft (1955) would include judging emotional expressions in photographs, drawings, and movies, personality matchings (J might be required to match some data concerning S with some other data concerning the same S), and prediction of behavior or life history data. Analytic judgments would include the rating and ranking of traits and personality descriptions, where J is provided with certain data about S.

Taft emphasizes the problem of vast individual differences among judges. He cites a study by F. H. Allport (1924) in a task involving the judging of emotional expression; it was found that some judges were superior at judging the intended emotion when using a naive type of intuitive method, while others were superior after they received training in the use of analytic methods of making judgments.

In reference to studies which have suggested that clinical psychologists are no better judges of people than are laymen, the literature reported by Taft (1955) suggests that most of the judgments were analytic tasks (e.g., fitting a diagnostic label on psychiatric patients; predicting Ss responses to test items). One study by Luft (1950) found that in comparing judgments of clinical psychologists with psychology students, the clinicians were significantly superior in predicting Ss response to a projective test, but there was no significant difference for predicting their responses on an objective test. According to the present author's definitions, a projective test would be more intuitive and less analytic than an objective test, because there would be a greater likelihood of unconscious inductions and synthesis of the cues influencing the judgment on the projective test. Hence, Luft's study might suggest that clinical psychologists and psychology students are at least equal in their ability to utilize analytic processes; however, clinicians are superior at predicting intuitively.

Other studies summarized by Taft (1955) suggest that physical scientists and personnel workers are better judges than psychology students or clinical psychologists in predicting inventory responses. Such studies may constitute evidence for the greater analytical ability of such non-clinicians--but give little information as to intuitive ability, or ability to make nonanalytic judgments.

It has been suggested earlier that intuitive judgments are not necessarily random, careless, or nonscientific judgments, but are judgments of a more complex nature than are analytic judgments. The type of complexity referred to involves configural processing of cues. Cohen (1973) states that incongruent information often results in the judge drawing on an independent dimension. Such a viewpoint is supported by the study by Gollin (1954), reported earlier, where judges observed a movie of a girl who was depicted as being both kind and promiscuous. Using a fictitious example based on Cohen's (1973) own handwriting and photograph study, Cohen states:

> "An individual who is seen as 'modest' on the basis of his photograph, but whose handwriting appears 'arrogant' may evoke in the judge the impression of overcompensation, thus leading to a judgment of 'tense'. We could consider it reasonable to suppose that a large number of these phenomena which are generally treated under the mysterious concept of 'intuition' . . . can be traced to general principles of this nature" (p. 176-177).

The study by Einhorn (1971), also reported earlier, lends support to the theory that intuitive judgments are more complex than analytic judgments. When 39 engineering students rank-ordered 15 jobs in terms of their attractiveness, given two, four, or six pieces of information about each job, it was found that a conjunctive model (where a person must have a certain minimum ability on all the attributes) fit the data better than a linear model. In a much more analytic task (choosing among applicants to a graduate school in psychology), there was no systematic use of either the linear or the non-linear models. Since a conjunctive model is essentially configural (all the attributes must approach a minimal standard before the effects of single attributes are considered). Einhorn's study gives evidence to support greater configurality for intuitive tasks.

One of Brehmer's studies (1971c) might also support the task differences theory. Brehmer found a cue configuration by task predictability interaction, suggesting that the linear model fits better in a condition where task predictability is high than in a condition where it is low. Since an intuitive task is generally more vague than an analytic task, and cues may be less relevant in an intuitive task, it is likely that such tasks would have a lower task predictability and lower reliabilities. Brehmer (1971c) also found that while the fit of the linear model was equally good for all four types of cue configurations that he used, the fit was better for high predictability conditions than for low predictability conditions.

No studies have been conducted for the specific purpose of investigating the amount of linear variance and the amount of configural variance accounted for in the judgment of two different tasks. In the study described below, an attempt to reveal differences between what has been termed an analytic judgment task and what has been termed an intuitive judgment task has been made. The hypothesis is that judgments in the intuitive task condition will show a smaller proportion of variance in their main effects, and a large proportion in configural effects or interactions.

While the definition of intuitive and analytic judgments given above do not concern themselves specifically with the task differences (but rather, attempt to describe the processes used in making the judgments), intuitive tasks can probably be thought of as being more vague, while analytic tasks might be more closely related to quantitative predictions than to judgments. The first pilot study described below attempts to support such a view.

If intuitive tasks are assumed to be more vague, it would also be hypothesized that reliability, in terms of test-retest consistency, would be lower in the intuitive condition.

RESEARCH

Two experiments and two pilot studies were conducted in order to investigate differences in the way judges process information in different kinds of judgment tasks. The first pilot study attempted to distinguish between intuitive and analytic tasks, while the second was conducted as a result of disproportionate weightings that judges were giving to the cues previously selected for the judgments. The four studies are described below.

PILOT STUDY 1

The purpose of the first pilot study was to determine which judgmental tasks are considered intuitive and which ones are generally considered analytic.

Method

Fifty-three students registered in a second year social psychology course were given a questionnaire in which they were asked to indicate for each of seven judgment tasks whether they would most likely make the judgment intuitively or analytically. If they were undecided, they were asked to indicate this.

The subjects were told to assume that they would be given all the information needed to make the judgment, and that all people who were to be judged would be undergraduate students at the University of Alberta. The judgments to be made were as follows:

Likelihood of succeeding in university (getting degree)

How sociable is this person?

What is his or her earning potential?

Is he or she stable or unstable?

How popular is this person?

Is he or she basically a "good" person?

How happy is this person?

Included with the questionnaire were definitions of intuitive and analytic judgments. The definitions were as follows:

> "Intuitive judgments are usually made quickly, and require a minimal amount of detailed information. They may be based on metaphorical thinking, and/or may entail the ability to synthesize heterogeneous (and sometimes disparate) elements into a whole. They are generally considered 'commonsense' judgments".

