LINEAR AND NONLINEAR CUE UTILIZATION IN THE IDENTIFICATION OF INDIVIDUAL MEMBERS OF TWO BIVARIATE NORMAL POPULATIONS

CHRISTOPHER DRACUP

Department of Psychology

University of Stirling

Submitted in partial fulfilment of the requirements of the degree of Doctor of Philosophy

May 1975

Acknowledgments

Thanks are due to Mr. M. F. Moore for his invaluable help and advice over the last three years, to Mr. A. Annan for his help with the equipment, and to all the staff and postgraduates of the Department of Psychology for their encouragement and support. I would like to thank Dr. Y. Topsever, Mr. L. Kirkman and Mrs. J. Dennison for their help in the preparation of the thesis. My particular thanks are due to my wife, parents, brothers and sisters, without whom it would not have been possible.

SUMMARY

An attempt was made to investigate the decision processes of subjects in a bivariate decision making task, similar to that facing a medical specialist who is required to classify a patient as belonging to one of a number of possible disease populations on the basis of the patient's scores of two predictor cues. It was felt that such tasks had been largely neglected in experimental psychology, where the tendency has been towards requiring subjects to learn relationships between continuous predictor variables and a continuous criterion, rather than between continuous predictor variables and a categorical criterion.

When the relationship between the predictor variables is the same in both the populations to be discriminated, the best decision function is based on a linear combination of the cues (Fisher's Linear Discriminant Function). It was found that the decisions of those subjects who learned to use the cues in a way which was at all valid in such situations, could be well approximated by a model which weighted the two cues equally in a linear combination and based it's decisions on the result.

When the relationship between the predictor variables differs from one population to the other, however, the best decision function becomes more complex, including terms in the squares and cross-products of the cues. It was felt that such situations are particularly relevant to medical decision making where clinicians have frequently claimed that the "pattern" of scores of a patient is important, not just the individual scores on each cue. It was found that if differences in cue inter-correlation were large, then subjects seemed to include in their

decision processes, some nonlinear term to take account of this fact. If, however, differences in cue intercorrelation were only moderate, or if the correlations involved were large but negative, this seemed to go unnoticed by the subjects and did not lead to any reliance on nonlinear terms.

The results show that previous findings in "real life" tasks, that decision making processes could be adequately represented as linear combinations of cues, may be due more to the linear nature of the tasks than to any predisposition towards linear processes on the part of human decision makers, and that the statistical properties of "real life" tasks must be more thoroughly investigated before it is assumed that they require nonlinear decision processes.

CONTENTS

	PAGE
ACKNOWLEDGMENTS	ii
SUMMARY	iii
GENERAL INTRODUCTION	1
LITERATURE REVIEW	6
"REAL LIFE" TASKS	6
LEARNING TASKS	32
EXPERIMENT I	
DECISIONS BETWEEN BIVARIATE NORMAL POPULATIONS WITH UNEQUAL MEAN VECTORS BUT EQUAL COVARIANCE MATRICES	62
EXPERIMENT II	
DECISIONS BETWEEN BIVARIATE NORMAL POPULATIONS WITH UNEQUAL MEAN VECTORS AND UNEQUAL COVARIANCE MATRICES	124
EXPERIMENT III	
FURTHER INVESTIGATION OF DECISIONS BETWEEN BIVARIATE NORMAL POPULATIONS WITH UNEQUAL MEAN VECTORS AND UNEQUAL COVARIANCE MATRICES	181
BIBLIOGRAPHY	249
APPENDIX A STATISTICAL FORMULATION OF BRUNSWICK'S LENS MODEL	259
APPENDIX B THE GENERATION OF SAMPLES FROM BIVARIATE NORMAL POPULATIONS	264
APPENDIX C STATISTICAL APPENDIX	266

GENERAL INTRODUCTION

This research is concerned with the ability of human subjects to learn to combine information from more than one unreliable cue, in order to classify a person or object as belonging to one of a 'finite' number of populations. The experiments to be described were intended to represent what is traditionally regarded as being the situation facing a physician engaged in medical diagnosis. The medical specialist, when faced with the task of diagnosing the complaint of a patient, has a number of information sources open to him (e.g. records, test results). The information from these sources must be combined in some way so as to enable him to decide into what diagnostic category the patient should be placed.

It should be pointed out that much of medical decision making is not concerned with diagnosis of the above kind. The approach to many problems would appear to be far more empirical; a choice between a number of alternative actions e.g. administration of one drug rather than another, drug therapy rather than surgery; with future actions dependent on the outcome of previous ones. Such an approach is adopted since there will not, in general, be a one to one relationship between disease and treatment; one treatment being a possible cure of a number of complaints and one complaint being susceptible to a number of different treatments. In such a context then it can be seen that diagnosis, as described above, may only have a secondary role. The prime aim being to find a treatment which will bring relief to a suffering patient. The pure diagnosis task is in fact probably most familiar to workers in psychological medicine and it is in this field that most of the psychological investigation has been conducted.

What research has been carried out into the behaviour of subjects in tasks of the above kind, in an attempt to discover the way in which such decisions are reached, and the accuracy of the final diagnosis, has, by the very nature of the task, had to leave variables uncontrolled e.g. the experiences of the judges, the accuracy and degree of feedback (often due to the lack of an objective criterion), the representativeness of the samples presented etc. For this reason, in the experiments to be described, pre-existing decision processes were not investigated, rather, subjects were provided with situations in which they could learn the relationships between cues and criterion.

Previous research on the ability of subjects to learn the relationships between cues and criterion has differed from that reported here in a number of ways. The most fundamental difference being that in the present experiment there exist two populations which are to be discriminated whereas in previous studies the task facing the subject has been the subdivision or partitioning of one general population. It is felt that the difference in approach is probably a reflection of two different kinds of task occurring in real life. For example, if one is required to distinguish between intelligent and unintelligent faces it would seem more reasonable to regard these as members of one overall population of faces rather than regard "intelligent" faces as one population and "unintelligent" faces as another. If, on the other hand, one was required to distinguish between the faces of mongols and the faces of normals, it would be far more reasonable to regard these as two distinct populations. The distinction being made is between those differences due to the action of a large number of factors, as in the determination of intelligence, and those due to some underlying fundamental cause (probably physiological or genetic) as in the determination of mongolism. This distinction is not however always easy to make, as is evidenced by the conflict between the "type" and "trait" theories of personality (c.f. Bischof 1964). The former regard such groups as schizophrenics, neurotics etc., as populations in their own right, whereas, the latter regard them simply as the extreme tails of the population of normal individuals.

It seems particularly appropriate to regard decision categories as distinct populations, when there is a reason to believe that the interrelations between particular variables differs from one category to another, clinicians, for instance, have often claimed that they consider the "pattern" of scores of patients, not just the individual score values, (c.f. Meehl (1954); Hoffman, Slovic and Rorer (1968)). We might consider why clinical psychologists should go to the trouble of joining up the scores of an individual on the sub-tests of the M.M.P.I.

distinguish between intelligent and unintelligent faces it would seem more reasonable to regard these as members of one overall population of faces rather than regard "intelligent" faces as one population and "unintelligent" faces as another. If, on the other hand, one was required to distinguish between the faces of mongols and the faces of normals, it would be far more reasonable to regard these as two distinct populations. The distinction being made is between those differences due to the action of a large number of factors, as in the determination of intelligence, and those due to some underlying fundamental cause (probably physiological or genetic) as in the determination of mongolism. This distinction is not however always easy to make, as is evidenced by the conflict between the "type" and "trait" theories of personality (c.f. Bischof 1964). The former regard such groups as schizophrenics, neurotics etc., as populations in their own right, whereas, the latter regard them simply as the extreme tails of the population of normal individuals.

It seems particularly appropriate to regard decision categories as distinct populations, when there is a reason to believe that the interrelations between particular variables differs from one category to another, clinicians, for instance, have often claimed that they consider the "pattern" of scores of patients, not just the individual score values, (c.f. Meehl (1954); Hoffman, Slovic and Rorer (1968)). We might consider why clinical psychologists should go to the trouble of joining up the scores of an individual on the sub-tests of the M.M.P.I.

or the 16 P.F. to form a personality profile, if they did not think that this extra cue (i.e. the pattern of the scores) was of some predictive use to them.

The configural and curvilinear utilization of cues is a topic which has been of great interest to researchers investigating decision making behaviour in real life tasks, but these researchers, have always been at the disadvantage of not knowing whether the tasks required non-linear utilization of cues, and if they did, what manner of non-linearity was required. One advantage of the present experiments is that the research can be designed so as to imply the validity of certain non-linear cues known to the experimenter, without making the tasks unnecessarily artificial.

In the following pages, three experiments are described which were designed to investigate the effects of various features of a two population discrimination task on a number of behavioural indices, but before these are described in detail, the papers forming the psychological background of the present research will be reviewed. First, the research dealing with behaviour in real life tasks will be discussed, its methods, aims and limitations; and then that research dealing with the ability of subjects to learn relationships between cues and criterion will be discussed. These pieces of research will be drawn upon in the individual introductions to the experiments, but are reviewed together at this point in order to make their presentation more systematic. The notation followed throughout this introduction is that provided by the statistical formulation of Brunswicks Lens Model, and a brief outline of this is given in Appendix A.

LITERATURE REVIEW

The psychological research into this area of decision making may be divided into two categories:-

1 That research dealing with "real life" situations, where the judges are assumed to possess already, decision processes which relate the predictor cues to the criterion.

2 That research dealing with the behaviour of subjects in situations where no such process can be assumed to exist, but where subjects are given an opportunity of developing suitable decision processes over the period of experimentation.

"Real Life" Tasks

In the real life situations, it is not always clear whether the experimenter regards the task as one in which a decision must be made between two or more populations, or one in which the decision categories are simply subdivisions of one population, though in most of the experiments one of the two alternatives is clearly a better representation of the situation. Such a distinction might not lead researchers in this area to alter their experimental design or methods of data analysis to any great degree, but it could have made clear some of the drawbacks of those experiments attempting to simulate such situations in learning tasks.

The earliest research efforts of psychologists in the field of clinical judgements yielded rather discouraging results. For many tasks in psychological medicine the amount of training and experience of the judge was found not to relate to his judgemental accuracy (c.f. Goldberg 1959; Johnston and McNeal, 1967; Luft, 1950; Oskamp, 1962). Equally disheartening, there now exist a number of studies demonstrating that the amount of information available to the judge is not related to the accuracy of his resulting inferences. (c.f. Goldberg, 1968).

As a consequence of these sorts of findings, the focus of research turned from such validity studies to investigations of the process of clinical inference, the aim of which is to "represent" (or "simulate" or "model") the hidden cognitive processes of the clinician as he makes his judgemental decisions. Hopefully, with a better understanding of the processes, clinical training procedures may be improved, and judgemental accuracy increased.

This research then, attempts to discover some manipulation or transformation which, when applied to the predictor cues, yields responses or predictions identical to those of the human judge. If such a transformation can be found, then it is implicitly assumed that similar processes take place within the clinician.

Hoffman (1960) has pointed out some of the possible pitfalls in such an approach. He distinguishes between "isomorphic" and "paramorphic" representations of human decision processes (borrowing the two terms from minerology). An "isomorphic" representation of a process is a true model in every sense. It predicts all the reactions of the process perfectly, under all conditions. As such, it is the unattainable end of all scientific investigation. However, if the model is able to predict the judges responses well (but not perfectly) it is said to be a "paramorphic" representation of the judgemental process. The model helps to account for or "explain" many, though not all, properties or characteristics of the process. The model is also useful in making predictions concerning the outcomes of certain other tests which may be employed. But, as with chemical analysis, the mathematical description of judgement is inevitably incomplete, for there are other properties of judgement still undescribed, and it is not known how completely or accurately the underlying process has been represented.

It is possible that two or more models may be capable of accounting for judgement variance with equal efficiency. Consider for example, a given model which is highly accurate in predicting judgements from the information given. In this sense, we may be said to have "described" or "characterized" the judgement process, but one important qualification is necessary. Even in the hypothetical situation in which prediction is perfect, one cannot conclude that the mental process has been "discovered" even among sets of mathematical models which are ostensibly different, there may be some which are in fact equivalent with respect to explanatory power.

For example - let us assume that for a particular judge, his judgements can be predicted from X_1 and X_2 with 95%

accuracy by the following equation.

$$\hat{Y}_s = +\sqrt{X_1^2 + 2X_1X_2 + X_2^2}$$

The right hand term is simply the positive square root of the binomial$(X_1 + X_2)^2$. Since $X_1 + X_2 = +\sqrt{X_1^2 + 2X_1X_2 + X_2^2}$ it follows that

$$\hat{Y}_s = X_1 + X_2$$

will account for the judgements equally well. It is therefore, no more reasonable to conclude that the judge is in fact "using" one particular combination of the information than it is to conclude that he is "using" the other. Different criteria must be established before such a choice may be intelligently made.

The question arises, what sort of judgemental models should be tried? Since introspective accounts have been interpreted as implying curvilinear, configural and sequential judgement processes, (e.g. McArthur 1954, Meehl 1954, 1960; Parker 1958) one possible strategy is to begin with fairly complex representations, perhaps with an eye to seeing how they may be simplified. For example:- Kleinmuntz (1968) had a clinical psychologist "think aloud" into a tape recorder, as he made judgements about the adjustment of college students, on the basis of their M.M.P.I. profiles. Kleinmuntz then used these introspections to construct a computer program simulating the clinician's thought processes. The resulting program was a complex sequential (e.g. hierarchical or tree) representation of the clinician's verbal reports. Such an approach has many points of interest, but parsimony is one of the aims of science, and the next approach to be discussed is far more acceptable on this ground.

At the other end of the complexity scale, the investigators at two major research centres - The Oregon Research Institute and the Behavior Research Laboratory of the University of Colorado - have concentrated attention on a very simple model which could be modified to take in more complex terms, if this was shown to be necessary. They likened the judgemental process to a simple linear weighting of the cue dimensions:-

$$\hat{Y}_s = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_N X_N$$

Such a model had been used by Meehl (1954) in a comparison of the accuracies of clinical and actuarial prediction and had been found to perform significantly better than the human judge. Now, however, the model is being used not to predict some objective criterion, but to predict the judgements of the clinician.

In the light of the previously mentioned claims of clinicians that their cognitive processes are complex ones, involving the nonlinear utilization of cues, it would be expected that the linear additive model would provide a poor representation of their judgements. Consequently, it might be anticipated that it would be necessary to introduce into the model, mathematical expressions to represent such processes.

The curvilinear utilization of a cue may be represented by including in the model a term in the square or higher power of that cue.

e.g. :-

$$\hat{Y}_s = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_N X_N + \alpha_1 X_1^2$$

Just which powers of which cues will be included will depend on the type of curvilinearity claimed. For example, if it was claimed that a salesman was more likely to be successful if he was of average intelligence, but less likely to be successful if of high or low intelligence, a term in the square of the intelligence cue might be used to represent this claim in predicting the personnel selectors estimates of success of applicants.

Clinicians frequently call attention to their use of cues in (what is apparently) a curvilinear fashion, but even more frequently do they claim to use cues in (what seems to be) a configural (or interactive) manner, i.e. they claim that the interpretation of a given dimension is conditional upon the values of other dimensions. For example, a physician might feel that body temperature (X_2) is positively related to the likelihood of some disease if a patient has a high score on some other variable (X_1), while body temperature has lesser relevance to this decision if the patient has a low score on variable X_1 . Such claims might be represented in the model by the inclusion of cross product terms. Clearly the weight given to body temperature increases as X_1 increases.

β_2 might be linearly related to X_1 , in which case

$$\beta_2 = \alpha_0 + \alpha_1 X_1$$

so $$\hat{Y}_s = \beta_1 X_1 + (\alpha_0 + \alpha_1 X_1) X_2 + \dots + \beta_N X_N$$

or $$\hat{Y}_s = \beta_1 X_1 + \alpha_0 X_2 + \alpha_1 X_1 X_2 + \dots + \beta_N X_N$$

More complex terms may of course be included e.g. terms in logarithms, terms in cross products of powers of cues or any other transformation of the cues which might be felt to represent the types of nonlinearity claimed.

It is hoped then, that the inclusion of such curvilinear and interactive terms will, if they have any validity, produce significantly better predictions of the judges responses, thus justifying the claims of clinicians that they do not combine information in a simple, linear way.

Before we go on to look at some of the tools used by researchers in their attempts to model the judgemental process in such efforts, we must first discuss a problem which arises from the use of the term "Linear Model".

To the statistician a "linear model" is a model which is linear in its parameters. (c.f. Mood and Graybill, 1963) For example:-

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \qquad (1)$$

$$Y = \beta_1 X_1 + \beta_2 X_2^2 + \beta_3 X_1 X_2 \qquad (2)$$

are both linear models as the parameters (β_is) are linearly related to Y.

$$Y = X^{\beta}$$

is not a linear model however since Y is not a linear function of β.

In the psychological literature however, the linear - nonlinear distinction differentiates models which are linear or nonlinear in the predictor variables. Model (1) above, therefore, is regarded as linear, but model (2) is regarded as nonlinear since it includes terms in cross products of variables, and terms in powers higher than one.

The linear - nonlinear distinction made in the psychological literature will be followed here. In cases where the statistical distinction is referred to, this will be made explicit.

Models which are linear in their parameters, may be evaluated using the technique of multiple regression. This technique finds those parameters of the model in question which best predict (to a least squares criterion) the dependent variable and gives a measure of the precision of prediction in R^2 , the amount of variance in the dependent variable which can be explained by the model. Even with models which are nonlinear in their parameters, it is often possible to find transformations which permit the estimation of the parameters by multiple regression.

R^2 may be used as a guide to what terms should be included in modelling a judgemental process. If the judge is using cues in a curvilinear or configural manner then the inclusion of cross products and power terms in the model should increase R^2 significantly over its value with purely linear terms. A significant increase in R^2 has therefore often been regarded as sufficient grounds for including an additional term in the model.

The analysis of variance is another technique, developed to test models linear in their parameters, which has been used in research of this kind. The predictor variables become the factors of the analysis of variance and particular values of these variables become the levels. If the levels are chosen according to prescribed rules, then tests of significance of curvilinearity may be carried out using the method of orthogonal polynomials. Configural use of cues will be shown by significant interaction terms.

One or the other of these two techniques has been used by the vast majority of researchers in this field. We now turn to the findings of these researchers in what was to become almost a frantic search for non-linearity.