> "Analytic judgments are usually made after much deliberation, and utilize an analytical thinking process; i.e., break down the components of the decision and weigh the evidence or the consequences of the decision (or possible consequences). They are logically deduced or consciously induced, and a fair amount of information is generally necessary before an analytical judgment is made".

Results

Ss considered "likelihood of succeeding in university" to be the judgment most likely to be made analytically. Table 1 shows that out of 53 Ss, 70% considered success in university a judgment most likely made analytically, 11% considered it most likely made intuitively, while the remainder were undecided. The earning potential judgment was the only other judgment considered more often to be analytic than intuitive, with 58% judging it as analytic, 11% intuitive, with 30% undecided.

The judgments most often considered intuitive were "How popular

TABLE 1

Results of Pilot Study 1: Likelihood of making judgments analytically or intuitively.

Judgment	Analytic	Intuitive	Undecided*
Success in university	37 (70%)	6 (11%)	10 (18%)
Sociability	8 (15%)	40 (75%)	5 (09%)
Earning potential	31 (58%)	6 (11%)	16 (30%)
Stability	16 (30%)	31 (58%)	6 (11%)
Popularity	11 (21%)	37 (70%)	5 (09%)
Goodness	5 (09%)	37 (70%)	11 (21%)
Happiness	9 (17%)	33 (62%)	11 (21%)

*Round off error of percentages results in failure of some judgments to add up to 100%.

is this person?", "How sociable is this person?", "Is he or she basically a 'good' person?", and "How happy is this person?". Table 1 illustrates the number and the proportion of subjects who judged each task as intuitive, analytic, or undecided.

Discussion

The results are not surprising if one accepts the distinctions discussed earlier between analytic and intuitive judgments, regarding relationship to the criterion. It was suggested that making a narrow prediction about an aspect of a person's future behavior would most likely employ an analytic procedure, while intuitive methods would likely be used in judging the whole person. The criterion for an intuitive prediction is generally more vague than the criterion for an analytic prediction or judgment. Sociability, happiness, popularity, and goodness are far more vague, encompassing the "whole person", (they are classed as traits) than are success in university and earning potential. The latter group can be empirically measured in time, while the former cannot. The latter suggest a more specific behavioral prediction; while the former are definitely judgments.

It is interesting to note that the relationship between criteria and type of judgment process considered most likely used is clearcut in the pilot study data, even though this criterion relationship was not part of the definition given to the subjects. The definition only defines the process, and describes the amount of information needed to make the judgment. This suggests some empirical basis for the discussion of intuitive and analytic tasks and was carried out in the previous section. The findings, however, are not consistent with the part of Taft's (1955) conception of nonanalytic or intuitive judgments that includes the clinician's predicting of behavior. However, Taft may be stressing qualitative prediction of behavior (e.g., what will the patient do next?) rather than behavior with clearcut alternatives (will he or won't he?), as is found in the "success in university" prediction, or than quantitative predictions about a specific behavior (e.g., how much of this behavior will be involved), as is found in the "earning potential" judgment.

The results of this pilot study determined the judgments that were to be made in the following experiments. Since presenting the judges with more than two sets of judgments would consume too much time, it was decided to use only the "success in university" judgment to represent the analytic task, and "sociability" to represent the intuitive task.

EXPERIMENT 1

Method

Fifty students enrolled in an introductory psychology course at the University of Alberta participated in this experiment as part of their course requirements. Each subject or judge (J) was given a questionnaire consisting of 64 profiles. Js were told that each profile represented one fictitious first year undergraduate at the University of Alberta. Five cues were presented as a part of each profile, with the score on each cue given as high or low. Each J was asked to make a judgment on a 9-point scale for each fictitious profile on "How likely is this person to succeed in university"? and "How sociable is this person?" A sample of three of the personality profiles is presented in the appendix. Three profiles, each one followed by two judgments to be made, were presented on each page. A random order of presentation of these profiles was determined, and the same order was used for all Js.

The cues used were: High-school grade point average, number of close friends, score on anxiety test, score on intelligence test, and score on test of dominance. It was believed that High-school grade point average, and score on intelligence test would be most relevant to the analytic judgment, while number of close friends and score on test of dominance would be most relevant to the intuitive judgment. Score on anxiety test was believed to be equally relevant to both judgments.

The design was a completely crossed 2^5 factorial design with each cue configuration presented twice, resulting in a total of 64 randomly ordered cue configurations. For each cue configuration, two judgments were to be made.

An analysis of variance was conducted for each subject for each of the two tasks. The relative importance of each cue (ω^2) was calculated for each subject; the mean proportion of variance ($\overline{\omega}^2$) was also calculated. The ω^2 for all main effects, and all interactions for each task (intuitive and analytic) across the 64 cue configurations (for each subject) was calculated, as well as the test-retest reliabilities (correlations between the judgments of the two presentations of the same cue configuration) for each subject on each of the two tasks. T-tests for correlated data were conducted on the differences between the intuitive and the analytic task regarding mean proportion of variance ($\bar{\omega}^2$) explained by interactions and by main effects. A t-test on the mean reliabilities (using z-transformations) between each task was also conducted. The number of significant 2-way, 3-way, and 4-way interactions for each condition was also calculated.

Results

The hypothesis that the intuitive task would result in a more configural judgment strategy than the analytic task was not supported. There was no difference between the proportion of variance accounted for by significant <u>interaction</u> effects between the two tasks, with the interaction component accounting for approximately .036 (3.6%) of the variance for each task.

There was, however, a small, but significant difference ($p<.001$ using a t-test for correlated samples) between the proportion of variance accounted for by totalling all the significant <u>main</u> effects. This difference favored the hypothesis that judges utilize a greater proportion of main effects in an analytic than in an intuitive task. In the analytic task, the significant main effects accounted for 75.1% of the total variance (average); in the intuitive task, they accounted for an average of 68.4% of the total variance.

The proportions of variance explained by each main effect, their total, and by all the interactions for the analytic and intuitive task, averaged across the fifty judges, is shown in Table 2. As well, the

TABLE 2

Mean Proportion of Variance Explained* by Significant Main Effects and Interactions for Analytic and Intuitive Tasks: Experiment 1.

EFFECT	A	B	C	D	E	TOT.	INT.	REL.
Analytic	.328	.020	.031	.343	.029	.751	.036	.781
Intuitive	.024	.552	.024	.041	.043	.684	.036	.714

*Using the $\bar{\omega}^2$ statistic (Hays, 1963). N = 50.

Meaning of Cues:
- A: Grade Point Average
- B: Number of Close Friends
- C: Anxiety
- D: Intelligence
- E: Dominance

TOT: mean proportion of variance explained by the total significant main effects.

INTER: mean proportion of variance explained by sum of the interactions.

REL: mean reliability.

mean test-retest reliabilities between the two tasks is shown. There was a significant difference ($p<.01$), using a t-test for correlated samples on transformed (Fishers Z) scores, between the test-retest reliabilities for the intuitive and the analytic task, with the analytic task being somewhat more reliable. The analytic task had a test-retest correlation, or reliability, of .781, and the intuitive task had a reliability of .714.

It should be noted that in the analytic task condition, judges concentrated on two cues--intelligence ($\overline{\omega}^2=.343$) and highschool grade point average ($\overline{\omega}^2=.328$), while the judges concentrated most of the weighting for the intuitive task on one cue--number of close friends ($\overline{\omega}^2=.552$).

The number of significant 2-way, 3-way, and 4-way interactions for all the judges for each of the two tasks was also calculated. These results are shown in Table 3. Contrary to the hypothesis, the intuitive task produced fewer interactions than did the analytic task. This difference, however, is not significant--in fact, a sign test shows an insignificant but opposite effect; i.e., 20 judges produced more interactions in the intuitive than in the analytic task, and 17 judges produced more interactions in the analytic task. Thirteen had the same number of interactions (most often zero) in both tasks. Any possible differences contrary to the hypothesis are found only in the 2-way interactions.

Looking at the specific 2-way interactions involved, a comparison between the two tasks show that the interactions each J used are quite

TABLE 3

Number of Significant 2-cue, 3-cue, and 4-cue interactions for each of the two tasks

INTERACTIONS	2-cue	3-cue	4-cue	Total
TASK				
Analytic	47	29	6	82
Intuitive	35	28	6	69

logical in light of the main effects concentrated on. Since the A effect and the D effect was strongest for the analytic task, it was hypothesized that an AxD interaction would be more pronounced in the analytic than in the intuitive task, and that 2-cue interactions with no A's or D's in it would be stronger for the intuitive task than for the analytic. Since B was the only consistently strong main effect in the intuitive task, it was hypothesized that any 2-cue interactions containing B components would be greater for the intuitive task than for the analytic, while interactions involving no B effects would be greater in the analytic task. The differences are shown in Table 4. All these hypotheses were supported at least at the 5% level of significance, using a t-test for correlated samples.

Because of the fact that the reliabilities were different, it was thought that an analysis of interactions assuming no error might reveal a significant difference between the two tasks. Thus, the interaction component was expressed as the sum of the significant interactions for each subject divided by the sum of the significant interactions plus significant main effects. Using this formula, the mean proportion of the variance, assuming no error, for the interaction components for the analytic and for the intuitive task was .043 and .056 respectively (only the data for subjects having reliabilities greater or equal to 150 in both tasks were used). The difference, however, is significant.

No significant sex differences were found in the utilization of linear and configural effects.

TABLE 4

Selected 2-cue Interactions*: Comparisons Between the Two Tasks.

INTERACTION	AxD	No A or D	B	No B
TASK				
Analytic	.012	.003	.006	.017
Intuitive	.001	.008	.013	.006

*Numbers represent $\bar{\omega}^2$ for the significant interactions for the 50 judges. All differences between analytic and intuitive tasks are significant at the .05 level.

The Pearson Product-Moment correlation between intuitive and analytic tasks across subjects was insignificant (r=.17), suggesting that there was no general classification of some judges as "intuitive" on both tasks, and others as "analytical".

Discussion

The data suggest that there are some differences in the methods used by judges in processing the cues for the different tasks. The fact that there was a significant difference in the total proportion of main effects utilized between the two tasks could imply that the judges were not using a simple linear combination of cues in the intuitive task to the same extent as in the analytic task. Another possible explanation for these results is that the greater error component in the intuitive task (due to lack of reliability) robbed the intuitive task of much of the variance to be explained by main effects. However, if this were the case, the variance explained by configural components in the intuitive task should have been deflated as well. Such was not the case. A third possible explanation is that the use of only one cue in the intuitive task affected the total proportion of main effects used in that task. However, it would appear that if a judge uses only one cue, his responses would more likely be more reliable than less reliable, and the proportion of variance explained by main effects would likely be greater rather than smaller.

Unfortunately, the fact that there was no significant difference in the proportion of variance explained by interactions makes the data less clearcut. The nonsignificant interaction difference combined with the significant reliability difference (the intuitive task was significantly less reliable) might suggest that many of the judges, while using a slightly more intuitive or nonadditive approach for the intuitive task (as suggested by the lower reliance on main effects in making the sociability ratings) are unable to use it effectively, thus resulting in a greater inconsistency (unreliability) for the intuitive task.

However, even disregarding error, there is still no significant difference in the extent of the magnitude of the significant interactions between the two tasks. Hence, other possible reasons for the failure to support the major hypothesis must be searched for, including the possibility that the hypothesis is incorrect; i.e., that there is no distinction to be made between the processing of cues for analytic and intuitive judgments--or, more extreme yet, that judges do not process cues more configurally for one type of task than for any other task. Another possibility is that the ANOVA design is not powerful enough to reveal differences in the reliance of judges on complex interactions.