It was found in study after study that the accuracy of the linear model in predicting judgements was almost always at approximately the same level as the reliability of the judgements themselves. The introduction of curvilinear and configural terms into the basic model rarely served to significantly increase goodness of fit. Hammond and Summers (1965) reviewed a series of studies in which the same general finding emerged from a number of different judgement tasks, and across a considerable range of judges. The simple linear model appeared to characterise quite adequately the judgemental processes involved, in spite of the reports of the judges that they were using cues in a highly configural manner.

It is possible of course, that the particular tasks studied, did not call for the use of cues in a configural or curvilinear way. It was necessary therefore, to find tasks where non-linear cue utilization was most likely to be required, since it is in such tasks that configural judgemental processes were likely to be developed.

Hoffman, Slovic and Rorer (1968) felt that they had found such a task in the diagnosis of benign versus malignant gastric ulcers. They had been assured that there were seven major signs which could be seen on X-rays of gastric ulcer patients, and that the diagnosis was only possible through configural use of the seven cues.

A six factor experiment was set up, five of the cues taking only two levels, present or absent, the sixth variable (a combination of two of the original cues), took three levels. Nine clinicians made diagnoses of the resulting $2^5 \times 3$ possible combinations of the cues on a seven-point scale, ranging from "definitely benign" through "uncertain" to "definitely malignant". Each of the 92 resulting combinations was in fact judged twice by each clinician thus allowing an assessment of the reliability of their decisions.

The inferences of each judge were analysed by the analysis of variance model in order to ascertain the proportion of variance in his diagnoses associated with each of the six possible main effects (the use of cues in a linear fashion), the fifteen two-way interactions, twenty three-way interactions, fifteen four-way interactions, six five-way interactions, and one six-way interaction (the configural use of cues).

The major finding of this study was the small amount of variance due to even the largest interaction effect i.e. 3%. The largest main effect, in general accounted for 10 to 40 times as much of the total variance in subjects' judgements as the largest interaction. On average about 90% of a judge's reliable variation in judgements could be accounted for by a linear combination of the individual symptoms, disregarding the interactions. However, many of the interaction terms were significant, showing that the judges used cues in a configural manner, they just did not appear to use them very much in this way. However disheartening these results may have been, indices of interjudge agreement were even more discouraging. The median value was .38 and this was not due to the unreliability of the judgements, which were in general highly reliable. This level of agreement is not much better than that found among judges in tasks in clinical psychology.

Another task expected to yield high reliance upon configural use of cues was the decision whether or not to grant temporary liberty to a psychiatric patient (Rorer, Hoffman, Dickman and Slovic 1967). Six cues, of binary nature were used in this task (e.g. Does the patient have a drink problem? Yes/No), yielding 2^6 or 64 possible cue combinations, which were presented twice to each judge. Six physicians, twelve nurses, three clinical psychologists, and three psychiatric social workers served as judges. Each judge decided whether each one of the 128, presumably real patients should be granted the privilege of leaving the hospital for eight hours on a weekend. Again the data were subjected to the analysis of variance and the proportions of variance due to main (linear) and interaction (configural) effects were computed.

The results were found to be of a very similar nature to those of the previous study. On average, less than 2% of the variance of the judgements was associated with the largest interaction terms. The percentages ranging from virtually zero for some subjects to 6% for others.

Wiggins and Hoffman (1968) set out to investigate the way in which the clinical psychologist draws inferences concerning neuroticism versus psychoticism from M.M.P.I. profiles, and particularly the extent to which these processes were nonlinear in nature.

They used the data originally collected by Meehl and Dahlstrom (Meehl 1959), from thirteen Ph.D. clinical psychologists and sixteen predoctoral trainees. These twenty nine judges were given seven samples of M.M.P.I. profiles and were required to sort the members of each sample on an eleven-point scale ranging from "neurotic" through "neutral" to "psychotic". Each profile consisted of scores on eight clinical scales and on three validity scales. The judges were told no more about the samples than that they represented males who were under psychiatric care, and who had been diagnosed as psychotic or neurotic.

The judgements of the clinical psychologists were regressed onto three models of the judgemental process which are described overleaf.

1 The familiar Linear Model.

$$\hat{Y}_s = \beta_1 X_1 + \beta_2 X_2 + \ldots. + \beta_{11} X_{11}$$

2 The Quadratic Model.

$$\hat{Y}_s = \sum_{i=1}^{11} \beta_i X_i + \sum_{i=1}^{11} \alpha_i X_i^2 + \sum_{\substack{i=1 \\ i \neq j}}^{11} \sum_{j=1}^{11} \gamma_{ij} X_i X_j$$

This model clearly contains both curvilinear and configural terms.

3 The Sign Model.

This model consisted of a linear combination of 70 clinical signs previously described by Goldberg (1965). A sign was defined as "any scale score or combination of scores however simple or complex, which can be specified precisely (i.e. non-judgementally). An operational definition of a sign in this sense is any index which can be programmed for a computer." (p.4.).

Although the variables in both the preceding models were signs in Goldberg's sense, the Sign model differs in important respects from these models by the rational nature of its signs. The two previous models are simply mathematical expressions of the first or second order, not rationally designed models of this particular situation. The signs used were compiled from the empirical clinical literature, from personal communication with M.M.P.I. experts, from clinical folklore etc. and the final model was a combination of both simple and configural terms.

An iterative multiple regression program was used to evaluate the models, in which zero weights were given to any terms which did not contribute significantly to the fit i.e. did not significantly change R_s^2. This led to a large number of terms of the Quadratic and Sign models being discarded.

The multiple correlation coefficients were reported for each judge, for each of the three models. The interpretation of these however is made very difficult by the exclusion of any information about the number of terms finally included in each of the three models for a particular subject, though even if this information had been given, it is doubtful if the degrees of freedom could be used in any meaningful way considering the way in which cues to be included in final models were selected. This process capitalises on chance variations in the data, and is presumably more open to error in cases with an initially large pool of possible cues.

The estimates of the parameters of the models, obtained from the first three samples, were used to predict the judgements of the subjects on the four later samples, resulting correlation coefficients between the predicted and actual responses were averaged over the four samples and used as indicators of the adequacy of each model. With regard to this index, Wiggins and Hoffman found that:-

(a) the Linear model was equal to or superior to the Quadratic model for 23 of the judges (though the highest difference in R_s

An iterative multiple regression program was used to evaluate the models, in which zero weights were given to any terms which did not contribute significantly to the fit i.e. did not significantly change R_s^2 . This led to a large number of terms of the Quadratic and Sign models being discarded.

The multiple correlation coefficients were reported for each judge, for each of the three models. The interpretation of these however is made very difficult by the exclusion of any information about the number of terms finally included in each of the three models for a particular subject, though even if this information had been given, it is doubtful if the degrees of freedom could be used in any meaningful way considering the way in which cues to be included in final models were selected. This process capitalises on chance variations in the data, and is presumably more open to error in cases with an initially large pool of possible cues.

The estimates of the parameters of the models, obtained from the first three samples, were used to predict the judgements of the subjects on the four later samples, resulting correlation coefficients between the predicted and actual responses were averaged over the four samples and used as indicators of the adequacy of each model. With regard to this index, Wiggins and Hoffman found that:-

(a) the Linear model was equal to or superior to the Quadratic model for 23 of the judges (though the highest difference in R_s

in favour of the quadratic model was only .03).

(b) the Linear model was equal to or superior to the Sign model for 17 judges (the highest difference in R_s in favour of the sign model being .04).

Their conclusion then that "The present research suggested that many clinicians utilize cues configurally in diagnosing M.M.P.I. profiles as psychotic or neurotic and from the magnitude of the cross validated multiple correlations between M.M.P.I. scales and judgements it was assumed that these judges performed at a high level of consistency" (p. 76), can only be accepted with the highest degree of caution. How likely is it that differences in R_s of .04 and .03 would reach statistical significance? And how can a quadratic model, which includes all the terms of the linear model, be meaningfully regarded as being a less good representation than the linear model? The methods of model evaluation and cue selection seem to require examination, before such conclusions may be drawn.

Despite the wealth of evidence showing that the judgements of human decision makers, in the above tasks may be quite adequately predicted by a quite simple linear combination of the cues, the leading researchers in this field refuse to accept the possibility that man might be nothing more than a "linear model". (c.f. Hoffman 1968; Anderson 1968, 1972). Anderson (1968) states:- "The model always fits the data quite well, but there are almost always small, significant discrepancies. Inspection of the data has failed to reveal the origin of the discrepancies; they may reflect some fundamental error in the model, or may result from remaining shortcomings in experimental technique" (p. 736). We might add to this, remaining shortcomings in methods of data analysis.

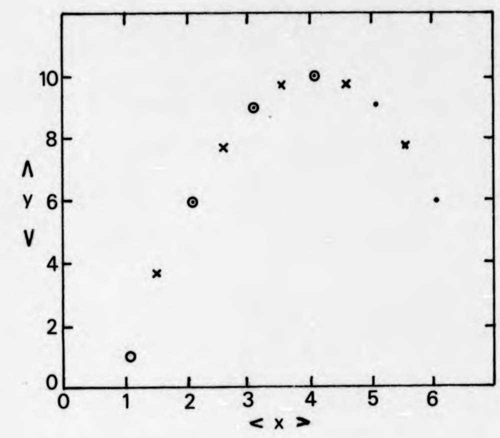
It is felt that the ubiquity of the linear model is due more to the statistical properties of the model itself rather than any relationship it might have to underlying psychological processes. It is likely that the power of the linear model to predict observations generated by a large class of non linear processes will tend to obscure all but the gross types of non-linear judgemental processes.

Yntema and Torgerson (1961) for example have shown that the analysis of variance is not necessarily a very good indicator of interactions. They constructed artificial data in which the dependent variable was related to the three independent variables in a purely interactive way:-

$$Y_{ijk} = ij + jk + ik$$

(where i, j and k were in integers from 1 to 7). Standard three way analysis of variance of the $7 \times 7 \times 7 = 343$ observations showed that the three main effects accounted for 94% of the variance leaving only 6% for the interactions.

Green (1968) also argues that in an important sense linearity is contributed by this sort of analysis rather than being an inherent property of the data.



The above "data" can be described exactly by the function $y = 10 - (x - 4)$. From one point of view this is purely a second degree system with no linear component. If the data points . are fitted using the standard method of orthogonal polynomials, all the variance will be described by the quadratic term, none by the linear. If, however, a curve is fitted to the x's a different pattern emerges: the linear component now accounts for 42% of the variance, the quadratic for 58%, the usual interpretation being that the data can be described by a weighted sum of a straight line and a parabola symmetric on the range of data points. It is obvious in the present case that such a combination results in a curve that is a segment of another parabola. What is often overlooked is that this is always the case. Any curve that is a weighted sum of a straight line and a parabola is a segment of another parabola. Are the data partly linear and partly quadratic, or completely quadratic? The answer is that they are both, these are two alternative descriptions of the same data. To take the process one step further, consider the o points. Here 92% of the variance is attributable to a linear component, only 8% to a quadratic component!

More recently Dawes and Corrigan (1974) have set out some of the conditions under which the linear model can be expected to fit data well, although those data may have come from a non-linear source.

Slovic (1966) used the method of multiple regression in a slightly different way from that discussed above. He set out to examine the influence of one aspect of the pattern of cues, namely "cue consistency", upon the manner in which these cues are used by judges.

Subjects were required to judge the intelligence of a person on the basis of quantified information from nine cues. Previous research had shown that subjects rely primarily on just two of the cues, I) High School Grade (HSR) and II) English Effectiveness (EE). On average, the linear use of these two cues accounted for about 80% of the total linearly predictable variance and about 60% of the variation in a judges predictions. Though some judges relied on HSR more than EE and vice versa, the judgements of more than 85% of subjects were correlated significantly and positively with both of the cues. This led Slovic to define consistency in this situation as a function of the absolute percentile difference between HSR and EE.

Each subject judged 75 profiles, after which certain subsets of the judgements were selected for separate analysis. Fifteen profiles, all exhibiting a percentile difference berween HSR and EE that exceeded 40 were singled out as inconsistent profiles and another 15 profiles with percentile differences between 10 and 20 inclusive were to represent a relatively consistent group. The two subsets were matched in terms of mean and standard deviation on all scales including HSR and EE, but the intercorrelation of HSR and EE in the consistent group was + . 80, whereas the corresponding correlation in the inconsistent group was - . 50.

Two separate multiple regression analyses were performed on the data from each judge; one for the set of consistent profiles and one for the inconsistent set. A modification of Hoffman's (1960) index of relative weight ($W_i = \beta_i r_i / R_s^2$) was calculated to assess the extent to which the judge was using the i th cue over the sets of consistent and inconsistent profiles (the index actually used was $R_s^2 W_i$ or $\beta_i r_i$).

The linear consistency, R_s of the subjects was quite high for both sets of profiles, though slightly higher for the consistent ones. Just what this implies is not clear (c.f. Schenck and Naylor (1968)). The sum of the relative weights for the seven additional cues was found to be significantly greater for the inconsistent set of profiles, and this was interpreted as showing that these cues were relied upon more when HSR and EE were inconsistent than when they agreed.

Consistency had little effect upon the average relative weight given to the more important of the two critical cues, but the mean relative weight for the lesser of the two critical cues was lower for inconsistent profiles than for consistent ones.

Slovic concluded, that there is a tendency for subjects to rely on both HSR and EE when the cues agreed with one another, but when HSR and EE were contradictory, subjects tended to use only one of these cues, the lesser one being excluded from consideration, they also tended to rely more upon the other seven cues.

This study is interesting, since, by an unconventional use of the multiple regression technique, it appears to have shown quite definite configural use of cues. It is unfortunate that Slovic did not attempt to carry out a multiple regression analysis on the whole of a subject's responses at once, in order to discover whether there was a significant increase in fit on the inclusion of a configural term in HSR and EE. Had no such significant improvement been found this would have added more weight to the arguments suggesting that linear statistical techniques tend to obscure non linearity in judgemental processes.

In all the studies so far reviewed, the subjects have been required to make their responses on a certainty scale and these responses were then treated as interval data in the subsequent analyses. In a number of tasks this introduction of a scale of certainty may be new to the judge, who in his day to day life might only make nominal classification e.g.

neurotic - normal, benign - malignant. Rodwan and Hake (1964) report a piece of research which required subjects to make only nominal responses, in an attempt to compare the judgements of human subjects, with those made by another linear statistical technique, Fisher's "Linear Discriminant Function" (LDF), (Fisher 1936).

They felt that a subject's judgements of certain events or stimuli might be described adequately if it was assumed that the judgement was a result of the subject having combined the values of the stimulus on what were felt to be the relevant dimensions in a linear manner. Any decision about or classification of the stimuli would be made on the basis of this linear combination.

Linear discriminant function analysis assumes that the decision to be made is between two multivariate normal distributions with the same covariance matrix. It yields weights for the variables which will best discriminate between the two populations and requires as data, not a continuous dependent variable (like multiple regression) but a simple dichotomous variable - population I or population II. (See the Statistical Appendix).

An experiment was performed, following the design of Brunswick and Reiter (1938) in which subjects were required to classify a number of schematic faces as "intelligent" or "not intelligent". The faces varied along four dimensions X_1 length of nose; X_2 length of chin; X_3 distance between eyes; X_4 length of forehead, each of which had four possible values. The same 81 faces were classified by a subject under a number of conditions varying in the a priori probability of an intelligent face. LDF was applied to the decisions of each subject singularly, and it was found that, for each subject one set of weights could adequately describe his behaviour in all conditions i.e. the subject could be represented as using the same linear combination of variables in all conditions, though the optimal cutoff value on this linear combination varied between conditions. Unfortunately, no test of goodness of fit is reported, nor does the significance of the derived weights appear to have been investigated. These weights are presumably even more subject to chance variations than those obtained through multiple regression, since the scale of the responses is so much weaker, it may well be that some dimensions received weights which did not differ significantly from zero.

Rodwan and Hake appear to have disregarded one assumption of the LDF, which is that the covariance matrices of the two populations are assumed to be equal (i.e. the populations have the same shape). This assumption would imply that the subjects' responses "intelligent" and "not intelligent" could be regarded as being samples from two multivariate populations of differing means but equal variances and covariances of the variables. This is extemely unlikely, particularly when it is considered that the subject is expected to have carried out a linear discriminant function analysis himself thus truncating the "intelligent" population to those observations falling above the cutoff and the "not intelligent" population to those falling below.

These considerations made the interpretation of Rodwan and Hakes findings very difficult. It would appear though that even if the judges do act as LDFs, then the application of the statistical method of linear discriminant function analysis using their responses as data, is not an appropriate test of the model.

Rodwan and Hake draw attention to the formal similarity to the LDF of the Theory of Signal Detectability (TSD), which was developed in the fields of audition and vision and later as a general model of psychophysics (Tanner and Swets 1954; Swets, Tanner and Birdsall 1961). Both models assume normal distributions of variables for the events to be distinguished. However, they claim that TSD does not enable the experimenter to determine the number of and the weights for, the attributes which the judge uses in making decisions.

It is interesting to note, in this connection, that a formulation by Green (1964) of a model to describe subjects' performance in a multiple component signal detection task, bears a most striking resemblance to the LDF although it was developed out of TSD. The model, which appeared to adequately describe data collected in tasks requiring the detection of a two tone signal, could be represented as a linear combination of the strengths of the two components of the signal and a cutoff somewhere along the resulting axis (the exact position of which depended on payoffs, prior probabilities etc.).

All the models of judgemental processes so far discussed, have been "linear models" in the statistical sense. Einhorn (1970) examined the ability of two statistically non linear models to reproduce the rankings by judges of the suitability of applicants to graduate school. Each judge was given the percentile scores of each of 20 hypothetical applicants on three predictor variables, only two of which were considered in the analysis.

The first, "disjunctive" model was designed to give a high value to the applicants who had a high score on any one or more of the indicators and was formulated in the two cue case as:-

$$\hat{Y}_s = \left(\frac{1}{a - X_1}\right)^{\beta_1}\left(\frac{1}{a - X_2}\right)^{\beta_2}$$

The β_is being weightings of the two indicators (estimated by a log transformation of the equation) and the "a" being an arbitrary parameter set to a value greater than the maximum that either X_i could obtain.

The second "conjunctive" model allotted high value only to those applicants who scored highly on all indicators and was formulated as:-

$$\hat{Y}_s = (X_1)^{\beta_1}(X_2)^{\beta_2}$$

The β_is again being weighting parameters estimated through log transformations.

Einhorn reported that one of his subjects appeared to be using the disjunctive rule, and one the conjunctive rule, when these models were tested against the simple linear model. For the third subject however, the linear and conjunctive formulations gave very similar fits and no meaningful decision could be made between them. Einhorn did not give details of his method of analysis, but it must be assumed that since his subjects were asked only to rank order the applicants, that only rank order properties of the responses were used.