One common criticism that is frequently made of the ANOVA design in measuring configural thinking is that there is a certain number of interaction components that are expected to appear by chance, and that the percentage of variance explained by the configural components may be nothing more than another source of error. However, in the present study, it was the two-way interactions that accounted for the majority of configural effects, and the particular interactions revealed were meaningful. Analysis of the two cue interactions showed that the cues that revealed the largest main effects for each task accounted for more of the configural effects for that particular task than for the other task. For example, the AxD interaction (intelligence by highschool grade point average) was significantly stronger for the "success in university" judgment than for the "sociability" judgment.

Referring back to Table 2, it can be seen that the analytic task elicited a tendency on the part of the judges to concentrate on two cues; while judges tended to focus on only one cue for the intuitive task. It is likely that rather than being a result of task differences, the differential weighting of cues is probably a function of the cues used for the two tasks. While "dominance" may be a good cue for judging a person's degree of sociability in real life, it did not function as a relevant cue in the experimental situation. It is possible that the word "dominance" does not have a strong impact on judges.

It was hypothesized that the cues would be distributed more evenly if they were chosen less arbitrarily. A second pilot study was conducted in order to discover two cues that are highly relevant to each task, plus one that is equally relevant for judging both tasks.

PILOT STUDY 2

Method

Thirty-two first year psychology students were asked to rate thirteen different cues on a 9-point scale, as to how helpful each cue would be in making each of the two judgments (i.e., success in university and sociability). Judges were told to assume that each cue would be presented as being either high or low.

Results and Discussion

Two cues that were rated high for the analytic task but low for the intuitive task, and two cues high for the intuitive but low for the analytic task were chosen. Also chosen was one cue that was rated as equally relevant to both tasks. The cues with the highest "success in university minus sociability" ratings were "score on intelligence test" and "high school grade point average"--the same cues as were chosen for the first experiment. The cues with the highest "sociability minus success in university" rating were "score on extraversion test", "number of acquaintances", and "number of close friends". Since it was felt that "number of acquaintances" was too similar in meaning to "score on extraversion test", the former was not used as a cue for the second experiment. The fifth cue selected for the experiment was "score on test of optimism-pessimism", which achieved an average rating for both the tasks. For use in the experiment, the cues selected from the pilot experiment were simplified (e.g., "score on test of optimism-pessimism" was shortened to read "optimism").

Of the two cues used for the first experiment but not the second, "score on anxiety test" (supposedly the cue that was to be equally applicable to both tasks in the first experiment) was rated slightly more useful for the "success in university" task. "Score on test of dominance" was rated the fourth most useful cue when ratings of usefulness on the "success in university" task was subtracted from ratings on the "sociability" task. Hence, "score on test of dominance", while it was rated a better cue for judging sociability than judging "success in university", was not rated nearly as useful as "number of acquaintances", "number of close friends", or "score on extraversion test".

EXPERIMENT 2

Method

A second experiment was conducted using a new set of cues, which were constructed as a result of the second pilot study. The cues used were: (a) Intelligence, (b) Number of close friends, (c) Optimism, (d) High school grade point average, and (e) Extraversion.

Other than the substituting of two new cues, the experimental procedure was the same as for Experiment 1. However, only 13 judges (volunteers) were available for this experiment.

Results

Since only 13 judges were used, no conclusive data were obtained. There were no significant differences in interactions or main effects between the intuitive and the analytic tasks; any possible differences were in the opposite direction from the hypothesis, and in the opposite direction from the differences found between the main effects and the reliabilities found in the first study. There was, however, a more even usage of cues for the two tasks than there was in the first experiment. In both tasks there was, as expected, a major concentration on two cues, a minor concentration on a third, with two cues being largely ignored by the judges. These results can be seen in Table 5.

The highest proportion of the interactions was accounted for by

TABLE 5

Mean Proportion of Variance Explained* by Significant Main Effects and Interactions for Analytic and Intuitive Tasks: Experiment 2

EFFECT / TASK	A	B	C	D	E	TOT.	INT.	REL.
Analytic	.392	.014	.075	.265	.016	.763	.038	.810
Intuitive	.021	.424	.050	.019	.274	.789	.031	.816

*using the $\bar{\omega}^2$ statistic (Hays, 1963). N=13

Meaning of Cues:
- A: Intelligence
- B: Number of Close Friends
- C: Optimism
- D: Grade Point Average
- E: Extraversion

TOT: mean proportion of variance explained by the total significant main effects.

INTER: mean proportion of variance explained by sum of the interactions.

REL: mean reliability.

the AxD interaction in the intuitive task ($\bar{\omega}^2$=.015), and by the BxE interaction in the analytic task ($\bar{\omega}^2$=.005). Two-way interactions accounted for a mean of .022 and .014 of the variance for the analytic and intuitive tasks, respectively, while three-way interactions accounted for a mean of .016 and .017 of the variance. No interaction effects between the two tasks, and no differences between specific interactions, approached significance.

Since the results, if anything, were incompatible with the first study, there seemed little use in re-analyzing the interactions, main effects, and reliabilities of the two studies combined. However, as explained in greater depth in the discussion section, a comparison of judges (with each judge-task combination treated as one "judge") who weighted one main effect .50 or above, with no other main effect weighted above .10, with those who used four or more main effects significantly, revealed a significant difference ($p<.02$), with judges who used several cues also making greater use of interactions. The $\bar{\omega}^2$ for interactions for judges using four or more cues was .039 ($n=53$), while the $\bar{\omega}^2$ (interactions) for those using primarily one cue ($n=35$) was .020. There was no significant differences for reliabilities or for main effects between judges using four or more cues, and those using primarily one cue.