Goldberg (1971) reanalysed the data of Meehl and Dahlstrom (Meehl 1959) previously analysed by Wiggins and Hoffman (1968). The task facing the 29 clinical psychologists being to rate M.M.P.I. profiles from a number of samples on a 11 - step forced normal distribution ranging from "certainly neurotic" to "certainly psychotic".

Goldberg tested the goodness of fit of Einhorn's formulations of the conjunctive and disjunctive models for these data and compared them with the simple linear model and two other models.

(1) The logarithmic model

$$\hat{Y}_s = \sum_{i=1}^{11} \beta_i \log X_i$$

(2) The exponential model

$$\hat{Y}_s = \prod_{i=1}^{11} e^{\beta_i X_i}$$

which were intended as "controls" for the conjunctive model, to check whether any possible incremental validity of the conjunctive over the linear model might simply stem from the logarithmic transformation of the cues alone (logarithmic) or of the judgements alone (exponential).

Since it was felt that the responses here could be regarded as interval data, the models were suitably transformed in order to yield (statistically) linear models and were then analysed by standard multiple regression techniques.

The linear model was found to provide a better representation of the diagnostic judgements of these psychologists than did either the conjunctive or disjunctive models, and of the five models utilized in the study only the logarithmic model provided the linear model with any real competition. It was also found that the constant "a" in the disjunctive model was not as arbitrary as Einhorn had felt, and that the accuracy of the disjunctive model did vary as a function of the value of this parameter.

Despite these findings, it is felt that the disjunctive and conjunctive models should not be disregarded as possible representations of human judgemental processes, possessing as they do, more psychological "face" validity than the linear model. However, more investigation is required into the requirements of various statistical techniques with regard to strength of date (e.g. ordinal versus interval) and of the effects of various transformations on the validity of techniques.

LEARNING TASKS

All of the studies, so far reviewed, have attempted to investigate judgemental processes which were assumed to exist before the commencement of experimentation, i.e. it was assumed that the subjects knew the distinctions to be drawn (the response scale), and that they possessed some process, which allowed them to make this distinction on the basis of the cues provided. Wiggins and Hoffman (1968) assumed that their clinicians knew, not only what "neurotic and psychotic" implied, but also what "certainly neurotic", "neutral" and "certainly psychotic" meant. Slovic (1966) assumed that his subjects knew what the "intelligence" scale on which they were required to respond meant. Hoffman, Slovic and Rorer (1968) assumed their physicians knew what was meant by "definitely benign", "uncertain", "definitely malignant" etc.

These situations were designed to be representative of the "real life" situations facing decision makers, but to what extent do they fulfil this aim? Such response scales as the ones above are probably never used by the decision maker in his day to day work. It is not likely that psychologists ever make responses on a scale ranging from "certainly neurotic" to "certainly psychotic", or that they ever judge large numbers of profiles in a short time, or even that they are required to make a dichotomous response on the basis of M.M.P.I. profiles of known psychiatric patients. Similar criticisms apply to the diagnosis of malignant versus benign ulcers, though in this case, the dichotomous classification is familiar to the physician, but on the basis of X-ray plates, not data summarised from such plates. Finally, how often are college administrators required to judge intelligence from various indices of past performance, and how much opportunity would they have to validate their estimates?

The point being made is that these tasks may not in fact be as familiar to the judges, as was intended and that the sorts of strategies generated to deal with such situations may not be representative of those actually applied in real life. We are particularly likely to get simple strategies developing in such situations as no feedback or knowledge of results is provided against which the validity of judgements can be estimated. The judges are being asked to make judgements, which are possibly novel to them, in situations which only bear some resemblance to those with which they are familiar, and are given no chance to test their responses. Is it surprising then that there is little evidence of a complex judgemental process in such situations?

The researchers in the next area to be reviewed made their tasks one degree more abstract than the ones described above, in order to overcome the above criticisms (but also it must be added, to make it possible to use some newly developed techniques which will probably never be usable in a "real life" situation).

The response and cue scales were not expected to have a case, the dichotomous classification is familiar to the physician, but on the basis of X-ray plates, not data summarised from such plates. Finally, how often are college administrators required to judge intelligence from various indices of past performance, and how much opportunity would they have to validate their estimates?

The point being made is that these tasks may not in fact be as familiar to the judges, as was intended and that the sorts of strategies generated to deal with such situations may not be representative of those actually applied in real life. We are particularly likely to get simple strategies developing in such situations as no feedback or knowledge of results is provided against which the validity of judgements can be estimated. The judges are being asked to make judgements, which are possibly novel to them, in situations which only bear some resemblance to those with which they are familiar, and are given no chance to test their responses. Is it surprising then that there is little evidence of a complex judgemental process in such situations?

The researchers in the next area to be reviewed made their tasks one degree more abstract than the ones described above, in order to overcome the above criticisms (but also it must be added, to make it possible to use some newly developed techniques which will probably never be usable in a "real life" situation).

The response and cue scales were not expected to have a relationship within the "mind" of the subject before the commencement of experimentation. The researchers attempted to teach the relationship to the subject by providing feedback or knowledge of results. The response scales were not, in general, scales of likelihood between two populations (e.g. neurotic - psychotic, benign - malignant) or even known traits (e.g. intelligence) but tended to be unspecified numerical or spatial scales representing some abstract variable Y, the value of which, the subject was to predict as accurately as possible.

A distinction may be made between those tasks in which the criterion is perfectly predictable from the predictor cues, and those in which the relationship is only probabilistic. This distinction is only important in the one predictor cue case where no interchange of ideas has taken place between researchers in the two areas. In multiple cue research, however, some experiments deal with both perfect and imperfect relationships so the distinction loses its importance.

Research into the ability of subjects to predict the value of one cue from another to which it is perfectly linearly related has been carried out mainly by Scandinavians, and has been referred to as "learning linear functions" (Carroll 1963) i.e. learning functions of the form.

$$Y = a + bX$$

Subjects are usually presented with paired samples from the X and Y variables during training trials and then on test trials only the value of X is presented and the subjects are requested to predict Y. (c.f. Eisler and Spolander (1970), Bjorkman (1965)). X and Y are, in effect, perfectly correlated (r = 1.00, or - 1.00 if "b" is negative).

Practically all the research carried out in this field has expressed the X and Y variables as distances marked off on lines, and this does of course place a number of restrictions on the generality of the conclusions drawn.

A number of researchers (Bjorkman, 1965, Brehmer 1971, Carroll 1963, Eisler and Spolander 1970) have investigated the effect of the sign of the slope parameter "b" upon the accuracy of subjects' predictions. The most general finding which emerged was that linear functions with negative slope parameters, were consistently less well learnt than those with positive slope parameters. Eisler and Spolander (1970) felt that the X - Y differences, which were confounded with sign of slope, may have given rise to the poorer learning in the negative slope tasks, rather than any intrinsic difficulty of negative slope functions.

De Klerk, Oppe and Truijens (1972), using an anticipation paradigm in which subjects predicted Y from X on training trials, but were given immediate knowledge of the correct value of Y, were able to find no effect of the "a" parameter, supporting a previous finding of Eisler and Spolander (1970). The "b" parameter had a significant effect, with performance when $b = .85$ significantly better than when $b = 1.25$ (only positive values of b were investigated). They suggested that more research should be carried out, to investigate the extent to which the effect of "b" might be reduced by providing subjects with visual aids or by using other types of stimulus material, such as numbers instead of lines.

Unlike the research into the learning of linear functions, that into the learning of imperfect linear relationships and estimating such relationships, has practically always used numerical materials.

The whole area of man's ability to deal with correlated variables will be seen to be of relevance to the present research, and for this reason, interest will not be restricted solely to those experiments in which subjects are required to predict one cue Y from another X when the two are imperfectly related.

i.e. $$Y = a \pm b X + \epsilon$$

where ϵ is unpredictable error.

Smedslund (1963) and Jenkins and Ward (1965) found that subjects performed rather poorly when asked to estimate the degree of relationship in a 2 x 2 contingency table. It appeared that they paid almost sole attention to confirming positive instances (+ +), disregarding information from the disconfirming (- +) and (+ -) and the confirming negative (- -) cells.

		Variable I -	Variable I +
Variable II	+	+ -	+ +
	-	- -	- +

Smedslund called for research into the performance of subjects in a situation using continuous variables and it is to the findings of those researchers who took up the plea that we now turn.

In an experiment in which subjects were required to give estimates of their certainty that a particular sample of sequentially presented X - Y pairs (two integers, one black, one green, between 1 and 10 inclusive) was drawn from a population with either a positive or negative correlation. Beach and Scopp (1966) found that subjects "Clearly ...... did better as the magnitude of the sample correlation increased, although there was a tendency to call all low correlations negative and thus there was a slightly higher chance of being correct when a low negative correlation occurred." (p. 34).

In a somewhat similar experiment Erlick and Mills (1967) required subjects to observe sequences of the 20 (X - Y) pairs sampled from populations of varying correlations. X and Y took integer values 1 to 5 inclusive and were presented by pointers on two circular dials. Subjects then estimated the relatedness of the two variables and it was found that these estimates were closely related to the sum of the discrepancies between X and Y over the sequence. Presumably this implies some process such as $\sum|X-Y|$ and this is just the cue that Eisler and Spolander (1970) were to feel caused the difference between subjects' accuracy in learning linear functions of positive and negative slopes. Erlick and Mills (1967) reported a significantly greater error rate for negative correlations than for positive ones and that one group of subjects had a marked tendency to make positive responses (unlike those of Beach and Scopp (1966)).

Two papers dealt with the ability of subjects to predict one of the variables from the other. Gray, Barnes and Wilkinson (1965) followed the usual linear function paradigm. A number of X - Y pairs (both X and Y took integer values 1 to 9) were presented numerically, physically (as lengths of lines) or both. On test trials only X was presented and subjects were required to predict Y. The samples of X - Y pairs were of varying degrees of correlation (.96, .75, and .44) but all were positive. It was found that the subjects' predictions appeared much more to match the environmental situation than to optimize the accuracy of prediction, i.e. the subjects' predictions Y_s tended to be correlated with X to a similar degree as the actual Y s, whereas an optimizing strategy, the one which linear regression analysis prescribes, would imply a perfect correlation between Y_s and X. At low levels of correlations, subjects' responses tended to show more dependence on X than actually existed between X and Y, this finding was also reported in the next paper to be reviewed.

Naylor and Clark (1968) followed an "anticipation" paradigm; on all trials the value of the X variable was displayed and the subject was required to predict Y, after which the correct value of Y was displayed. Both X and Y were presented as two digit integers and both variables were normally distributed with mean 50 and standard deviation 10. Groups of subjects were run for 200 trials under one of nine experimental conditions (representing correlations of .80, .60, .40, .20, .00, - .20, - .40, - .60, - .80).

Performance of subjects in the positive correlation conditions was consistently superior to that of those in the negative correlation conditions (the measure of accuracy used was the correlation of subjects' predictions with the actual criterion values). The accuracy of prediction was also greater for higher degrees of correlation than for lesser degrees.

The correlations of subjects' predictions with X were of similar magniture to the actual correlations between X and Y for the high positively correlated conditions, but at lower positive correlations subjects' estimates were much more dependent on X than was Y. (Indeed, in the .00 correlation condition the correlation between subjects' predictions and X was still as high as .40). With negative correlation conditions, however, dependence of the subjects predictions on X was consistently less than between X and Y. Naylor and Clark proposed that this result was due to subjects approaching such prediction tasks with a general set to regard all relationships as positive and that this attenuates their achievements with negatively correlated cues.

In all the research on correlations which has been reviewed, the constant "a" in the equation

$$Y = a + b X + \epsilon$$

has been set equal to zero. Not only this, but "b" has only taken the values + 1 and - 1. It is of course an open question what effect other values of these parameters may have, but from the findings in linear function research, we might expect "a" to have little or no effect, and the effect of varying values of "b" to be small when compared to the effect of sign "b".

The lesser ability of subjects to deal with negatively related cues, pervades these areas of research. But it is felt, that this inability may be due more to the "scaling" of the two cues, than to anything inherently difficult in negative relationships. The point may be illustrated by the following sequences of pairs:-

(a) (1,1), (2,2), (3,3), (4,4)

(b) (1,4), (2,3), (3,2), (4,1)

(c) (1,-1), (2,-2), (3,-3), (4,-4)

(d) (1,-4), (2,-3), (3,-2), (4,-1)

Subjects have been shown to find situations of the (b) type more difficult than those of the (a) type (the sign of "b" is positive in (a) but negative in (b). However, is there any reason to believe that subjects will not behave as well in situations of type (c) as in (a) ((c) has a negative slope) or that they will behave better in situations of type (d) than of type (b), ((d) has positive slope, (b) negative).

Situations of the (b) and (d) kind are not easily expressed when the cues take the form of distances marked on lines, but there is no reason why numerical tasks should not be constructed in this way. It is felt that if such tasks were investigated, the conclusion about negative relationships would have to be reformed and replaced by one which restricted the difficulty of negative relationships to situations in which both cues are either positive or negative.

We now turn to look at tasks which require subjects to learn to predict the criterion Y from more than one predictor variable X. As stated earlier, there is no reason to discuss the cases in which Y is perfectly predictable from the X's, separately from those in which Y and the X's are only probabilistically related, as such a division does not occur in the literature as it does in the one cue case.

The earliest research of this character, by Smedslund (1955), was strongly influenced by the description of Brunswick and Herma (1951) of the way in which subjects combine cues in perceptual learning. Two (or three) pointers, whose directions changed from trial to trial, were presented to the subjects. The experimenter defined a variable Y by expressing the values of the pointers numerically, averaging them and adding a small randomly determined component in order to make the relationship probabilistic.

i.e. $Y = \frac{1}{3} X_1 + \frac{1}{3} X_2 + \frac{1}{3} X_3 + \epsilon$

The subject reported his estimates of Y on each trial by sliding a knob along an unmarked scale and feedback was provided by the experimenter moving the knob to the setting representing the predetermined value of Y. Smedslund was able to conclude that subjects can learn to utilize many probabilistic cues simultaneously in making such predictions.

Summers (1962) felt that Smedslund's design was not completely satisfactory as a particular cue was always associated with the same validity so that the effect of validity could not be studied independently of the effects of saliency or other cue characteristics. He designed an experiment in which correlations were imposed between each of three simultaneously presented cues (the orientation, colour, and area shaded of isoceles triangles, each one taking one of eight discrete values) and a predicted variable, line length, whose magnitude varied with the magnitude of all three cues.

The correct line length was determined by the equation

$$Y = 2.0X_1 + 1.5X_2 \text{ and } 1.0X_3$$

where the X_is are the three cues. Different subjects were then under six conditions - in the first X_1 was colour, X_2 was orientation, X_3 was area shaded - in the second X_2 became area shaded, X_3 orientation and X_1 was colour - and so on until every cue had taken all three possible weightings (2, 1.5 and 1) in combination with all possible weightings of the other cues.

Summers expected that the order of the response weightings, (i.e. those weights obtained on regressing the subjects predictions Y_s on to the X_is) would come to conform to the order of the actual cue weightings, and that the magnitudes of the response weightings would approach those of the cue weightings. The results, in general supported these expectations, the subjects responded simultaneously and differentially to the multiple cues, and cue utilization was found to be roughly proportional to cue validity throughout the learning trials.

The effect of distribution of cue weightings on learning, was studied in more detail by Uhl (1963). He designed a task in which three interval scale cues (expressed as three rows of nine lights) were perfectly linearly related to a criterion scale (a fourth row of lights). The values taken by the three cue dimensions were determined by a table of random normal numbers, with standard deviation equal to 2, mean 5, and intercorrelations of the cues approximately zero.

Different subjects were run under seven experimental conditions which differed in the degree to which the criterion was related to each of the three cues, but in all conditions the criterion was perfectly predicted by some linear combination of the cues. The distribution of cue weights varied from, approximately equal weightings of all cues in one condition to an almost complete reliance on only one of the cues in another.

It was found, that subjects had most difficulty in predicting accurately from multiple stimuli, differing only moderately in cue weightings, they had less difficulty with cues of equal weighting, but performed most accurately with highly disparately weighted cues. This last finding is related to another conclusion of the study, that the smaller the number of relevant cues, the more accurate were subjects' predictions.

Uhl included an eighth condition where the correct response was not related to any of the cues. Here, he found that subjects responded to the stimuli in a relatively non-random way. For the last block of trials, the mean multiple correlation between the cues and the subjects' predictions (R_s) was + .40. This is of very much the same order as was to be reported by Naylor and Clark (1968) with one predictor cue, when R_e equalled zero.

The effect of varying degrees of predictability (of the criterion) was investigated by Dudycha and Naylor (1966). Only two predictor cues were used, two normally distributed variables of mean 50 and standard deviation 10, whose intercorrelation was zero. The cue weightings of these two cues were varied over conditions yielding R_e s ranging from .998 - .457, representing combinations of cues with validities of (.80, .60) and (.40, .20) respectively.

They found that the more predictable was the criterion (i.e. the greater R_e), the greater was the average achievement (r_a) and the greater was the linear dependence of the subjects' responses (R_s). Taking into consideration, the results of their previously mentioned single cue experiment, they were able to conclude that when the validity of one cue was large (e.g. .8) the addition of a second cue of validity less than .6 decreased performance. When the first cue had only a validity of .4 however then an additional cue of validity greater than .2 increased performance. A cue of validity .2 was always detrimental to performance, presumably because it drew attention from more relevant cues.

Having calculated G, the correlations between the best linear combinations of cues for the prediction of the criterion and the best linear combination for predicting the subjects' responses, they comment, that the high values of G "which were consistently obtained regardless of the particular experimental conditions involved are in many respects truly remarkable. This indicates that humans tend to generate "correct" strategies, but then, in turn fail to use their own strategy with any great consistency .... One is left with the conclusion that humans may be used to generate inference strategies, but that once the strategy is obtained the human should be removed from the system and replaced by his strategy!" (p. 127). In the days of "boot strapping" (Dawes and Corrigan, 1974) this suggestion is no longer so surprising as it was in 1966.

Azuma and Cronbach (1966) designed a task in which Y was perfectly predictable by a linear combination of two of four presented variables.

$$Y = \frac{2}{3} X_1 + \frac{1}{3} X_2$$

Cues X_3 and X_4 were irrelevant. The two relevant cues were the horizontal positions of a circle and a cross on a card, the vertical positions being the two irrelevant dimensions (each cue took one of four possible values). Rather than actually predict Y, subjects were required to state which of four standard cards had a Y value most similar to that of the card on each trial. On training trials, immediate feedback was given, none however was given on test trials to discourage subjects from trying new rules during testing, thus giving a "purer" measure of performance.