Discussion of Both Experiments

The fact that the differences in the utilization of the main effects between the analytic and intuitive tasks was significant in the first experiment, but was insignificantly reversed in the second experiment, suggests that the differences in utilization of main effects may have been due to the relevance of the cues provided, rather than to the analytic and intuitive task distinction. The first study showed that in the intuitive task, the judges tended to rely primarily on one cue, while in the analytic task, two cues were used fairly equally. In the second study, the cues were constructed so that the judges would base their judgments fairly equally on two cues for both tasks, with one additional cue being used to a lesser extent. Table 4 suggests that this attempt to control cue utilization was successful. In this second study, the difference in the use of main effects between the analytic and intuitive task disappeared. The significant reliability difference that was seen in the first experiment also disappeared in the second one. Thus, it is possible that in the first study, the tendency to put all the weighting on one cue (which is, of course, related to the relevance of the cues to the task) resulted in less reliability and a smaller concentration on main effects, than when two or more cues were used. To test this, further analysis of the data was conducted in order to see if very simple utilization of data (i.e., reliance on smaller number of main effects and/or configural effects) is followed by a decrease in reliability, which in turn, restricts the ability of the data to reveal the total utilization of main effects. Such a phenomenon is noted by Schaeffer and Jackson (1970) who found that ". . . the most consistent judge in this study . . . is the one who spreads his money out over as many different cues, and their interactions as possible" (p. 17). They report, however, that in the Hoffman, Slovic, and Rorer (1968) study,

> ". . . the lawful variance appeared to be most directly under the control of the variance associated with significant main effects, while negatively related to the number of main effects. The most consistent and reliable judge, then, was the one who put all his money, as it were, on one or two diagnostic cues, and let these determine his judgments in a highly regular manner" (p. 17).

The proportion of variance accounted for by all the main effects and all the interactions for judges who used primarily one effect (i.e., had one main effect weighted .50 or above, with no other main effect weighted above .10), was tested against the proportion of variance for main effects and interactions for judges who used four or more main effects significantly. Since this test was not interested in differences between analytic and intuitive tasks, each subject-task combination was treated as one judge. In this analysis, it was possible for some subject-task combinations to be used for both groups (e.g., it is possible for a person to weight one cue on one of the tasks .50 or greater with no others weighted above .10, yet also significantly weight at least three of the other cues). Also, some subject-task combinations were excluded (i.e., those that did not weight any cue .50 or beyond, and did not significantly utilize four of the five cues).

The analysis was done across the first and second study. The total number of subject-task combinations using primarily one main effect was 35, while 53 judges used four or five of the cues significantly.

Results from a t-test for independent samples indicated no differences in the total number of main effects used, or in the reliabilities, between judges using four or more cues and those using primarily one cue. However, a significant difference ($p<.02$) was found for the use of interactions, with the judges using four or more cues significantly also using interactions more heavily ($\bar{\omega}^2=.039$) than judges who based their judgments primarily on one cue ($\bar{\omega}^2=.020$). This result suggests that judges are most likely to utilize complex interactions if the task encourages the use of several cues. The results are also consistent with the use of configural thinking as measured by the ANOVA technique, as a measure of cognitive complexity, since the number of different cues utilized, as well as the proportion of configural variance utilized by the judge, can indicate the complexity of his judgments. However, the low, insignificant correlation (.17) in the first experiment, between intuitive and analytic judgments provides evidence that the tendency to use interactions is not consistent for each subject across the two tasks, and suggests the importance of task differences. Thus, cognitive complexity may be more a function of task differences than personality differences.

While such results have interesting implications, the most important experimental hypothesis investigated in the present study is that there are some kind of task differences that will cause a judge to process cues differently in one task than in another task. The purpose of the type of experiment reported in this paper is to find two different tasks that will reveal these differences. At the moment, there is too much evidence for task differences to seriously dispute their existence. Such evidence has been revealed from the Einhorn (1971) study, and the comparison of figures from the Hoffman, Slovic, and Rorer (1968) and the Schaeffer and Jackson (1970) studies--reported in the latter study--where certain kinds of tasks resulted in higher interaction components than did other tasks.

It is of interest to compare the strength of the configural component for the two tasks reported above with the configural components in similar studies by Schaeffer and Jackson (1970), Hoffman et al (1968), Schaeffer and Saidman (1971), and Slovic (1966). Such a comparison is made in Table 6.

In the present study (Exp. 1), the average magnitude of the interactions for both the intuitive and analytic task was about .036--much lower than the proportion of the variance explained by interactions in the judgment of creativity as reported by Schaeffer and Jackson (.095). Even when interactions disregarding error were calculated (interactions divided by interactions plus main effects), the proportion of variance accounted for by interactions is .043 and .056 for the analytic and intuitive tasks, respectively--still less than the average magnitude for the Schaeffer and Jackson study. For Schaeffer and Jackson, the error component was the 6th and 7th order interactions, while in the present study, it was the response error due to presenting each set of cues twice. A more important difference concerns the judges used in the two studies. The present study used students enrolled in an introductory psychology course, while Schaeffer and Jackson used eight fine arts students and faculty members, and two engineering students. It is likely the cognitive processes of fine arts students would be less linear than that of the average first year university student, if we assume that artists are more intuitive than the average student, and if we assume that intuitive thinking is less

TABLE 6

Comparison of the Results of the Two Present Experiments with those of Schaeffer and Jackson (1970), Hoffman, Slovic, and Rorer (1968), Schaeffer and Saidman (1971), and Slovic (1969).

STUDY	1st exp Ana	1st exp Int	2nd exp Ana	2nd exp Int	S&J	HS&R	S&S	SLOVIC
EFFECTS*								
N	50	50	13	13	10	09	42	02
Largest Main	1.00	1.00	.86	.89	.85	.92	.74	.80
Average Main	.75	.68	.76	.79	.66	.71	.50	.76
Largest Inter.	.11	.14	.16	.09	.41	.03	.44	.07
Average Inter.	.04	.04	.04	.03	.10	.02	.16	.06

* The effects refer to significant ω^2 across cues.