Considerable individual differences in ability to solve this problem were found to exist amongst the subjects. Seven of the sixteen subjects eventually learned to perform almost perfectly, whereas the others learned to degrees varying from well to not at all. Azuma and Cronbach felt that if a subject did not learn a correct rule at an early stage in the experiment, then he probably learnt a false one which he found hard to discard. Performance on training trials was significantly worse than that on test trials, justifying the distinction made between the two.

On asking subjects for verbalizations of their rules, the authors were surprised to find that no subject had in fact discovered the correct rule. What they had done, was to form simple rules for particular situations i.e. for particular cue combinations, and the rules were not formulated in any mathematical way. Such a finding brings home in full Hoffman's (1960) distinction between "isomorphic" and "paramorphic" representations, but the generation of such ad hoc rules is probably less likely in tasks of a less discrete nature.

Practically all the experiments investigating the ability of subjects to learn the relationships between predictor cues and a criterion Y have used predictor cues which were generated so as to be independent (uncorrelated or orthogonal). In such situations, the square of the multiple correlations coefficient between the predictors and the predicted variable Y (i.e. R_e^2) is equal to the sum of the squares of the individual correlations between each predictor and Y (if the cues receive equal weight). If, however the predictors are intercorrelated, this relationship breaks down and R_e^2 will not be equal to $\sum r_{ei}^2$. With reference to the use of orthogonal predictors, Naylor and Schenck (1968) felt that "While such "nicety" of experimental design has merit for examining certain experimental questions, it is nevertheless inconsistent for research based on Brunswick's model of probabilistic functionalism. One of the most important concepts of Brunswick's position is the notion of "representative design". That is, our experimental situations should involve sampling of those environments to which we might wish to generalise." (p. 48 - 49). Since it is extremely unlikely that there is no redundancy in the cues in any real life situation, Naylor and Schenck set out to study the effect of systematically manipulating cue redundancy upon the behaviour of subjects in order to discover if previous findings with orthogonal cues would hold up.

Two normally distributed predictor cues (means equal to 50, standard deviations equal to 10) were presented as two-digit numbers related probabilistically (R_e took values .5, .7 and .9) to the criterion of Y, also a normally distributed variable (mean 50, standard deviation 10) expressed as a two-digit number. Three levels of predictor cue intercorrelation were investigated (r = .00, .40, and .80) at each level of task predictability, R_e.

Each subject received 200 trials in eight blocks under one of the nine resulting conditions, using the anticipation paradigm, no distinction being made between training and test trials. Subjects behaviour was considered in terms of the three indices r_a, R_s and G.

It was found that performance, in terms of r_a, increased with both degree of task determinancy R_e and degree of predictor intercorrelations. The linear dependence of subjects responses R_s followed both these trends, but the matching index G increased significantly only with cue intercorrelations. They concluded that the most striking feature of the data was the substantial moderating influence of cue intercorrelation upon the various performance indices, and they felt that this result implied that multiple cue data based on orthogonal cues may indeed not generalize so readily to more representative learning situations in which certain degrees of cue redundancy exist. Cue redundancy appeared not only to have a direct influence upon performance but also appeared to interact with task predictability having less effect with tasks of lower predictability.

In a later paper, (Schenck and Naylor, 1968) it was suggested that the effect of cue intercorrelation on the linear consistency R_s index may have been more apparent than real. They found that if one assumes a linear model of some specified degree of fit for a subject, then any increment in cue intercorrelation will necessarily yield an increase in the goodness of fit of that model. In fact, this is an oversimplification of the situation on their part, and holds only if both cues between which the increment in correlation takes place are positively or negatively weighted. If one cue receives a negative and the other a positive weight, then an increase in the correlation between them will actually lead to a reduction in R_s. However, the point is taken that intercorrelations between predictors may lead to difficulties in the interpretation of certain behavioural indices, which do not exist with orthogonal predictors, but this should surely be expected in the light of difficulties occurring in say, the analysis of variance with correlated variables due to different numbers of subjects in each cell.

In view of the large amounts of energy which have been expended in searching for evidence of non-linear cue utilization in "real life" judgemental situations, it comes as some surprise to discover the scarcity of research on non-linear cue utilization in learning tasks. As far as we know, only one such paper exists, that of Hammond and Summers (1965). They felt that it was not very useful trying to discover non-linear inference processes in tasks which had not been shown to be non-linear in nature. In view of Brunswick's (1956) admonition that tasks should be representative of a wide range of conditions, the performance of subjects in situations involving non-linear, as well as linear relations should be investigated before concluding that the process of inductive inference is primarily linear. In short, having discovered what apparently is a strong tendency for human subjects to utilize the data from linear relations in a highly linear manner, it remains to investigate whether subjects utilize the data from non-linear relations in a linear or non-linear manner.

An experiment was performed in which two predictor cues X_1 and X_2 were related to the criterion Y. X_1 was linearly related to Y but X_2 was related to Y as a sine function.

i.e. $Y = X_1 + \text{sine } X_2$

Thus Y was completely determined by X_1 and X_2 but a multiple regression of Y on X_1 and X_2 would not yield $R_e = 1.00$. (The regression of Y on X_1 and a new variable Z, defined by $Z = \text{sine } X_2$, would yield a multiple correlation $R_e = 1.00$).

Subjects were tested under one of three experimental conditions, which differed in the amount of prior information about the combination rule, which was given to the subjects:-

(1) Subjects were just instructed to predict Y from X_1 and X_2

(2) Subjects were told the "theoretical" structure (linear and non-linear) of the task.

(3) Subjects were told how the "theoretical" structure applied to the specific task.

The predictor cues were presented as scores on two ten point scales which were printed on cards. The criterion was printed on the back of the cards. One hundred trials were split into five blocks and each subject on each trial viewed the two predictor scores, made his prediction of Y then turned over the card and noted the true criterion value.

Achievement in this task (in terms of r_a) was quite high but significant differences existed between the three levels of task information. The relative contributions of the linear and non-linear components of performance were about equal, except in the minimal information condition, where the main contribution came from the linear component.

The mean linearity R_s was less than optimal and quite low (R_s^2 ranged from .21 - .42 compared with the optimal value of .50). This finding is important because, although the results of previous studies indicate a high degree of linearity on the part of subjects in multiple cue probability learning tasks, these results indicate that the propensity for a highly linear, additive response system, is contingent upon the subject being presented with a highly linear task.

In all the experiments so far reviewed on the behaviour of subjects in single or multiple cue learning tasks, the samples of observations shown to the subjects, are best regarded as being drawn from one multivariate population which

(2) Subjects were told the "theoretical" structure (linear and non-linear) of the task.

(3) Subjects were told how the "theoretical" structure applied to the specific task.

The predictor cues were presented as scores on two ten point scales which were printed on cards. The criterion was printed on the back of the cards. One hundred trials were split into five blocks and each subject on each trial viewed the two predictor scores, made his prediction of Y then turned over the card and noted the true criterion value.

Achievement in this task (in terms of r_a) was quite high but significant differences existed between the three levels of task information. The relative contributions of the linear and non-linear components of performance were about equal, except in the minimal information condition, where the main contribution came from the linear component.

The mean linearity R_s was less than optimal and quite low (R_s^2 ranged from .21 - .42 compared with the optimal value of .50). This finding is important because, although the results of previous studies indicate a high degree of linearity on the part of subjects in multiple cue probability learning tasks, these results indicate that the propensity for a highly linear, additive response system, is contingent upon the subject being presented with a highly linear task.

In all the experiments so far reviewed on the behaviour of subjects in single or multiple cue learning tasks, the samples of observations shown to the subjects, are best regarded as being drawn from one multivariate population which may or may not be normal. These tasks, clearly bear a closer relationship to such "real life" tasks as that of Slovic (1966) in which subjects were required to make judgements of intelligence, on the basis of a certain set of cues, than they do, for instance, to the diagnosis of malignant versus benign ulcers (Hoffman, Slovic and Rorer, 1968) or of neurotic versus psychotic psychiatric patients (Wiggins and Hoffman, 1966). The latter two tasks are certainly better regarded as decisions between two populations rather than the partitioning of one population. A difference between these "real life" tasks also occurs in the type of feedback which could be given to the judge in a real life situation. Intelligence can be measured (though with some inaccuracy) on a scale which can be regarded as an interval scale - thus, data of an interval nature could be given as feedback if such a task ever occurred in real life. However with differential diagnosis, the feedback received would normally be of a much weaker, nominal kind i.e. malignant or benign, neurotic or psychotic, and, not until a great deal more is known about the ecological side of Brunswick's lens will interval data of the form X% certainty of malignant ulcer, Y% certainty of neurotic, be available. This later point is one of the main drawbacks of Hammond et al's (1964) formulation of Brunswick's model. This formulation assumes not only an interval scale for the criterion Y but an interval scale which can be measured objectively. The model therefore is only of use with such (unlikely) tasks as the prediction of intelligence from M.M.P.I. profiles and it is unlikely that it will ever be of use in such situations as the differential diagnosis of malignant versus benign ulcers, or neurotic versus normal personality profiles. It might be argued that much of the research into multiple cue learning tasks has taken the form which it has, not because of any desire for representative design, but more in an effort to collect data which will fit into models such as Hammond et al's, without any difficulty. It is felt then that such research cannot always be justified on the grounds that it is an attempt to simulate but control real life decision making tasks, and it does not follow that strategies used by judges in such situations will be in any sense representative of "real life" decision making processes.

We now turn to the two papers which have attempted to simulate situations in which the basic task is to discriminate between two normally distributed populations. The feedback given is nominal i.e. population I or population II but, because of the known properties of the populations it is possible to compare the judges estimates of "certainty" or "probability" with those logically derived, using the parameters of the populations:- this last point is a real advantage over real life tasks and is a benefit gained by taking the decision making task into the laboratory, where a greater degree of control of samples, judges' experience etc. can be applied.

Lichtenstein and Feeney (1968) constructed a task in which subjects were to learn to classify observations as belonging to one of two bivariate normal populations. Subjects were shown cards, representing maps of the positions of two cities A and B and also an asterisk marking the point at which a bomb, dropped from an enemy aircraft, had exploded. Subjects were required to identify the city at which the bomb had been aimed. They were told that, because of bombing errors the bomb rarely landed directly on the target city, but the errors were unbiased in that the bomb might just as likely miss it's target in any direction, and that it was more likely that the bomb would fall near it's target than fall far from it. One hundred and fifty samples were drawn from each of two bivariate normal distributions, one with mean at the position of city A and the other with mean at the position of city B. The dispersion matrices of the two distributions were identical, the two variables, X - horizontal position and Y - vertical position, had the same variance and were uncorrelated. The cities were 1.25 standard deviations apart and both lay on the same vertical axis (i.e. the vertical position of the bomb blast (y) had no predictive validity).

The resulting 300 bomb positions were presented to the subjects in random order, and on all trials subjects were required to state at which city they thought the bomb was aimed. Immediate feedback (the correct city) was given for the first 150 trials, and, for the last 250 trials, subjects were required to assign percent certainties to the two targets i.e. divide 100 points between them in such a way as to express the subject's degree of confidence that each city in turn was the correct target. For the last 150 trials no feedback at all was given, and it was the data from these trials which were analysed.

The best decision function in this situation is that given by linear discriminant function analysis and reduces to drawing a line half way between the cities and at right angles to the line joining the city centres. This is the line of equal probability, where a bomb is as likely to have been aimed at city A as at city B. Subjects seemed to appreciate this fact and soon learned to assign all bombs to the left of this line to city A and all those to the right to city B. What they were not so good at was the assignment of probabilities to the two alternatives.

For a bomb landing at a point with co-ordinates (x,y) the probability of this bomb being aimed at city A can be shown to be

$$P_A = \frac{1}{1 + \exp\left[\frac{(x_B - x_A)(x - x_M)}{\sigma^2}\right]}$$

where σ^2 is the variance of both distributions, x_A and x_B are the x co-ordinates of the two cities and x_M is the x co-ordinate of a point half way between the two cities. The probability is independent of y and the farther away along the x axis the bomb falls from the two cities, the more extreme the posterior probabilities become.

When Lichtenstein and Feeney came to correlate the estimates of probabilities of their 11 subjects with those of the aforementioned normative model, they found that the estimates of only three of their subjects, were even moderately related to those of the model. From an investigation of the subjects' responses however, as well as comments made by subjects during and after the experiment, it was found that the subjects were being careful and consistent in their choice of probabilities, but they were not following the normative model. Many subjects gave probabilities close to .5 when the bomb fell to the extreme left or extreme right of the field and gave probabilities of .9 or above when the bomb fell very close to one city. Some subjects reported that they compared the distances of the bomb site from each of the cities and based their probability estimates on this comparison.

On the basis of this evidence, Lichtenstein and Feeney developed an alternative model with which to compare subjects' responses. The posterior probability of A, P_A, was assumed to be a function of the ratio of the two distances of the bomb site from the two cities.

i.e. $$P_A = \frac{D_A}{D_A + D_B}$$

where D_A is the distance of the site from city A and D_B the distance from city B.

The loci of the points of constant likelihood for this model are circles (except when $P_A = .5$ when they form the same straight line as prescribed by LDF analysis) with radius

$$\frac{D \ P_A (1 - P_A)}{2 P_A - 1} \qquad \text{when } P_A > .5$$

or

$$\frac{D \ P_A (1 - P_A)}{1 - 2 P_A} \qquad \text{when } P_A < .5$$

and with centre at $\quad x = x_A - \dfrac{D (1 - P_A)^2}{2 P_A - 1}$

and $\quad y = y_A \ (= y_B)$

where D is the distance between the two cities.

On comparing the predictions of this model with the estimates of the subjects, the responses of all subjects were found to correlate at least moderately and some quite highly with the predictions. (The correlation of the predictions of this model with those of the normative model was only .17).

It may be noticed, in passing, how well this new model fits in with one of the three heuristic rules which Tversky (1974) says are utilized by humans in estimating probabilities of events - what Tversky calls "Representativeness". Humans base their estimates of probabilities on the degree to which the event in question is representative of the possible class of events from which it is assumed to have come, this often takes the form of a consideration of the proximity of the event to the mean event of the assumed class. This is exactly the process being postulated here.

Lichtenstein and Feeney concluded that to rely too much on the comparison of subjects' behaviour with a normative model, may make subjects seem to be unreliable estimators of probabilities when, in fact, they may be using some model other than the one proposed, in a very reliable way. A more damning criticism of such an approach, however, is that it removes emphasis from the investigation of what subjects are doing and concentrates on what they are not.

This study raises a point which bears on the use of linear discriminant function analysis by Rodwan and Hake (1964) discussed earlier. The subjects in the present task soon learned to assign any bomb site on the left of the line prescribed by this normative model to city A and any on the right to city B. An analysis along the lines of that employed by Rodwan and Hake, taking as data only the nominal responses A and B would indicate that the subjects' responses were exactly the same as if they had taken a linear combination of the cues (cue y receiving a zero weight) and formed a cut off at the point of equal probability or to put it in simpler terms, the subjects acted just like the linear discriminant function. The analysis of the probability estimates however make this process very unlikely, the subjects were certainly not coming to conclusions about likelihood by a simple weighing of the cues, but this only becomes clear when their responses are stronger than a nominal scale.

We might ask ourselves why subjects made such poor (in the normative sense) estimates of probabilities, when they were after all given immediate feedback on 150 trials which would tend to disconfirm their estimates. The most likely explanation is that the structure of the task was such that they had very strong preconceptions about the probability distribution and hence were not much affected by disconfirming feedback (which after all was only of a nominal scale kind). If the task had been framed in a less concrete, more abstract setting, more accurate estimates may have been generated.

On later re-analysis of this data, Vlek and Van der Heijden (1969) found that the following modification of Lichtenstein and Feeney's model gave an even better fit to the subjects' estimates.

$$S\ L\ L\ R = \alpha \log D_B - \beta \log D_A + \gamma$$

Where S L L R stands for subjective log likelihood ratio defined as

$$S\ L\ L\ R = \log\left(\frac{P_A}{1 - P_A}\right) = \log\left(\frac{P_A}{P_B}\right)$$

α β and γ were estimated by multiple regression. The weights α and β did not generally differ significantly from one another and γ tended to be near zero. D_A and D_B therefore were in general equally important to the subjects. However, α and β were not in general equal to one as Lichtenstein and Feeney's model implied, showing that a subject's responses are better described when his own set of weights is taken into account than when α and β are set equal to one (assuming that the log transformations do not affect the meaning of R_S^2).

Vlek and Van der Heijden (1970) reported an experiment in which the abilities of subjects to utilize information from two cues in making discriminations between two bivariate populations were investigated. The stimuli (based on those of Rodwan and Hake (1964)) were schematic faces. However, the decision to be made was not between "intelligent" and "not intelligent" faces but between the members of two "families". The faces varied along two dimensions, the horizontal X_1 and vertical X_2 co-ordinates of the eyes, and the two families differed in mean position of the eye on both variables, though the dispersion matrices of both "families" were equal, the two cues being uncorrelated <u>within</u> both families.

An experimental block consisted of the presentation of 100 faces (50 from each family) in random order. After each presentation subjects assigned the face to family A or family B and were told immediately of the correct decision. For the next 50 trials (25 faces from each family in random order) no feedback was given, and subjects, besides giving the categorical response were required to assign subjective probabilities (as a percent estimate) to the two populations. In all, six such blocks were completed by each subject - 600 training trials and 300 test trials.

All probability responses were converted into probability of family B and then these were regressed onto the two variables X_1 and X_2. If the subjects had been behaving optimally then the weights assigned to X_1 and X_2 would have been equal. The method of regression used first estimated the β weights to a least squares criterion but then iterated these weights to what was basically a rank order criterion. A comparison of the weights of X_1 and X_2 over experimental blocks showed little change for the average subject whose weightings of X_1 and X_2 were approximately equal. The fit of the model to the subjects' responses seems to have been very good indeed. A loss function S was reported which was said to be roughly equivalent to $(1 - R^2)^{\frac{1}{2}}$ and was shown to decrease over the blocks, showing increasing similarity between subjects' responses and those resulting from a simple linear combination of cues. Values of S as low as .06 were reported,which imply R_S s of about .999, which is an extremely good fit.

We now turn to the first experiment of the present study, which is closely related to this last experiment of Vlek and Van der Heijden. It is in fact an attempt to discover to what extent their findings may be generalised to situations involving stimuli presented in a different modality and to tasks of varying degree of cue intercorrelation.

EXPERIMENT I

DECISIONS BETWEEN BIVARIATE NORMAL POPULATIONS WITH UNEQUAL MEAN VECTORS BUT EQUAL COVARIANCE MATRICES

INTRODUCTION TO EXPERIMENT I.