** Interactions refer to 2-way interactions only. Fractional replications design used.

linear. Thus, the difference in the results could be explained by a subject difference rather than a task difference. Also, Schaeffer and Jackson used seven cues, while the present study used five.

For both tasks in the present experiment, and of course in the creativity judgments from the Schaeffer and Jackson experiment, the average interaction is higher than for the interactions reported by Hoffman <u>et al</u> (1968) in the judgment of ulcer malignancy by nine radiologists. The average ω^2 for interactions in the Hoffman <u>et al</u> study was .017.

The average sum-of-main-effects component for the Hoffman <u>et al</u> data (.71), however, lay between the main effects of the present study's (Exp. 1) analytic and intuitive judgments. Thus, radiologists used main effects to the same extent as the first year psychology students did in judging success in university and sociability, but they used fewer interactions. A more accurate comparison between the studies could be done if Hoffman <u>et al</u> (1968) and Schaeffer and Jackson (1970) had used more subjects.

The Schaeffer and Saidman (1971) results, where <u>J</u>s rated musical preference of several compositions in which the metre, melody, harmony, and dynamics were varied, produced a total interaction component, averaged over 42 judges (including 1st year undergraduates and some students from a music class) of .158; an interaction component which is even higher than the summation of interactions for the creativity study by Schaeffer and Jackson (1970). In terms of "intuitiveness", the comparison of the interaction components across the present two studies, the Schaeffer and Jackson (1970) study, the Hoffman, Slovic, and Rorer

(1968) study, and the Schaeffer and Saidman (1971) study seems to suggest that the studies that involve the most intuitive tasks (judging creativity and rating musical selections in terms of preference) tend to show higher interaction components than those that are most analytic (medical diagnosis and predicting gradepoint average). The Slovic (1966) stockbroker study may be an exception, in that it is an analytic study that has shown high configural components. However, this could be partially due to his use of fractional replications design, which confounds several of the main effects and interactions, and could conceivably exaggerate the configural components.

Because of the exploratory nature of this research, various difficulties with the design have resulted in the meaning of some of the results being obscured, and may have contributed to the failure to achieve some of the hypothesized results.

One confounding factor concerns the relationship between the task to be judged and the cues that are given to the judge. Ensuring, through extensive pilot studies, that cues are equally relevant for each task could reduce what would otherwise be an intuitive task to one which is analytic, since the obscurity of the cues may be an important factor in the concept of intuition. It is conceivable that both tasks, in the second study in particular, may have been too analytic. The first study may have produced differences in main effects that are totally explainable by the relevance of the cues to the judgment, since the analytic task had two cues primarily used by most Js, while the intuitive task only had one.

It is possible that one of the reasons for the failure to get significant differences in the number of interactions between the two tasks concerns the fact the two tasks contained more similarities than differences. Both tasks are analytic in the sense that they require a conscious processing of cues in order to make a rating or prediction. Unfortunately, it would be difficult to set up a design in which one of the tasks (the intuitive) could simply require the giving of verbal impressions--a response that would be much more compatible with intuitive judgments. A simpler way of widening the distinction between analytic and intuitive tasks would be through the use of differential instructions (Brunswik, 1956). For example, one group of judges could be taught to make a judgment "intuitively", while another group can be taught to make either the same or a different judgment "analytically".

While the majority of subjects in the first pilot study did consider the "success in university" judgment to be analytic, and the "sociability" judgment intuitive, it is likely that many judges would find such a distinction more meaningful for another pair of judgments. Some of the vast individual differences seen in the present study could be eliminated if each judge were tested individually on a task that the individual judge considered either analytic or intuitive.

Finally, it should be stressed that making a total of 128 separate judgments, processing five cues for each judgment, takes a great deal of energy from each judge. During the experiments, there were indications of boredom and fatigue from many of the judges, no doubt making the judgments less meticulous than would be the case if the judgments were made outside of the experimental situation where the J makes a judgment about one person on a task that is fairly important to that particular judge. Somewhat more accurate results could be produced if each judge only had to judge the intuitive or the analytic task, and/or if fewer cues or cue configurations were given to each judge. Such a design would probably increase the reliabilities of the judgments.

CONCLUSION

This study was not concerned with explaining the nature of intuitive and analytic judgments; such a task is beyond the scope of this research. Its primary purpose was to find task differences in judgment strategy. When such differences are found, other studies will no doubt be conducted to determine the specific nature of these differences, and whether the analytic-intuitive distinction has any empirical basis.

From the above study, the significant main effects difference found from the first experiment, and the relationship suggested between main effects and interactions (i.e., that judges who used four or more significant main effects had a greater proportion of their variance explainable by interactions than did judges who based their judgments primarily on one cue) give some support to the existence of task differences affecting the judgment process. Comparisons between the above results and those of other studies also lends support.

There is evidence that the ANOVA method and the ω^2 statistic can be used to determine which kinds of judgments are generally made using a configural judgment process, and which ones are almost purely linear. If we choose to define judgments that utilize configural components as intuitive, and those that are primarily linear as analytic judgments, the use of the ANOVA may be able to shed some light on these two different kinds of judgment processes. Such research could be of value both in explaining what a lot of clinical psychologists claim they are doing, and may someday be used for teaching potential clinicians to make the most of the cues they are presented with. It may teach clinicians the situations where an actuarial method can be most efficient, as distinct from situations in which clinical intuition may be used.