As previously indicated, research into decision making tasks where the connection between the predictor and predicted variables was assumed to be known to the subject before the commencement of experimentation, had a number of drawbacks:- The true relationship may not be known, the subjects are often required to make decisions of a kind with which they are unfamiliar, the experiences of the judges are left uncontrolled (peculiarities of sampling may have caused the generation of unusual or even generally invalid "rules") etc.

Attempts to overcome problems of this kind led researchers to create artificial tasks in the laboratory, but these tasks are not representative of a large class of decision making situations facing many professionals in their daily lives, and the generalization of conclusions drawn on the basis of data collected in such tasks to real life situations may not be justified. The main drawbacks of these tasks are that they do not in any way resemble differential diagnosis i.e. the decision between two or more populations, but require the subject to respond on an interval scale (i.e. to partition one general population), they are often framed in a completely abstract way (i.e. the prediction of one undefined variable from other variables also undefined), the feedback given to the subject in no way resembles that received in many real life situations.

It was felt that the experiments of Lichtenstein and Feeney (1968) and Vlek and Van der Heijden (1970) overcame many of the objections to the artificial tasks mentioned previously, and that further experimentation along the same lines would not be amiss. Decisions between multivariate normal populations face many professionals in their day to day work. The situation is important enough to justify a chapter on discriminant function analysis in almost all text books of multivariate statistics, but, with the exception of these two papers, the situation has been totally neglected in experimental psychology.

Lichtenstein and Feeney's experiment brought to light an interesting point about the strength of the data which is collected from subjects. Using just the nominal responses it would have been impossible to have distinguished the processes of the judges from that of linear discriminant function analysis. It would have appeared that subjects were behaving optimally - forming the likelihood ratio in an appropriate way, and applying an optimal cutoff at $L(X) = 1$. However, when the subjects' estimates of probabilities, which were assumed to be of interval scale strength, were analysed, it was discovered that the processes giving rise to the subjects' responses were totally dissimilar from those of the linear discriminant function. It is essential therefore, that such interval data should be collected in future experiments if it is hoped to discover the underlying processes of human judges.

Vlek and Van der Heijden showed that, at least with cues presented spatially, which were orthogonal within the populations to be discriminated, subjects' responses (estimates of posterior probabilities) could be very well described by a model which linearly combined the scores on the two cues and based its predictions of probability and presumably its classifications on the value of this linear combination. However, it seems unlikely that subjects actually carried out the operation of weighting and adding the two cues in any real sense, it seems more likely that the orientation of the apparent linear combination was itself regarded as the <u>single</u> relevant cue. In a particular case, for example, when it appears from the regression analysis that a subject is basing his responses on a linear combination of the cues which weights them equally, it seems far more reasonable to assume that the subject is basing his responses simply on the diagonal displacement of the eyes, than actually weighting the two cues "in his head" and basing his decision on the resulting combination value. The problem is that the linear combination actually exists in a physical sense, it is the spatial displacement of the eye along some axis in the two dimensional space, it probably does not exist as the result of some manipulation by the subject of the two separate cue values. This task is not then so much an investigation of subjects' ability to combine information from two cues, as an investigation of their ability to discover the <u>one</u> most relevant cue.

This problem should be overcome by the use of numerically rather than spatially presented cues. Numerical cues also seem more representative of real life situations of the kind it is hoped to simulate and may allow the subject to be more precise in his predictions if he is utilizing some sort of more or less rigid rule. One practical advantage of numerical cues is their ease of presentation in an experimental context. Without computer control fine gradations in displacement are rather difficult to obtain, whereas a number may be expressed to any degree of accuracy which is desired. In the present research, all cues values were given to an accuracy of four decimal places in order to stress to the subject the continuous nature of the cues and to dissuade subjects from considering such cues as primeness, oddness etc., which they had been found to try with integer cues.

It was decided then to carry out an experiment similar to that of Vlek and Van der Heijden but with numerically presented cues. Subjects would be required to learn to discriminate between the members of two bivariate normal populations on the basis of two probabilistic numerical cues and would also be required to give an interval scale response of their certainty in the correctness of each response. For the reasons outlined by Azuma and Cronback (1966) it was decided to differentiate learning from test trials; no feedback being given on test trials so as to give a "purer" indication of performance.

Naylor and Schenck (1968) had shown that cue intercorrelation had a large moderating effect on a number of indices of performance in a multiple cue learning task and had stressed that if conclusions were to be at all generalizable then this factor must be investigated. There are few if any, decision making tasks in real life where the predictor cues are orthognnal and in view of Brunswick's plea for "representative design" it was decided that varying degrees of cue intercorrelation should be considered in the present experiment.

A two cue, two population decision making task was designed and presented to subjects in terms of a medical specialist classifying patients as having one of two diseases on the basis of two medical tests. Both cues were equally useful to the judge, the populations means being one standard deviation apart on each variable. Subjects were run for three sessions (each of 100 training and 50 test trials) under one of three conditions. differing in the degree of cue intercorrelation existing within the populations. All subjects in a condition received the same random sample of "patients" in the same random order, these being generated by computer before the experiment and punched on paper tape. (The method of generation of the samples is indicated in appendix B. A more detailed discussion of the statistical properties of the tasks appears in appendix C).

SUBJECTS

The subjects were 27 members of a Part I psychology course, 18 males and nine females, with an average age of 18 years. They were divided randomly into three groups, with the restraint that an equal proportion of males to females existed in each group i.e. six males and three females. They were not volunteers, except in as much as they chose to participate in this experiment rather than a number of possible alternative ones, in order to fulfil a course requirement. Each subject attended for three one-hour sessions, which took place as nearly as timetable limitations would permit, on consecutive days.

APPARATUS

The apparatus consisted of a Wang 700 B programmable calculator interfaced to a teletype, and a closed circuit television camera and monitor. The subject sat in one cubicle with the television monitor and Wang calculator before him, the teletype and television camera occupied an adjacent cubicle. The stimuli for each session had been previously generated by computer (using the method described in Appendix A) and were punched on a paper tape, which was read by the tape reader of the teletype. The stimuli and the feedback (when this was given) were printed by the teletype and presented to the subjects via the closed circuit television system. The subject made all his responses on the keyboard of the Wang calculator and these were punched on another paper tape by the teletype, and were later analysed by computer.

PROCEDURE

Each one-hour session consisted of 100 training trials followed by 50 test trials:-

The Training Trials. A sample of 100 observations, 50 from each of the two bivariate normal populations, were presented in random order. On each trial the subject was presented with two real numbers (four decimal places) the first representing the patients' scores on test I and the second his score on test II. The subjects decided which disease the patient had (1 or 2) and, having made up his mind stopped an electronic clock in the Wang by pressing the "STEP" button on the keyboard. He then "PRIMED" the machine, pressed button "1" or "2" (to represent his decision) and pressed the "GO" button. This last response caused the Wang to send an instruction to the teletype which caused the present stimulus pair to be removed and to be replaced by the correct response. The correct response was displayed for three seconds in which time the teletype punched the subject's response and his reaction time. When the three seconds were over the next stimulus pair was automatically displayed and the sequence was repeated, the stimuli being displayed for as long as it took the subject to respond.

The Test Trials. A sample of 50 observations, 25 from each of the two bivariate normal populations were presented in random order. On each trial, however, besides his dichotomous "1" or "2" response, the subject was required to indicate his degree of certainty about this response by entering on the Wang keyboard a number from 1 to 50 inclusive ("1" was to represent a guess i.e. complete uncertainty and "50" complete certainty). This response was made immediately following the "1" or "2" response and was followed by pressing the "GO" button. This last command caused the stimulus pair to be removed and the screen remained blank for three seconds whilst the subjects' responses and reaction time were punched on paper tape by the teletype. Following this, the next stimulus pair was displayed.

the machine, pressed button "1" or "2" (to represent his decision) and pressed the "GO" button. This last response caused the Wang to send an instruction to the teletype which caused the present stimulus pair to be removed and to be replaced by the correct response. The correct response was displayed for three seconds in which time the teletype punched the subject's response and his reaction time. When the three seconds were over the next stimulus pair was automatically displayed and the sequence was repeated, the stimuli being displayed for as long as it took the subject to respond.

The Test Trials. A sample of 50 observations, 25 from each of the two bivariate normal populations were presented in random order. On each trial, however, besides his dichotomous "1" or "2" response, the subject was required to indicate his degree of certainty about this response by entering on the Wang keyboard a number from 1 to 50 inclusive ("1" was to represent a guess i.e. complete uncertainty and "50" complete certainty). This response was made immediately following the "1" or "2" response and was followed by pressing the "GO" button. This last command caused the stimulus pair to be removed and the screen remained blank for three seconds whilst the subjects' responses and reaction time were punched on paper tape by the teletype. Following this, the next stimulus pair was displayed.

EXPERIMENTAL INSTRUCTIONS

The subject was asked to sit before the Wang calculator and the television monitor on which was displayed the first stimulus pair. On the first session the following instructions were given to the subject:-

"This experiment is a simulation of a medical decision-making task. You are to put yourself in the place of a medical specialist whose interest is restricted to two diseases which (for the sake of simplicity) we call Disease 1 and Disease 2. Patients will be referred to you when it is known that they have either Disease 1 or Disease 2, your job is to decide which one of the diseases a particular patient has. (He cannot have both diseases)."

"To help you make your decision, the score of each patient on two medical tests will be given to you. The test results are shown on the television screen in front of you. This is patient number one." (At this point the experimenter pointed to the "1" on the television screen). "This is his score on the first test." (The experimenter pointed to the first score). "This is his score on the second test." (The experimenter pointed to the second score).

"You need not worry about what the tests actually are, you could imagine them to be such tests as blood pressure, blood sugar concentration etc. Each of them, however, gives a score on a continuous scale and is useful in some way in predicting the disease of a patient. The scores are given to so many decimal points to stress the continuous nature of the tests."

"At present, of course, you have no idea which test scores are indicative of Disease 1 and which of Disease 2, but, for the first hundred trials of this experiment you will be told the correct disease of the patient as soon as you have made your diagnosis. Now, if you will guess the disease of the first patient here, you will see what happens." (The subject guessed Disease 1 or Disease 2 and the experimenter entered his response for him on the Wang keyboard. The correct Disease was shown on the monitor and after three seconds was replaced by the test scores of the second patient).

"When you have made up your mind which disease the patient has, press the "STEP" button. This stops a clock within the machine which is recording your decision time. Then "PRIME" the machine, press button "1" or "2" and then the "GO" button."

"Half the patients in each session have Disease 1 and the other half Disease 2. If you guess randomly, therefore, you will probably get about half of them right. However, you will not get all of them right, even if you are using the tests in an optimal way. This is because the tests are not perfect predictors of the diseases. I will try to make this clearer by giving you a similar example from outside medical decision making:-"

"Imagine that we take a person at random from the population and measure that person's height and weight. I give these two scores to you and ask you whether the person is Male or <u>Female</u>. You might expect males to be taller and heavier in general, but if I give you a large number of people, about whom you had to decide, you would not expect to get them all right, as there are quite a large number of small men, heavy women, tall women etc. You cannot predict sex perfectly from knowledge of just height and weight, though such knowledge will increase the accuracy of your predictions quite appreciably over the chance level."

"In a very similar way to this, Diseases 1 and 2 cannot be perfectly predicted from the two scores. You will be doing very well if you correctly diagnose 90% of the patients. I want you to perform as well as you can, try to find connections between the test scores and the diseases and use these to improve the accuracy of your diagnoses."

"Now before you start, I would just like to stress that <u>both</u> tests are useful in predicting <u>both</u> diseases, and that both diseases should be regarded as equally dangerous to the patient - you should not bias your responses in favour of either disease on that ground."

"Are there any questions?" (Questions were answered by the repetition of the appropriate part of the above instructions).

(The subject now made four or five responses in the presence of the experimenter to ensure that he had the "hang" of the equipment).

"Now there are 100 of these training trials, followed by 50 trials which are slightly different. In order to finish the whole session in an hour, you should respond at a rate of about three patients a minute. I will leave you now to complete these first 100 trials, after which I will come back to tell you about the test trials."

After the subject has completed the training trials, the experimenter returned to the cubicle and the following instructions were given.

"The next 50 trials will be slightly different. You probably found that certain patients were easier to diagnose correctly than were others. Well, I would like you to show me how confident you are about each of your diagnoses of the next 50 patients who also either have Disease 1 or Disease 2. You will press button "1" or "2" as before, but, immediately following this I want you to press a number between 1 and 50 to show me how certain you are about your diagnosis. (Pressing "50" means you are absolutely certain, pressing "1" means you are just guessing, not at all certain, pressing "25" would mean you are "half certain" if you like). Then press the "GO" button as before. You will not be told the correct response, but after a while the scores of the next patient will appear. The reason why you are not given the correct disease is so that you will not learn anything new in these trials, I just want to know what you have learned during the first part of the session."

"Again the 50 patients are made up of equal numbers from both diseases, presented in random order."

"Any questions?" (Questions were dealt with as above).

(The experimenter remained with the subject whilst he completed four or five trials, then left the cubicle).

On subsequent sessions the bare outline of the task was repeated to the subject, questions were answered as far as was possible. At the end of the third session, each subject was asked to write a few sentences on the manner in which he finally came to make his decisions and certainty estimates.

EXPERIMENT I

Results

Only the responses of the subjects on the 50 test trials of each session were subjected to analysis, since it was felt that performance on these trials would give a more pure measure of performance. An attempt had been made to measure subjects' decision times, but it was discovered at an early stage in the analysis that this measure was not significantly correlated with any of the variables under consideration. One possible explanation of this is that the equipment was not in fact capable of recording the time passing between the presentation of the stimulus pair and the subjects' "1" or "2" response, instead it recorded the time elapsing between presentation and the subject's pressing of the "STEP" button. The subjects had been instructed not to press this button until they had "made up their minds" about their response, but there is reason to believe that this instruction was not always obeyed and that the subjects tended to press "STEP" very soon after the presentation of the stimuli and then consider the implications of these stimuli for their diagnosis. For this reason no decision time data is reported here, the analysis being restricted to the categorical "1" - "2" response and the subjects' certainty responses.

The categorical "1" - "2" responses were analysed with respect to the decision of the normative model on each trial. Thus if a patient was drawn from Disease 1 but the normative model said that with these scores Disease 2 was more likely, if the subject placed this patient in Disease 2, his response was regarded as correct. This index when totalled over trials, besides being a purer measure of performance is unbiased by the cue intercorrelations in the disease populations, allowing subjects in all conditions to achieve a maximum score of 100% if all their diagnoses agreed with those of the normative model. Whereas total number correct in terms of the disease population from which each patient was drawn has a different maximum value in each condition thus rendering meaningful comparisons impossible. This index, shown in Tables 1 (a, b and c) was subjected to analysis of variance (Table 1 d) which indicated a significant effect of sessions only, which proved to be due to significantly better performance in session III than in session I and II.

The certainty responses, which could range from 50 (completely certain) to 1 (completely uncertain) for Disease 1, and from 1 (completely uncertain) to 50 (completely certain) for Disease 2 were transformed to yield a scale ranging from 1 (completely certain for Disease 1) to 100 (completely certain for Disease 2). The resulting values were correlated with the certainty estimates of the normative model and the correlation coefficients (Tables 2 (a,b and c)) provide a second index of performance. An analysis of variance on this measure showed an exactly similar pattern of significant results as the total number correct index (Table 2 d).

NUMBER OF CORRECT RESPONSES

(i.e. Agreements with the Normative Model)

Table 1a Condition I (0.0,0.0)

Subject	Session		
	I	II	III
1	25 n.s.	30	27 n.s.
2	36	38	38
3	40	40	41
4	43	47	48
5	35	36	42
6	39	37	45
7	36	36	38
8	26 n.s.	40	43
9	22 n.s.	26 n.s.	24 n.s.

Table 1b Condition II (0.7,0.7)

Subject	Session		
	I	II	III
10	34	28 n.s.	32
11	26 n.s.	25 n.s.	44
12	20 n.s.	15 n.s.	25 n.s.
13	45	43	46
14	26 n.s.	23 n.s.	25 n.s.
15	24 n.s.	25 n.s.	31
16	28 n.s.	26 n.s.	23 n.s.
17	33	47	47
18	23 n.s.	23 n.s.	25 n.s.

NUMBER OF CORRECT RESPONSES

(i.e. Agreements with the Normative Model)

Table 1c Condition III (0.9,0.9)

Subject	Session		
	I	II	III
19	27 n.s.	29 n.s.	24 n.s.
20	28 n.s.	27 n.s.	42
21	29 n.s.	29 n.s.	47
22	26 n.s.	43	42
23	21 n.s.	40	24 n.s.
24	26 n.s.	24 n.s.	44
25	29 n.s.	25 n.s.	26 n.s.
26	23 n.s.	42	49
27	35	29 n.s.	33

n.s. = Not significantly different from chance level.

Table 1d

ANALYSIS OF VARIANCE ON NUMBER OF CORRECT RESPONSES

Source	S.S.	d.f.	M.S.	F	P
Between Subjects	3753.06	26			
CONDITIONS	535.58	2	267.79	1.998	N.S.
Subjects within Groups	3217.48	24	134.06		
Within Subjects	2142.00	54			
SESSIONS	542.32	2	271.16	8.760	<.01
CONDITIONS X SESSIONS	113.83	4	28.46	<1	
SESSIONS X Subjects within Groups	1485.85	48	30.96		

SIGNIFICANCE OF DIFFERENCES BETWEEN INDIVIDUAL PAIRS OF SESSION MEANS

	MEAN	SESSION II	SESSION III
MEAN		32.33	36.11
SESSION I	29.81	2.44	6.30 *
SESSION II	32.33		3.78 *

Critical value of d = 3.32 at .05 Level.

* Significant at the .05 Level.

CORRELATIONS OF SUBJECTS' CERTAINTIES WITH THOSE OF THE NORMATIVE MODEL (r_a)

Table 2a Condition I (0.0,0.0)

Subject	Session		
	I	II	III
1	.134 n.s.	.349	.202 n.s.
2	.344	.764	.819
3	.764	.721	.794
4	.931	.883	.882
5	.592	.522	.852
6	.724	.473	.876
7	.616	.603	.657
8	.062 n.s.	.686	.684
9	- .037 n.s.	.177 n.s.	.085 n.s.