APPENDIX:

SAMPLE OF FIRST THREE PROTOCOLS (EXPERIMENT 1)

1. <u>low</u> High school grade point average
   <u>high</u> Number of close friends
   <u>high</u> Score on anxiety test
   <u>high</u> Score on intelligence test
   <u>low</u> Score on test of dominance

   How likely is this person to succeed in University?

   not at all / / / / / / / / / / extremely

   How sociable is this person?

   not at all / / / / / / / / / / extremely

2. <u>high</u> High school grade point average
   <u>high</u> Number of close friends
   <u>high</u> Score on anxiety test
   <u>high</u> Score on intelligence test
   <u>high</u> Score on test of dominance

   How likely is this person to succeed in University?

   not at all / / / / / / / / / / extremely

   How sociable is this person?

   not at all / / / / / / / / / / extremely

3. <u>low</u> High school grade point average
   <u>low</u> Number of close friends
   <u>high</u> Score on anxiety test
   <u>low</u> Score on intelligence test
   <u>high</u> Score on test of dominance

   How likely is this person to succeed in University?

   not at all / / / / / / / / / / extremely

   How sociable is this person?

   not at all / / / / / / / / / / extremely

REFERENCES

Allport, F. H. Social Psychology. Cambridge: Riverside Press, 1924.

Allport, G. W. Personality. New York: Holt, 1937.

Allport, G. W. The general and the unique in psychological science. In Eugene Southwell & Michael Merbaum (Ed.), Personality, readings in theory and research, Belmont, California: Wadsworth Co., 1965, 244-258.

Anderson, N. H. Application of an additive model to impression formation. Science, 1962, 138, 817-818.

Anderson, N. H. Averaging versus adding as a stimulus combination rule in impression formation. Journal of Experimental Psychology, 1965, 70, 394-400.

Anderson, N. H. Application of a linear-serial model to a personality-impression task using serial presentation. Journal of Personality and Social Psychology, 1968, 10, 354-362.

Anderson, N. H. Comment on "An ANOVA model for assessment of cue utilization in clinical judgment". Psychological Bulletin, 1969, 72, 63-65.

Anderson. N. H. Looking for configurality in clinical judgment. Psychological Bulletin, 1972, 78, 93-102.

Anderson, N. H., & Jacobson, A. Effects of stimulus inconsistency and discounting instructions in personality impression formation. Journal of Personality and Social Psychology, 1965, 2, 531-539.

Brehmer, B. Cognitive dependence on additive and configural cue-criterion relations. American Journal of Psychology, 1969, 82, 490-503.

Brehmer, B. Policy conflict as a function of policy similarity and policy complexity. UMEA Psychological Reports, 33, 1971(a).

Brehmer, B. Effects of cue validity and function form upon interpersonal learning to make inductive inferences. UMEA Psychological Reports, 34, 1971(b).

Brehmer, B. Cue utilization and cue consistency in multiple-cue probability learning. UMEA Psychological Reports, 44, 1971(c).

Brehmer, B. Cue utilization in multiple-cue probability learning tasks with intercorrelated cues. *UMEA Psycholøgical Reports*, *45*, 1971(d).

Brehmer, B. Concurrent learning of two single-cue probability learning tasks. *UMEA Psychological Reports*, *41*, 1971(e).

Brehmer, B. Effects of task predictability and cue validity on interpersonal learning of inference tasks involving both linear and nonlinear relations. *UMEA Psychological Reports*, *64*, 1973.

Brehmer, B. Inductive inference from uncertain information. *UMEA Psychological Reports*, *78*, 1974.

Brunswik, E. Representative design and probabilistic theory. *Psychological Review*, 1955, *62*, 236-242.

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

Bunge, M. *Intuition and science*. New Jersey: Prentice Hall, 1962.

Cohen, R. *Patterns of personality judgment*. trans. D. L. Schaeffer. New York: Academic Press, 1973.

Cox, D. R. *Planning of experiments*. New York: Wiley Press, 1958.

Dawes, R. M. The mind, the model, and the task. *Proceedings of the seventh annual Indiana theoretical and cognitive psychology conference*, Bloomington, Indiana, 1974.

Dawes, R. M., & Corrigan, B. Linear models in decision making. *Psychological Bulletin*, 1974, *81*, 95-106.

Dudycha, L. W., & Naylor, J. C. Some characteristics of the human inferential process in complex choice behavior situations. *Organizational Behavior and Human Performance*, 1966, *1*, 110-126.

Einhorn, H. J. The use of nonlinear, noncompensatory models in decision making. *Psychological Bulletin*, 1970, *73*, 221-230.

Einhorn, H. J. The use of nonlinear, noncompensatory models [illegible]tion of task and amount of information. *Organizational Beha[illegible] Human Performance*, 1971, *6*, 1-27.

Ewing, A. C. *Reason and intuition*. Annual Philosophical Lecture: Henriette Hertz Trust, British Academy, 1941.

Goldberg, L. R. Diagnosticians vs. diagnostic signs: the diagnosis of psychosis vs. neurosis from the MMPI. *Psychological Monographs*, 1965, 79.

Goldberg, L. R. Simple models or simple processes? Some research on clinical judgments. American Psychologist, 1968, 23, 483-496.

Goldberg, L. R. The search for configural relationships in personality assessment. The diagnosis of psychosis versus neurosis from the MMPI. Multivariate Behavioral Research, 1969, 4, 523-536.

Goldberg, L. R. The proliferation of personality scales and inventories: an historical analysis. In P. McReynolds (Ed.), Advances in psychological assessment, Vol. 2, New York: Science and Behavior Books, 1970.

Goldberg, L. R. Five models of clinical judgment: an empirical comparison between linear and non-linear representations of the human inference process. Organizational Behavior and Human Performance, 1971, 6, 458-479.

Gollin, E. S. Forming impressions of personality. Journal of Personality, 1954, 23, 65-76.

Grebstein, L. C. Relative accuracy of actuarial prediction, experienced clinicians, and graduate students in a clinical judgment test. Journal of Consulting Psychology, 1963, 27, 127-132.

Green, B. F. Jr. Descriptions and explanations: A comment on papers by Hoffman and Edwards. In Kleinmuntz, B. (Ed.), Formal representation of human judgment. New York: John Wiley & Sons, 1968.

Hammond, K. R. Probabilistic functioning and the clinical method. Psychological Review, 1955, 62, 255-262.

Hammond, K. R., Hursch, C. J., & Todd, F. J. Analyzing the components of clinical inference. Psychological Review, 1964, 71, 438-456.

Hammond, K. R., & Summers, D. A. Cognitive dependence on linear and non-linear cues. Psychological Review, 1965, 72, 215-224.

Hankin, H. Common sense and its cultivation. London: Kegan, Trench, Truber, & Co. Ltd., 1928.

Hays, W. L. Statistics for psychologists. New York: Holt, 1963.

Hoffman, P. J. The paramorphic representation of clinical judgment. Psychological Bulletin, 1960, 57, 116-131.

Hoffman, P. J. Cue-consistency and configurality in human judgment. In Kleinmuntz, B. (Ed.), Formal representation of human judgment. New York: John Wiley & Sons, 1968, 53-90.

Hoffman, P. J., Slovic, P., & Rorer, L. G. An analysis-of-variance model for the assessment of configural cue utilization in clinical judgment. Psychological Bulletin, 1968, 69, 338-349.

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

Jackson, I. Cue utilization and the process of judging creativity. Honors thesis, University of Alberta, Spring, 1970.

Johnson, D. M. The psychology of thought and judgment. New York: Harpers, 1955.

Kleinmuntz, B. The processing of clinical information by man and machine. In Kleinmuntz, B. (Ed.), Formal representation of human judgment. New York: John Wiley & Sons, 1968.

Knox, R. E., & Hoffman, P. J. Effects of variation of profile format on intelligence and sociability judgments. Journal of Applied Psychology, 1962, 46, 14-20.

Lampel, A. K., & Anderson, N. H. Combining visual and verbal information in impression formation. Journal of Personality and Social Psychology, 1968, 9, 1-6.

Lundberg, G. A. Case studies vs. statistical methods--an issue based on misunderstanding. Sociometry, 1941, 4, 379-383.

McKeachie, W. J. Lipstick as a detminer of first impressions of personality. Journal of Social Psychology, 1952, 36, 241-244.

Meehl, P. E. Configural scoring. Journal of Consulting Psychology, 1950, 14, 165-171.

Meehl, P. E. Clinical versus Statistical Prediction. Minneapolis: University of Minnesota Press, 1954.

Meehl, P. E. A comparison of clinicians with five statistical methods of identifying psychotic MMPI profiles. Journal of Consulting Psychology, 1959, 6, 102-109.

Meehl, P. E. The cognitive activity of the clinician. American Psychologist, 1960, 15, 19-27.

Naylor, J. C., & Wherry, R. J. The use of simulated stimuli and the "Jan"-technique to capture and cluster the policies of raters. Educational and Psychological Measurement, 1965, 25, 969-986.

Newell, A. Judgment and its representation: an introduction. In Kleinmuntz, B. (Ed.)., Formal representation of human judgment. New York, N.Y.: John Wiley & Sons, 1968.

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

Peterson, C. R., Hammond, K. R., & Summers, D. A. Optimal responding in multiple-cue probability learning. Journal of Experimental Psychology, 1965, 70, 270-276.

Rorer, L. G., Hoffman, P. J., Dickman, H. R., & Slovic, P. Configural judgments revealed. Proceedings of the 75th annual convention of the American Psychological Association, Vol. 2. Washington, D.C. APA, 1967, 195-196.

Rorer, L. G., & Slovic, P. The measurement of changes in judgmental strategy. American Psychologist, 1966, 21, 641-642.

Sarbin, T. R. Clinical psychology--art or science. Psychometrika, 1941, 6, 391-400.

Sarbin, T. R., Taft, R., & Bailey, D. E. Clinical inference and cognitive theory. New York: Holt, 1960.

Schaeffer, D. L., & Jackson, I. Judging creativity in art. Paper presented at Western Psychological Association, Los Angeles, 1970.

Schaeffer, D. L., & Saidman, L. F. Some dimensions of musical preference. Paper presented at Western Psychological Association, San Francisco, 1971.

Schumer, R., Cohen, R., & Schwoon, D. Some comments on linear models of diagnostic judgment. Trans: D. L. Schaeffer. Original version: Zeitschrift fur experimentelle und angewandte Psychologie, 1968, 15, 336-353.

Shanteau, J. C., & Anderson, N. H. Test of a conflict model for preference judgment. Journal of Mathematical Psychology, 1969, 6, 312-325.

Sidowski, J. B., & Anderson, N. H. Judgments of city-occupation combinations. Psychonomic Science, 1967, 7, 279-280.

Slovic, P. Cue consistency and cue utilization in judgment. American Journal of Psychology, 1966, 79, 427-434.

Slovic, P. Analyzing the expert judge: a descriptive study of a stockbrokers' decision processes. Journal of Applied Psychology, 1969, 53, 255-263.

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

Smith, H. C. Sensitivity to people. New York: McGraw-Hill, 1966.

Summers, D. A. The learning of responses to multiple weighted cues. Journal of Experimental Psychology, 1962, 64, 29-34.

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

Taft, R. The ability to judge people. Psychological Bulletin. 1955, 52, 1-21.

Wiggins, N., & Hoffman, P. J. Dimensions of profile judgments as a function of instructions, cue consistency, and individual differences. Multivariate Behavioral Research, 1968(a), 3, 3-20.

Wiggins, N., & Hoffman, P. J. Three models of clinical judgment. Journal of Abnormal Psychology, 1968(b), 73, 70-77.

Yntema, D. B., & Torgerson, W. S. Man-computer cooperation in decisions requiring common sense. IRE Transaction of the Professional Group on Human Factors in Electronics, 1961, 2(1).