Table 2b Condition II (0.7,0.7)

Subject	Session		
	I	II	III
10	.375	.110 n.s.	.288
11	- .062 n.s.	.156 n.s.	.797
12	- .282	- .371	- .069 n.s.
13	.842	.967	.932
14	- .047 n.s.	- .076 n.s.	- .001 n.s.
15	.192 n.s.	.052 n.s.	.164 n.s.
16	.038 n.s.	.156 n.s.	.162 n.s.
17	.184 n.s.	.868	.829
18	.057 n.s.	- .222 n.s.	.011 n.s.

CORRELATION OF SUBJECTS' CERTAINTIES
WITH THOSE OF THE NORMATIVE MODEL (r_a)

Table 2c   Condition III (0.9,0.9)

Subject	Session		
	I	II	III
19	.127 n.s.	.103 n.s.	- .022 n.s.
20	.075 n.s.	.112 n.s.	.865
21	.222 n.s.	.177 n.s.	.831
22	.101 n.s.	.863	.937
23	- .200 n.s.	.596	- .015 n.s.
24	.038 n.s.	- .054 n.s.	.639
25	.109 n.s.	.129 n.s.	.173 n.s.
26	- .071 n.s.	.647	.834
27	.689	.356	.543

Table 2d

ANALYSIS OF VARIANCE ON r_a

(Fisher's Z Transformation)

Source	S.S.	d.f.	M.S.	F	P
Between Subjects	16.670	26			
CONDITIONS	2.450	2	1.225	2.068	N.S.
Subjects within Groups	14.219	24	.592		
Within Subjects	8.302	52			
SESSIONS	2.322	2	1.161	9.566	< .01
CONDITIONS X SESSIONS	.326	4	.081	< 1	
SESSIONS X Subjects within Groups	5.583	46	.121		

SIGNIFICANCE OF DIFFERENCES BETWEEN SESSION MEANS

(FISHER'S Z TRANSFORMATION)

	MEAN	SESSION II	SESSION III
MEAN		.338	.746
SESSION I	.436	- .102	.316 *
SESSION II	.338		.408 *

Critical value of d = .157 at .05 Level.

* Significant at the .05 Level.

A multiple regression was carried out using the certainty values with the two cues X_1 and X_2 as predictor variables. The resulting multiple correlation coefficients (R_s for the subjects and R_e for the normative model) are shown in Tables 3 (a, b and c) and an analysis of variance on this index (Table 3 d) showed a significant effect of sessions only, with R_s lower in session I than in sessions II and III. The correlation of the cues X_1 and X_2 individually with the certainty values are displayed in Tables 4 (a (i), b (i) and c(i)) and Tables 5 (a (i), b (i), c (i)) respectively and F ratios to test the significance of the contribution of each term to the multiple correlation coefficients R_ss are shown in the adjacent Tables 4 (a (ii), b (ii), c (ii)) and Tables 5 (a (ii), b (ii) and c (ii)).

The correlations between the certainty values and a linear combination of the cues X_1 and X_2 which assigned equal weight to both (here called the Sum of X_1 and X_2 though it could equally well be regarded as the mean of their values) were found and are presented in Tables 6 (a (i), b (i) and c (i)). the significance of the better fit of the linear combination of the cues with estimated weights over this one with weights set equal was tested and the values of the F ratios are shown in Tables 6 (a (ii), b (ii) and c (ii)). an analysis of variance on these correlation coefficients (Table 6 d) shows no significant effect of condition though the magnitude of the correlations is significantly greater in session III than in the other two sessions.

The Lens Model indices were computed and are shown, G in

Tables 7 (a (i), b (i) and c (i)), GR_eR_s in Tables 7 (a (ii), b (ii) and c (ii)), C in Tables 7 (a (iii), b (iii) and c (iii)) and $C\sqrt{(1 - R_e^2)(1 - R_s^2)}$ in Tables 7 (a (iv), b (iv) and c (iv)). Analyses of variance on the G (Table 7 d) and C (Table 7 e) indices of linear and non-linear matching showed a significant effect of session only for both indices. Both G and C were significantly greater in session III than in session I but were not significantly different from either of these values in session II.

R_s THE CORRELATION OF SUBJECTS' CERTAINTY ESTIMATES WITH A LINEAR COMBINATION OF CUES WITH WEIGHTS ESTIMATED TO A LEAST SQUARES CRITERION

Table 3a   Condition I (0.0,0.0)

Subject	Session		
	I	II	III
1	.169 n.s.	.321 n.s.	.233 n.s.
2	.268 n.s.	.822	.817
3	.762	.693	.788
4	.907	.832	.872
5	.631	.580	.839
6	.732	.679	.827
7	.605	.559	.650
8	.304 n.s.	.651	.579
9	.048 n.s.	.587	.661
(Model)	(.975)	(.981)	(.972)

Table 3b   Condition II (0.7,0.7)

Subject	Session		
	I	II	III
10	.343 n.s.	.109 n.s.	.255 n.s.
11	.193 n.s.	.142 n.s.	.784
12	.552	.348	.823
13	.787	.918	.896
14	.092 n.s.	.101 n.s.	.329 n.s.
15	.416	.644	.765
16	.084 n.s.	.554	.188 n.s.
17	.170 n.s.	.837	.777
18	.098 n.s.	.279 n.s.	.114 n.s.
(Model)	(.982)	(.992)	(.988)

R_s. THE CORRELATION OF SUBJECTS' CERTAINTY ESTIMATES WITH A LINEAR COMBINATION OF CUES WITH WEIGHTS ESTIMATED TO A LEAST SQUARES CRITERION

Table 3c Condition III (0.9,0.9)

Subject	Session		
	I	II	III
19	.395	.597	.125 n.s.
20	.566	.907	.832
21	.227 n.s.	.577	.800
22	.065 n.s.	.843	.916
23	.257 n.s.	.595	.322 n.s.
24	.065 n.s.	.813	.629
25	.160 n.s.	.732	.238 n.s.
26	.513	.626	.803
27	.812	.367	.590
(Model)	(.984)	(.992)	(.991)

Table 3d

ANALYSIS OF VARIANCE ON R_s

(FISHER'S Z TRANSFORMATION)

Source	S.S.	d.f.	M.S.	F	P
Between Subjects	8.346	26			
CONDITIONS	.742	2	.371	1.172	N.S.
Subjects within Groups	7.604	24	.317		
Within Subjects	6.987	54			
SESSIONS	1.990	2	.995	10.223	<.01
CONDITIONS X SESSIONS	.326	4	.082	< 1	
SESSIONS X Subjects within Groups	4.671	48	.097		

SIGNIFICANCE OF DIFFERENCES BETWEEN INDIVIDUAL PAIRS OF SESSION MEANS (FISHER'S Z TRANSFORMATION)

		SESSION	
	MEAN	II	III
MEAN		.750	.820
SESSION I	.458	.292 *	.362 *
SESSION II	.750		.070

Critical value of d = .171 at the .05 Level.

* Significant at the .05 Level.

Table 3d

ANALYSIS OF VARIANCE ON R_s

(FISHER'S Z TRANSFORMATION)

Source	S.S.	d.f.	M.S.	F	P
Between Subjects	8.346	26			
CONDITIONS	.742	2	.371	1.172	N.S.
Subjects within Groups	7.604	24	.317		
Within Subjects	6.987	54			
SESSIONS	1.990	2	.995	10.223	<.01
CONDITIONS X SESSIONS	.326	4	.082	<1	
SESSIONS X Subjects within Groups	4.671	48	.097		

SIGNIFICANCE OF DIFFERENCES BETWEEN
INDIVIDUAL PAIRS OF SESSION MEANS
(FISHER'S Z TRANSFORMATION)

	MEAN	SESSION II	SESSION III
MEAN		.750	.820
SESSION I	.458	.292 *	.362 *
SESSION II	.750		.070

Critical value of d = .171 at the .05 Level.

* Significant at the .05 Level.

Table 4a (i) Condition I (0.0,0.0)

THE CORRELATION OF SUBJECTS' CERTAINTIES WITH THE CUE X_1

Subject	Session		
	I	II	III
1	- .056	- .137	- .092
2	- .215	- .766	- .726
3	- .539	- .462	- .713
4	- .631	- .634	- .644
5	- .512	- .538	- .673
6	- .638	- .677	- .698
7	- .312	- .344	- .584
8	.164	- .395	- .433
9	.032	- .542	- .454
(Model)	(- .635)	(- .660)	(- .759)

Table 4a (ii) Condition I (0.0,0.0)

F RATIOS TO TEST THE SIGNIFICANCE OF THE CONTRIBUTION OF THE X_1 TERM TO THE FIT OF THE LINEAR COMBINATION

Subject	Session		
	I	II	III
1	.13 n.s.	.74 n.s.	.15 n.s.
2	2.25 n.s.	80.79	60.54
3	30.82	16.59	51.90
4	99.60	54.75	56.83
5	19.69	19.35	53.45
6	39.93	40.28	56.20
7	6.62	6.82	22.79
8	1.51 n.s.	10.67	9.33
9	.45 n.s.	22.08	23.17
(Model)	(357.94)	(466.91)	(363.06)

Table 5a (i) Condition I (0.0,0.0)

THE CORRELATION OF SUBJECTS' CERTAINTIES

WITH THE CUE X_2

Subject	Session		
	I	II	III
1	- .161	- .298	- .227
2	- .165	- .346	- .490
3	- .553	- .545	- .448
4	- .668	- .576	- .687
5	- .382	- .252	- .606
6	- .376	.011	- .553
7	- .527	- .462	- .376
8	- .251	- .542	- .451
9	.036	.193	.399
(Model)	(- .756)	(- .766)	(- .725)

Table 5a (ii) Condition I (0.0,0.0)

F RATIOS TO TEST THE SIGNIFICANCE OF THE

CONTRIBUTION OF THE X_2 TERM

TO THE FIT OF THE LINEAR COMBINATION

Subject	Session		
	I	II	III
1	1.23 n.s.	4.42	2.28 n.s.
2	1.29 n.s.	12.91	19.96
3	32.58	24.21	13.93
4	112.17	44.05	68.13
5	10.60	3.37 n.s.	39.90
6	13.06	.26 n.s.	29.21
7	19.94	13.28	6.53
8	3.39 n.s.	21.88	10.47
9	.59 n.s.	3.72 n.s.	19.31
(Model)	(516.94)	(654.28)	(318.87)

Table 4b (i) Condition II (0.7,0.7)

THE CORRELATION OF SUBJECTS' CERTAINTIES

WITH THE CUE X_1

Subject	Session		
	I	II	III
10	- .336	- .103	- .251
11	.139	- .126	- .777
12	.430	.338	.398
13	- .742	- .838	- .852
14	.046	.071	- .114
15	- .038	.225	.128
16	- .063	.051	- .182
17	- .105	- .801	- .724
18	- .033	.124	- .074
(Model)	(- .913)	(- .919)	(- .912)

Table 4b (ii) Condition II (0.7,0.7)

F RATIOS TO TEST THE SIGNIFICANCE OF THE

CONTRIBUTION OF THE X_1 TERM

TO THE FIT OF THE LINEAR COMBINATION

Subject	Session		
	I	II	III
10	1.25 n.s.	.45 n.s.	1.03 n.s.
11	1.79 n.s.	.08 n.s.	25.87
12	19.65	1.62 n.s.	92.19
13	8.26	27.48	35.52
14	.39 n.s.	.00 n.s.	4.79
15	5.10	28.17	45.64
16	.32 n.s.	12.86	.41 n.s.
17	.08 n.s.	21.99	10.31
18	.13 n.s.	.47 n.s.	.62 n.s.
(Model)	(106.92)	(384.26)	(246.18)

Table 5b (i) Condition II (0.7,0.7)

THE CORRELATION OF SUBJECTS' CERTAINTIES

WITH THE CUE X_2

Subject	Session		
	I	II	III
10	- .307	- .050	- .211
11	.025	- .136	- .634
12	.119	.302	- .209
13	- .743	- .866	- .808
14	- .014	.101	.130
15	- .288	- .251	- .428
16	- .014	- .342	- .165
17	- .166	- .749	- .719
18	- .083	.262	.006
(Model)	(- .938)	(- .924)	(- .923)

Table 5b (ii) Condition II (0.7,0.7)

F RATIOS TO TEST THE SIGNIFICANCE OF THE

CONTRIBUTION OF THE X_2 TERM

TO THE FIT OF THE LINEAR COMBINATION

Subject	Session		
	I	II	III
10	.28 n.s.	.62 n.s.	.92 n.s.
11	.88 n.s.	.20 n.s.	1.29 n.s.
12	8.36	.36 n.s.	75.41
13	8.42	42.08	18.36
14	.30 n.s.	.24 n.s.	5.00
15	9.73	29.17	64.58
16	.15 n.s.	20.63	.11 n.s.
17	.87 n.s.	9.25	9.52
18	.40 n.s.	3.18 n.s.	.36 n.s.
(Model)	(167.54)	(410.89)	(285.93)

Table 4c (i) Condition III (0.9,0.9)

THE CORRELATION OF SUBJECTS' CERTAINTIES

WITH THE CUE X_1

Subject	Session		
	I	II	III
19	- .059	.022	.040
20	.051	.053	- .821
21	- .201	- .074	- .771
22	- .065	- .831	- .869
23	.182	- .593	- .057
24	- .050	- .098	- .592
25	- .099	.001	- .204
26	.150	- .618	- .780
27	- .766	- .367	- .578
(Model)	(- .965)	(- .973)	(- .964)

Table 4c (ii) Condition III (0.9,0.9)

F RATIOS TO TEST THE SIGNIFICANCE OF THE

CONTRIBUTION OF THE X_1 TERM

TO THE FIT OF THE LINEAR COMBINATION

Subject	Session		
	I	II	III
19	6.67	22.32	.74 n.s.
20	20.61	191.48	7.85
21	.04 n.s.	17.07	2.24 n.s.
22	.01 n.s.	7.19	3.26 n.s.
23	.58 n.s.	2.63 n.s.	5.09
24	.15 n.s.	83.24	.26 n.s.
25	1.12 n.s.	45.14	1.92 n.s.
26	16.77	1.96 n.s.	3.34 n.s.
27	37.25	1.48 n.s.	8.79
(Model)	(33.68)	(141.27)	(107.02)

Table 5c (i) Condition III (0.9,0.9)

THE CORRELATION OF SUBJECTS' CERTAINTIES

WITH THE CUE X_2

Subject	Session		
	I	II	III
19	- .190	- .224	- .015
20	- .147	- .322	- .800
21	- .225	- .302	- .790
22	- .063	- .817	- .910
23	.233	- .564	.082
24	- .033	.241	- .626
25	- .049	- .299	- .134
26	- .029	- .606	- .787
27	- .625	- .328	- .475
(Model)	(- .972)	(- .966)	(- .971)

Table 5c (ii) Condition III (0.9,0.9)

F RATIOS TO TEST THE SIGNIFICANCE OF THE

CONTRIBUTION OF THE X_2 TERM

TO THE FIT OF THE LINEAR COMBINATION

Subject	Session		
	I	II	III
19	8.50	25.94	.68 n.s.
20	21.93	218.34	2.72
21	.55 n.s.	23.12	6.06
22	.00 n.s.	3.33 n.s.	24.51
23	1.66 n.s.	.24 n.s.	5.27
24	.08 n.s.	89.98	3.48 n.s.
25	.77 n.s.	54.16	.73 n.s.
26	15.39	.83 n.s.	4.71
27	10.16	.02 n.s.	.96 n.s.
(Model)	(55.62)	(99.83)	(144.29)

Table 6a (i) Condition I (0.0,0.0)

THE CORRELATION OF SUBJECTS' CERTAINTIES

WITH THE SUM OF X_1 and X_2

Subject	Session		
	I	II	III
1	- .160	- .301	- .203
2	- .259	- .757	- .805
3	- .757	- .692	- .770
4	- .902	- .830	- .869
5	- .607	- .538	- .839
6	- .678	- .448	- .824
7	- .600	- .554	- .636
8	- .098	- .644	- .577
9	.048	- .229	- .069
(Model)	(- .974)	(- .979)	(- .972)

Table 6a (ii) Condition I (0.0,0.0)

F RATIOS TO TEST THE SIGNIFICANCE

OF THE BETTER FIT OF THE LINEAR COMBINATION WITH

ESTIMATED WEIGHTS OVER THAT OF THE SUM OF X_1 AND X_2

Subject	Session		
	I	II	III
1	.63 n.s.	.66 n.s.	.65 n.s.
2	.24 n.s.	14.92	2.85 n.s.
3	.84 n.s.	.21 n.s.	3.41 n.s.
4	2.11 n.s.	.51 n.s.	1.16 n.s.
5	2.31 n.s.	3.37	.00 n.s.
6	7.72	22.79	.76 n.s.
7	.45 n.s.	.40 n.s.	1.40 n.s.
8	4.28	.75 n.s.	.13 n.s.
9	0.00 n.s.	20.98	36.10
(Model)	(1.29)	(4.63)	(.50)

Table 6b (i) Condition II (0.7,0.7)

THE CORRELATION OF SUBJECTS' CERTAINTIES

WITH THE SUM OF X_1 AND X_2

Subject	Session		
	I	II	III
10	- .339	- .082	- .249
11	.081	- .141	- .761
12	.275	.344	.109
13	- .786	- .918	- .894
14	.014	.093	.006
15	- .185	- .022	- .156
16	- .038	- .164	- .187
17	- .146	- .833	- .777
18	- .064	.210	- .037
(Model)	(- .982)	(- .992)	(- .988)

Table 6b (ii) Condition II (0.7,0.7)

F RATIOS TO TEST THE SIGNIFICANCE

OF THE BETTER FIT OF THE LINEAR COMBINATION WITH

ESTIMATED WEIGHTS OVER THAT OF THE SUM OF X_1 AND X_2

Subject	Session		
	I	II	III
10	.16 n.s.	.25 n.s.	.14 n.s.
11	1.51 n.s.	.01 n.s.	4.23
12	15.51	.16 n.s.	96.79
13	.08 n.s.	.22 n.s.	.64 n.s.
14	.39 n.s.	.07 n.s.	5.68
15	7.86	33.20	63.68
16	.26 n.s.	18.99	.02 n.s.
17	.37 n.s.	1.05 n.s.	.00 n.s.
18	.26 n.s.	1.71 n.s.	.55 n.s.
(Model)	(.00)	(.00)	(1.17)

Table 6c (i) Condition III (0.9,0.9)

THE CORRELATION OF SUBJECTS' CERTAINTIES

WITH THE SUM OF X_1 AND X_2

Subject	Session		
	I	II	III
19	- .130	- .106	.013
20	- .054	- .141	- .831
21	- .217	- .195	- .799
22	- .065	- .843	- .911
23	.212	- .591	.012
24	- .042	.077	- .624
25	- .074	- .155	- .174
26	.057	- .626	- .803
27	- .703	- .355	- .540
(Model)	(- .984)	(- .991)	(- .991)

Table 6c (ii) Condition III (0.9,0.9)

F RATIOS TO TEST THE SIGNIFICANCE

OF THE BETTER FIT OF THE LINEAR COMBINATION WITH

ESTIMATED WEIGHTS OVER THAT OF THE SUM OF X_1 AND X_2

Subject	Session		
	I	II	III
19	7.75	25.16	.74 n.s.
20	21.91	213.78	.33 n.s.
21	.22 n.s.	20.83	.25 n.s.
22	.00 n.s.	.25 n.s.	2.67 n.s.
23	1.06 n.s.	.36 n.s.	5.43
24	.11 n.s.	90.48	.49 n.s.
25	.98 n.s.	51.72	1.31 n.s.
26	16.62	.08 n.s.	.02 n.s.
27	22.94	.48 n.s.	4.06
(Model)	(.29)	(1.65)	(.54)

Table 6d

ANALYSIS OF VARIANCE ON THE CORRELATIONS
OF SUBJECTS' CERTAINTIES WITH THE TERM $(X_1 + X_2)$

(FISHER'S Z TRANSFORMATION)

Source	S.S.	d.f.	M.S.	F	P
Between Subjects	13.587	26			
CONDITIONS	2.265	2	1.133	2.401	N.S.
Subjects within groups	11.322	24	.472		
Within Subjects	7.033	54			
SESSIONS	1.910	2	.955	9.482	< .01
CONDITIONS X SESSIONS	.288	4	.072	< 1	
SESSIONS X Subjects within Groups	4.835	48	.101		

SIGNIFICANCE OF DIFFERENCES BETWEEN
INDIVIDUAL PAIRS OF SESSION MEANS
(FISHER'S Z TRANSFORMATION)

		SESSION	
	MEAN	II	III
MEAN		- .448	- .665
SESSION I	- .290	- .158	- .375 *
SESSION II	- .448		- .217 *

Critical value of d = .175 at the .05 level.

* Significant at the .05 Level.

THE LENS MODEL INDICES

Table 7a (i) Condition I (0.0,0.0)

THE MATCHING INDEX G

Subject	Session		
	I	II	III
1	.931	.956	.882
2	.975	.895	.981
3	.997	1.000	.973
4	.998	.992	.998
5	.972	.902	.999
6	.940	.612	.994
7	.986	.997	.975
8	.287	.996	.999
9	- 1.000	.333	.081

Table 7a (ii) Condition I (0.0,0.0)

$G \times R_e \times R_s$

Subject	Session		
	I	II	III
1	.154	.301	.200
2	.254	.722	.779
3	.741	.680	.745
4	.882	.809	.846
5	.598	.513	.815
6	.671	.408	.799
7	.582	.547	.616
8	.085	.636	.562
9	- .046	.192	.052

THE LENS MODEL INDICES

Table 7a (iii) Condition I (0.0,0.0)

THE MATCHING INDEX C

Subject	Session		
	I	II	III
1	- .089	.261	.007
2	.414	.381	.295
3	.163	.292	.341
4	.521	.688	.313
5	- .031	.057	.287
6	.352	.454	.584
7	.190	.347	.229
8	- .131	.336	.637
9	.043	- .095	.189

Table 7a (iv) Condition I (0.0,0.0)

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

Subject	Session		
	I	II	III
1	- .020	.048	.002
2	.089	.042	.040
3	.024	.041	.049
4	.049	.074	.036
5	- .005	.009	.037
6	.054	.065	.077
7	.034	.056	.041
8	- .028	.050	.121
9	.010	- .015	.033

THE LENS MODEL INDICES

Table 7b (i) Condition II (0.7,0.7)

THE MATCHING INDEX G

Subject	Session		
	I	II	III
10	.986	.751	.974
11	- .414	.996	.966
12	- .493	- .988	- .110
13	.999	.999	.997
14	- .147	- .920	- .042
15	.451	.029	.226
16	.451	.289	.990
17	.861	.996	1.000
18	.659	- .750	.306

Table 7b (ii) Condition II (0.7,0.7)

$G \times R_e \times R_s$

Subject	Session		
	I	II	III
10	.332	.081	.245
11	- .079	.140	.747
12	- .267	- .341	- .089
13	.772	.910	.882
14	- .013	- .092	- .014
15	.184	.018	.171
16	.037	.159	.184
17	.144	.827	.768
18	.063	- .208	.035

Table 7b (iii) Condition II (0.7,0.7)

THE MATCHING INDEX C

Subject	Session		
	I	II	III
10	.239	.231	.289
11	.088	.124	.520
12	- .093	- .250	.226
13	.598	.732	.719
14	- .179	.131	.089
15	.049	.347	- .070
16	.004	- .025	- .146
17	.214	.591	.626
18	- .035	- .122	- .151

Table 7b (iv) Condition II (0.7,0.7)

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

Subject	Session		
	I	II	III
10	.043	.029	.043
11	.017	.016	.050
12	- .015	- .030	.020
13	.071	.037	.050
14	- .034	.016	.013
15	.009	.034	- .007
16	.001	- .003	- .022
17	.040	.041	.061
18	- .007	- .015	- .023

Table 7b (iii) Condition II (0.7,0.7)

THE MATCHING INDEX C

Subject	Session		
	I	II	III
10	.239	.231	.289
11	.088	.124	.520
12	- .093	- .250	.226
13	.598	.732	.719
14	- .179	.131	.089
15	.049	.347	- .070
16	.004	- .025	- .146
17	.214	.591	.626
18	- .035	- .122	- .151

Table 7b (iv) Condition II (0.7,0.7)

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

Subject	Session		
	I	II	III
10	.043	.029	.043
11	.017	.016	.050
12	- .015	- .030	.020
13	.071	.037	.050
14	- .034	.016	.013
15	.009	.034	- .007
16	.001	- .003	- .022
17	.040	.041	.061
18	- .007	- .015	- .023

Table 7c (i) Condition III (0.9,0.9)

THE MATCHING INDEX G

Subject	Session		
	I	II	III
19	.341	.154	- .085
20	.109	.133	.997
21	.960	.316	.999
22	.994	1.000	.996
23	- .833	.995	- .056
24	.636	- .072	.994
25	.449	.190	.718
26	- .097	1.000	1.000
27	.858	.972	.908

Table 7c (ii) Condition III (0.9,0.9)

$G \times R_e \times R_s$

Subject	Session		
	I	II	III
19	.133	.091	- .011
20	.061	.120	.822
21	.214	.181	.793
22	.064	.836	.905
23	- .211	.588	- .018
24	.040	- .058	.620
25	.071	.138	.169
26	- .049	.621	.796
27	.686	.354	.531

THE LENS MODEL INDICES

Table 7c (iii) Condition III (0.9,0.9)

THE MATCHING INDEX C

Subject	Session		
	I	II	III
19	- .027	.107	- .083
20	.097	- .146	.588
21	.046	- .039	.482
22	.210	.392	.618
23	.063	.077	.023
24	- .015	.050	.191
25	.158	- .096	.032
26	- .143	.259	.489
27	.035	.018	.115

Table 7c (iv) Condition III (0.9,0.9)

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

Subject	Session		
	I	II	III
19	- .004	.011	- .011
20	.014	- .008	.043
21	.008	- .004	.038
22	.038	.027	.033
23	.011	.008	.003
24	- .003	.004	.020
25	.028	- .009	.004
26	- .022	.026	.039
27	.004	.002	.012

Table 7d

ANALYSIS OF VARIANCE ON THE MATCHING INDEX G

(FISHER'S Z TRANSFORMATION)

Source	S.S.	d.f.	M.S.	F	P
Between Subjects	141.347	26			
CONDITIONS	11.843	2	5.922	1.097	N.S.
Subjects within Groups	129.504	24	5.396		-
Within Subjects	91.000	54			
SESSIONS	15.089	2	7.545	4.991	<.05
CONDITIONS X SESSIONS	3.347	4	.837	< 1	
SESSIONS X Subjects within Groups	72.563	48	1.512		

SIGNIFICANCE OF DIFFERENCES BETWEEN
INDIVIDUAL PAIRS OF SESSION MEANS.
(FISHER'S Z TRANSFORMATION)

	MEAN	SESSION II	SESSION III
MEAN		1.463	2.081
SESSION I	1.029	.434	1.052 *
SESSION II	1.463		.618

The critical value of d = .676 at .05 Level.

* Significant at the .05 Level.

Table 7e

ANALYSIS OF VARIANCE ON THE MATCHING INDEX C

(FISHER'S Z TRANSFORMATION)

Source	S.S.	d.f.	M.S.	F	P
Between Subjects	3.999	26			
CONDITIONS	.272	2	.136	< 1	
Subjects within Groups	3.727	24	.155		
Within Subjects	2.419	54			
SESSIONS	.538	2	.269	7.353	<.01
CONDITIONS X SESSIONS	.126	4	.032	< 1	
SESSIONS X Subjects within Groups	1.755	48	.037		

SIGNIFICANCE OF DIFFERENCES BETWEEN INDIVIDUAL PAIRS OF SESSION MEANS (FISHER'S Z TRANSFORMATION)

	MEAN	SESSION II	SESSION III
MEAN		.211	.309
SESSION I	.109	.102	.200 *
SESSION II	.211		.088

The critical value of d = .105 at the .05 level.

* Significant at the .05 level.

Rank order equivalents of some of the foregoing indices were also computed. Tables 8 (a, b and c) show the rank order correlation of the subjects' certainties with those of the normative model. An analysis of variance (Table 8 d) on these coefficients showed exactly the same pattern of significant differences as the product moment index, with no significant effect of conditions but a higher correlation in session III than in either sessions I or II.

A rank order multiple regression of the subjects' certainties on the two cues X_1 and X_2 was carried out, the resulting correlation coefficients being shown in Table 9 (a, b and c). An analysis of variance on these coefficients showed no effect of conditions but the correlations were significantly lower in session I than in the other two sessions.

A rank order G, i.e. the rank order correlation between the prediction of the best linear combination of cues for predicting the rank order of the subject's certainties and those of the best linear combination for predicting the rank order of the certainties of the normative model, was found and is shown in Tables 10 (a, b and c). The analysis of variance (Table 10 d) on this measure shows no effect of either conditions or sessions.

RANK ORDER CORRELATIONS OF SUBJECTS' CERTAINTIES WITH THOSE OF THE NORMATIVE MODEL

Table 8a Condition I (0.0,0.0)

Subject	Session		
	I	II	III
1	.188	.426	.139
2	.329	.751	.833
3	.777	.726	.875
4	.941	.885	.887
5	.572	.555	.867
6	.819	.521	.931
7	.601	.592	.654
8	.021	.658	.777
9	- .111	.194	.222

Table 8b Condition II (0.7,0.7)

Subject	Session		
	I	II	III
10	.346	.110	.375
11	- .068	.197	.812
12	- .268	- .373	- .072
13	.852	.939	.947
14	- .014	- .000	- .023
15	.173	.017	.197
16	.028	.115	.157
17	.199	.857	.842
18	.055	- .158	.012

RANK ORDER CORRELATIONS OF SUBJECTS' CERTAINTIES
WITH THOSE OF THE NORMATIVE MODEL

Table 8c Condition III (0.9,0.9)

Subject	Session		
	I	II	III
19	.139	.136	- .031
20	.098	.158	.871
21	.229	.158	.870
22	.147	.879	.962
23	- .161	.642	- .043
24	.077	.085	.874
25	.090	.149	.162
26	- .085	.691	.778
27	.722	.334	.534

Table 8d

ANALYSIS OF VARIANCE ON THE RANK ORDER CORRELATIONS OF SUBJECTS' CERTAINTIES WITH THOSE OF THE NORMATIVE MODEL (FISHER'S Z TRANSFORMATION)

Source	S.S.	d.f.	M.S.	F	P
Between Subjects	17.893	26			
CONDITIONS	2.804	2	1.402	2.230	N.S.
Subjects within Groups	15.089	24	.629		
Within Subjects	10.091	54			
SESSIONS	3.099	2	1.549	11.188	<.01
CONDITIONS X SESSIONS	.345	4	.086	<1	
SESSIONS X Subjects within Groups	6.647	48	.138		

SIGNIFICANCE OF DIFFERENCES BETWEEN INDIVIDUAL SESSION MEANS (FISHER'S Z TRANSFORMATION)

		SESSION	
	MEAN	II	III
MEAN		.501	.803
SESSION I	.330	.171	.473 *
SESSION II	.501		.302 *

Critical value of d = .205 at .05 Level.

* Significant at the .05 Level.

THE RANK ORDER CORRELATION BETWEEN SUBJECTS' CERTAINTY ESTIMATES WITH A LINEAR COMBINATION OF CUES WITH WEIGHTS ESTIMATED TO A RANK ORDER CRITERION

Table 9a Condition I (0.0,0.0)

Subject	Session		
	I	II	III
1	.340	.443	.188
2	.340	.861	.874
3	.777	.734	.891
4	.956	.893	.891
5	.610	.558	.873
6	.827	.744	.944
7	.610	.605	.703
8	.305	.671	.784
9	.119	.576	.383
(Model)	(1.000)	(1.000)	(1.000)

Table 9b Condition II (0.7,0.7)

Subject	Session		
	I	II	III
10	.370	.270	.384
11	.306	.213	.890
12	.546	.407	.872
13	.854	.944	.953
14	.107	.034	.419
15	.858	.720	.813
16	.052	.505	.198
17	.213	.868	.853
18	.093	.322	.144
(Model)	(1.000)	(1.000)	(1.000)

THE RANK ORDER CORRELATION BETWEEN SUBJECTS' CERTAINTY ESTIMATES WITH A LINEAR COMBINATION OF CUES WITH WEIGHTS ESTIMATED TO A RANK ORDER CRITERION

Table 9c Condition III (0.9,0.9)

Subject	Session		
	I	II	III
19	.460	.604	.103
20	.568	.909	.879
21	.305	.487	.874
22	.157	.879	.963
23	.383	.650	.739
24	.116	.870	.876
25	.152	.769	.284
26	.524	.691	.783
27	.826	.356	.571
(Model)	(1.000)	(1.000)	(1.000)

ANALYSIS OF VARIANCE ON THE RANK ORDER R_s

(FISHER'S Z TRANSFORMATION)

Source	S.S.	d.f.	M.S.	F	P
Between Subjects	11.908	26			
CONDITIONS	.613	2	.307	<1	
Subjects within Groups	11.295	24	.471		
Within Subjects	9.813	54			
SESSIONS	2.876	2	1.438	10.434	<.01
SESSIONS X CONDITIONS	.321	4	.080	<1	
SESSIONS X Subjects within Groups	6.616	48	.138		

SIGNIFICANCE OF DIFFERENCES BETWEEN
INDIVIDUAL PAIRS OF SESSION MEANS

(FISHER'S Z TRANSFORMATION)

		SESSION	
	MEAN	II	III
MEAN		.823	1.006
SESSION I	.548	.275 *	.458 *
SESSION II	.823		.183

Critical value of d = .204 at .05 Level.

* Significant at the .05 Level.

RANK ORDER MATCHING INDEX G

Table 10a Condition I (0.0,0.0)

Subject	Session		
	I	II	III
1	.398	.975	.916
2	.978	.809	.934
3	1.000	.998	.967
4	.999	.997	.993
5	.975	.981	.992
6	.989	.776	.969
7	.946	.996	.915
8	.305	.987	.980
9	- .997	.563	.953

Table 10b Condition II (0.7,0.7)

Subject	Session		
	I	II	III
10	.991	.725	.999
11	- .189	.996	.927
12	- .431	- .992	- .010
13	.995	.995	.995
14	.702	- .942	.050
15	.275	- .151	.251
16	.232	.070	.621
17	.553	.996	.990
18	.897	- .831	.286

Table 10c Condition III (0.9,0.9)

Subject	Session		
	I	II	III
19	.632	.286	- .854
20	- .030	.111	.987
21	.781	.512	.999
22	.998	1.000	.997
23	- .708	.998	- .067
24	.854	- .179	.995
25	.892	.085	.582
26	- .030	1.000	.999
27	.885	.942	.935

Table 10d

ANALYSIS OF VARIANCE ON THE RANK ORDER MATCHING INDEX G

(FISHER'S Z TRANSFORMATION)

Source	S.S.	d.f.	M.S.	F	P
Between Subjects	120.853	26			
CONDITIONS	13.688	2	6.844	1.533	N.S.
Subjects within Groups	107.165	24	4.465		
Within Subjects	97.924	54			
SESSIONS	5.630	2	2.815	1.538	N.S.
CONDITIONS X SESSIONS	4.435	4	1.109	<1	
SESSIONS X Subjects within Groups	87.858	48	1.830		

EXPERIMENT I

Conclusions

The most all pervading aspect of the data collected in this experiment is that a large proportion of subjects never learned to use the cues in a valid manner. Two subjects in Condition I (0.0,0.0), five in Condition II (0.7,0.7) and three in Condition III (0.9,0.9) gave certainty estimates whose correlation with those of the normative model did not exceed what might be expected by chance, sampling from a population with zero correlation, throughout the three sessions. An almost identical pattern occurs when the number of correct responses of each subject is compared with the number expected from a chance process which randomly allocates each "patient" to either of the disease populations with equal probability. The most likely explanation of this poor performance is a lack of motivation amongst the subjects who were not volunteers but were participating in the experiment in order to fulfil a course requirement. The remainder of the subjects appeared to be operating at a higher level than would be expected by chance and some obtained quite high correlations with the certainties of the normative model and made large numbers of correct "diagnoses".

In view of these large differences in performance between the subjects in each group it is hardly surprising that no significant effect of conditions was found on any of the accuracy indices (total number correct and the product moment and rank order correlations of the certainties of the subjects with those of the normative model). These indices did however, increase significantly over sessions with more accurate performance in session III than in the previous sessions, showing that some learning did take place in the course of the experiment, though as we have seen some subjects did not improve at all, nor reach a level of performance higher than that expected by chance.

The multiple correlation of the subjects' certainties with the two cues X_1 and X_2 also showed an increase over sessions but no effect of conditions. The certainties of a number of subjects in each condition (in general those who were not performing above a chance level) were not correlated significantly with the two cues, though those of other subjects were highly related to a linear combination of the cues. Most subjects with high multiple correlations with the two cues also had a high correlation between their certainties and a linear combination of the two cues which weighted them equally (i.e. the sum of X_1 and X_2). One or two subjects however, notably Subject 9 condition I (0.0, 0.0), Subject 12 condition II (0.7,0.7) and Subject 20 condition III (0.9,0.9) have in some sessions high values of R_s but low values of correlations with $(X_1 + X_2)$ and with the certainties of the normative model. The difference in the multiple correlation and the correlation with $(X_1 + X_2)$ resulting in high and significant F ratios. This is easily explained if the beta weights (not presented) given to the two cues are examined. The certainty estimates of all these subjects had significant correlations with a linear combination which assigned a positive weight to one cue and a negative weight to the other. These subjects seem to have subtracted one cue from the other (the weights were not in general significantly different from one another in absolute magnitude) and to have based their certainty estimates and decisions on the result. Although this process is linear, it has little or no validity thus explaining their poor performance.

The certainties of those subjects who were performing above chance level could generally be predicted almost as well by a linear combination of the two cues with equal weights as by a linear combination with weights estimated from the data. In practically all cases, if one of the cues contributed significantly to the fit of the linear combination (with estimated weights) then the other cue also contributed significantly. There is no evidence to show that subjects used only one cue, or that they used one cue more than the other in making their certainty estimates.

The indices of the Lens Model do not help us very much in our investigation of subjects' behaviour in such tasks as these. The GR_eR_s and $C\sqrt{(1 - R_e^2)(1 - R_s^2)}$ terms cannot be meaningfully compared as the linear predictability of the task R_e varies from condition to condition (and session to session.) Both the G and C matching indices show no significant differences between conditions though both increase over sessions. One drawback of the Lens Model analysis is that it assumes that the subjects' estimates are of exactly the same variable as the one predicted by the ecological model. In the current experiment subjects' certainty responses were regarded as being exactly comparable to those of the normative model given by :-

$$\text{Certainty of Disease 1 at } (x_1,x_2)$$
$$= \frac{\text{Prob. of Disease 1 at } (x_1,x_2)}{\text{Prob. of Disease 1 at } (x_1,x_2) + \text{Prob. of Disease 2 at } (x_1,x_2)} \times 100$$

Subjects, however, were given no feedback as to the accuracy of their certainty estimates and though it is probable that their estimates are monotonic with their subjective certainty (e.g. their estimates might represent simply a constant times their subjective probability of Disease 1, or a constant times the log likelihood ratio of their subjective probabilities etc.) there is no guarantee that they are estimating what we hope they are. It is possible that the interval properties required of the subjects' responses by the Lens Model may not exist, and thus that any apparent non-linearity in their decision processes may be a result of an inappropriate response scale.

An examination of R_e (the linear predictability of the system) shows that some of the tasks are apparently more linearly dependent than others. Yet we know that an equal weighting of the two cues can predict perfectly the log likelihood ratio, which is monotonic with certainty. Thus when a rank order multiple correlation procedure was applied to the certainties of the normative model, by changing the beta weights found by least squares so as to minimise the difference in rank order of the predicted certainties of the normative model, and its actual certainties, weights could be found which made the rank order correlation perfect in all conditions i.e. all the tasks were equally well predicted by a linear combination of the cues, to a rank order criterion. It is only when the supposed interval qualities of the certainty data are used that R_e drops below unity.

The rank order procedure was also applied to the certainties of the subjects, changing the weightings of the two cues X_1 and X_2 so as to minimise the differences in rank order of the predictions of this linear combination and the rank order of the subject's certainties. The resulting correlation coefficients, however, do not differ greatly from their product moment equivalents and on applying analysis of variance are found to have a very similar pattern of significance. Similar results were found with a rank order matching index G as were found with the product moment G, but when we come to look at the index of non-linear matching, C, the two approaches lead to quite different conclusions.

The value of C must be zero for all conditions in the present task, if we are working with the rank order version of the Lens Model. Since the rank order of the certainties of the normative model is perfectly predictable from a linear combination of the two cues, there is nothing left to be explained by non-linear cue utilization, i.e. nothing with which to correlate that part of the subjects' responses which cannot be put down to using the cues linearly, C therefore must be zero.

These tasks then are not linear if our criterion of linearity is that

$$Y_e = \beta_1 X_1 + \beta_2 X_2$$

However, if

$$Y_e \neq \beta_1 X_1 + \beta_2 X_2$$

but $$Z = \beta_1 X_1 + \beta_2 X_2$$

and Y_e and Z have the same rank order i.e. are monotonic with one another, then the tasks may be regarded as being linear to a rank order criterion, since a linear combination of the two cues, though it cannot predict the <u>values</u> of Y_e exactly can predict the rank order of the values without error.

The product moment version of the Lens Model is unable to distinguish between decision processes which are structurally non-linear (i.e. include non-linear terms) and those which appear non-linear because the response scale is not a linear function of the cues, but is just monotonic with such a function. Thus, though many of the C values of subjects in this experiment are quite large, this may be due to peculiarities of their response scale rather than non-linear decision processes.

In this task, if subjects had been asked to state their subjective log likelihood ratio (since the log likelihood ratio estimates of the normative model are perfectly predictable by a linear combination of the cues to a least squares criterion) then C would have been zero in all tasks even with the product moment conception of the Lens Model. However, we may ask what "subjective log likelihood ratio" means to the average Part I psychology student - the answer is probably "absolutely nothing", so such an approach is not likely to bear fruit. Nor is the transformation of subjects' estimates on some other scale to "subjective log likelihood" ratio likely to be much use, since there is no guarantee that subjects are making estimates on the scale we think they are, nor that the transformed data have the necessary interval properties for the analysis.

It is felt then that this rank order approach may be of some use, particularly in comparing data from tasks which may be of different levels of predictability to a least squares criterion but which are equally predictable to a rank order one. In the present tasks, though, the rank order indices of linear predictability of subjects' responses and of linear matching do not differ markedly from their product moment counterparts (either in magnitude or in terms of significant effects of conditions or sessions), the rank order index of non-linear matching must clearly be zero in all conditions whereas it may be other than zero under the product moment conception.

In summary then, many of the subjects never learned to use the cues in an appropriate way, though some of these were using the cues consistently in an inappropriate way i.e. they appeared to have learnt an invalid rule and not to have been able to discard it, in a similar manner to some of the subjects of Azuma and Cronbach (1966). Of those subjects who did learn to use the cues appropriately there was strong reason to believe that both cues were weighted equally, the subjects probably basing their decisions and certainty responses on the sum or average of the two cues. Cue intercorrelation was not found to have a significant effect on either performance indices or on any of the indices of matching or linear dependence. This could either be due to the great error variance resulting from the poor performance of some subjects or it might be that cue intercorrelation has not such a strong moderating effect as was felt by Schenck and Naylor (1968).

EXPERIMENT II

DECISIONS BETWEEN BIVARIATE NORMAL POPULATIONS WITH UNEQUAL MEAN VECTORS AND UNEQUAL COVARIANCE MATRICES

INTRODUCTION TO EXPERIMENT II

We have seen how clinicians claim to use cues in complex non-linear ways, how they claim that it is the "pattern" or relationship between scores on different tests which they feel to be important, not just the scores on each test viewed singly and independently. Their claims seem to imply that they perceive differences in the inter-relations of the cues between the various diagnostic groups. We might feel it surprising then that only one paper has dealt with the behaviour of subjects in a learning task so designed as to create different relationships in the cues from one diagnostic alternative to another.

Schum (1966) states "One issue not investigated in previous studies but of considerable importance in an evaluation of human inferential skills concerns interdependencies among data upon which inferences are to be based. Part of the task of weighing evidence for inferential or diagnostic purposes includes appraisal of the joint impact of various items of evidence being considered. Frequently, knowledge of the form, extent, and probable cause of inter-relationships among two or more items of evidence makes the joint impact of the evidence different from the total impact from successive but independent consideration of each item. In some cases the dependence, or more appropriately non-independence, of two or more items of data is conditional upon the truth of some hypothesis (H_i) being considered. In other words the factor or process mediating the non-independence is the truth of one of the hypotheses being entertained to explain the occurrence of the data." (pp. 401 - 402).

He placed subjects in a simulated "threat diagnosis context, and required them to estimate, on the basis of a set of data ("scenario") about a potential enemy's troop movements, the probability that an attack rather than a "peaceful" exercise was being mounted. During the experiment 240 scenarios were presented to each subject, each scenario consisting of six items of information (two of which had very little validity). The subject made his estimate of the probability of attack for each of the scenarios and was immediately informed of the correct hypothesis. the six variables listed in each scenario were of a categorical nature, taking one of only three or four possible values. For example, data class one (D_1) was the number of "heavy tank units" in the vicinity, and took one of the values 1, 2, 3 or 4.

Three groups of subjects were used, one group under each of three conditions. For groups 1 and 2 the data sources were conditionally non-independent, though they differed in the exact form of non-independence. For group 3, all data sources were conditionally independent. For groups 1 and 2, when H_a (attack) was true, data classes D_3 and D_4 exhibited interdependence and when H_b (no attack) was true data classes D_1 and D_2 exhibited interdependence. For group 3 all six data classes were independent regardless of which hypothesis was true. The figure below shows an extreme case (not investigated by Schum) where neither D_1 nor D_2 used singly has any predictive validity but used together they increase accuracy quite substantially. (It is hoped that this extreme example will clearly illustrate what is meant by conditional non-independence).

$\underline{H_a}$

D_2 \ D_1	1	2	3	4	
1	.10	.00	.00	.00	.10
2	.00	.10	.00	.00	.10
3	.00	.00	.20	.00	.20
4	.00	.00	.00	.60	.60
	.10	.10	.20	.60	1.00

$\underline{H_b}$

D_2 \ D_1	1	2	3	4	
1	.01	.01	.02	.06	.10
2	.01	.01	.02	.06	.10
3	.02	.02	.04	.12	.20
4	.06	.06	.12	.36	.60
	.10	.10	.20	.60	1.00

In the above example, the knowledge that D_1 takes the value 1 does not help us to distinguish between the two states of nature. Similarly, the knowledge that D_2 takes the value 1 helps us not at all in making the decision. However the knowledge that D_1 takes the value 1 <u>and</u> D_2 takes the value 1 makes H_a ten times as likely as H_b.

Subjects' estimates of probabilities were compared with the actual probabilities calculated through Bayes theorem, and Schum was able to conclude that the group 1 subjects with conditionally non-independent cues made estimates of the probabilities of "attack" and "no attack" which were very close to the actual values. The subjects of both groups 1 and 2 were found not to have been processing the non-independent data as if they were independent, implying that they had perceived and utilized the conditional non-independence of the cues.

It is questionable if similar results would have been obtained if the subjects had not been required to keep written records and if their attention had not been drawn to the likely locus of any non-independence in the data sources. Hammond and Summers (1965) had found subjects not to utilize a non-linear cue if they were not informed of its non-linearity, it is possible that similar results might hold for conditional non-independence.

This last experiment dealt with conditional non-independence in categorical cues, what sort of conditional non-independence might occur between continuous cues in the sort of situation with which Experiment I dealt? One obvious way in which the two populations might differ in cue interdependence, is in the degree of correlation existing between the two predictor variables. (In Experiment I, though cue correlation differed from condition to condition, it was equal for both populations in any one condition).

The normative model in Experiment I was provided by Linear Discriminant Function analysis, but this procedure is based on the very assumption that the two populations to be discriminated are of equal covariance matrix i.e. the same cue intercorrelations exists in both. It is possible however to deduce the best decision function in this situation (for a detailed analysis see Appendix C), and this is found to be a linear (in the statistical sense), combination of not only the scores on the variables, but also of the squares and cross products of the scores. The log likelihood ratio is given by a function of the

It is questionable if similar results would have been obtained if the subjects had not been required to keep written records and if their attention had not been drawn to the likely locus of any non-independence in the data sources. Hammond and Summers (1965) had found subjects not to utilize a non-linear cue if they were not informed of its non-linearity, it is possible that similar results might hold for conditional non-independence.

This last experiment dealt with conditional non-independence in categorical cues, what sort of conditional non-independence might occur between continuous cues in the sort of situation with which Experiment I dealt? One obvious way in which the two populations might differ in cue interdependence, is in the degree of <u>correlation</u> existing between the two predictor variables. (In Experiment I, though cue correlation differed from condition to condition, it was equal for both populations in any one condition).

The normative model in Experiment I was provided by Linear Discriminant Function analysis, but this procedure is based on the very assumption that the two populations to be discriminated are of equal covariance matrix i.e. the same cue intercorrelations exists in both. It is possible however to deduce the best decision function in this situation (for a detailed analysis see Appendix C), and this is found to be a linear (in the statistical sense), combination of not only the scores on the variables, but also of the squares and cross products of the scores. The log likelihood ratio is given by a function of the form:-

$$\ln L(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1^2 + \beta_4 X_2^2 + \beta_5 X_1 X_2$$

If more than two cues are involved, similar terms in their squares and cross product must be included, the function remaining throughout, what is termed a "Quadratic Function".

The above function may look familiar, we did in fact come across it when reviewing Wiggins and Hoffman's (1968) paper in which they used it to model the behaviour of subjects asked to estimate the probability of M.M.P.I. profiles being from psychotic rather than neurotic patients. Wiggins and Hoffman, however, felt it was "simply a mathematical expression involving all possible first and second order terms" and contrasted it with the sign model which "involves rationally chosen variables" (p. 72). It appears however, that the model is not quite so arbitrary as was felt, if decisions must be made between multivariate normal populations possibly differing in cue independence (and the diagnosis of psychotics vs. neurotic profiles could certainly come under this category) then this simple expression follows as logically as does the Linear Discriminant Function under slightly differing circumstances.

Brunswick's plea for representative experimental design again stimulates us to be interested in situations of the above kind. Researchers in "real life"decision making behaviour went to great pains to discover naturally occurring tasks which required the non-linear use of cues for accurate performance. But, having found tasks, reportedly non-linear in nature, they did not investigate these tasks statistically to discover if such claims were warranted and, if so, what non-linear processes were valid in these situation. Rather, they frequently altered the situations to suit their research tools, and possibly as a result of this, were unable to find any indication of a great reliance on anything other than linear cues. If, however, non-linear tasks are constructed in the laboratory, all the parameters of the populations are known to the experimenter, he knows which cues are valid and which are not and he also has complete control of subjects' experiences.

In the literature on learning multiple cue relations, only Hammond and Summers (1965) investigated behaviour in a non-linear task, including as it did a dependence between the predicted variable and the size of one of the predictors.

Hammond and Summers, however, had no particular reason for choosing a non-linear term of this kind, it was not felt to be particularly representative of some certain kind of real life task, it was however mathematically convenient having a zero correlation with the "raw" predictor variable. The experiment to be described, on the contrary, was designed to represent a class of decision making tasks, requiring non-linear cue utilization, which could easily occur in real life. The non-linearity of the situation is not something artificial which is added on to a linear component, but is an integral part of the task following logically from its underlying structure.

The experiment consisted of three different conditions varying in the difference between the two populations to be distinguished in terms of cue intercorrelation. In all conditions, the mean value of population 2 was 9.5 on both variables, with standard deviation 2.0 and intercorrelation zero. The mean value of population 1 was 7.5 on both variables, with standard deviation 2.0, but with zero intercorrelation in condition I, and intercorrelation of + .70 in condition II and +.90 in condition III. (The tasks are discussed in greater detail in Appendix C.)

A pilot study has revealed that with high differences in correlation, subjects claimed that they look at a transformation of the two cues which may be represented mathematically as ABS $(X_1 - X_2)$ i.e. the magnitude of the difference between the two cues, disregarding the sign of the difference. This is a non-linear combination of X_1 and X_2 and it was intended to consider the conditions in terms of the correlations between this cue and the subjects' certainty estimates as well as the linear model examined in the last experiment.

The experiment was designed in order to discover whether subjects modify their decision strategies so as to take account of differences in cue intercorrelation between the two populations to be discriminated, or whether they persist in their use of a linear combination of the two cues as they appeared to do in Experiment I. A large number of conditions could not be run, but it was hoped that the three investigated here would give some indication of the level of difference in correlation which is necessary to cause subjects to discard the simple linear model.

It was known from the experiments on the ability of subjects to estimate one correlated variable from another that subjects can perceive quite accurately the high range of positive correlations, but since the prime aim of the task in this situation is not to predict one cue from the other, but to predict some other criterion from both, it is possible that even quite high correlations might be overlooked.

Subjects

Due to the relative lack of success, in using as subjects Part 1 psychology students who were compelled to participate in the first experiment, it was felt necessary to find subjects whose motivation could be relied upon. A sufficient degree of motivation was assured by the use of subjects who were known personally to the experimenter and who were paid for their participation (at the rate of 30p per hour). This however, severely reduced the pool of possible subjects and it was decided to use a within subjects design, each subject taking part for three hours under all three conditions. Six female students, who were either honours students or postgraduates in psychology acted as subjects. Their average age was approximately 21 years. All possible six orders of presentation of the three conditions were used, one subject under each order.

Apparatus

The same apparatus as was employed in Experiment I was used in this study, though the paper tapes on which were punched the stimuli were obviously different.

Procedure

This was the same as in Experiment I, but after the subject had completed the three sessions of each condition, she was told that the next session would be different and that the rules she had so far developed may not be valid in the next session. At the end of the third session of each condition, the subject was asked to write a few sentences on the way in which the categorical decision and certainty estimates were made.

EXPERIMENT II

RESULTS

Again only the data collected in the 50 test trials of each session were analysed, with most of the analyses based on the certainty estimates again transformed to yield a scale ranging from 1 (complete certainty of Disease 1) to 100 (complete certainty of Disease 2).

The number of correct responses (i.e. the number of categorical responses which agreed with those of the normative model) was found for each subject in each session and are presented in Tables 11 (a, b and c). No significant differences occurred between conditions but performance in session II was significantly better than in the other two sessions (Table 11 d). The correlation of the subjects' certainties withthose of the normative model, displayed in Tables 12 (a, b and c), were found not to be affected significantly by either conditions or sessions (Table 12 d).

NUMBER OF CORRECT RESPONSES

(i.e. Agreements with the Normative Model)

Table 11a Condition I (0.0,0.0)

Subject	Session		
	I	II	III
1	49	48	48
2	40	47	40
3	43	48	46
4	45	46	41
5	42	43	42
6	44	41	41

Table 11b Condition II (0.7,0.0)

Subject	Session		
	I	II	III
1	43	37	39
2	42	42	43
3	46	46	42
4	34	45	36
5	43	47	45
6	40	44	35

Table 11c Condition III (0.9,0.0)

Subject	Session		
	I	II	III
1	45	46	42
2	44	49	40
3	44	45	42
4	45	48	40
5	44	49	44
6	40	40	45

Table 11b

ANALYSIS OF VARIANCE ON THE NUMBER OF CORRECT RESPONSES

Source	S.S.	d.f.	M.S.	F	P
Subjects	86.53	5			
Sessions	102.37	2	51.19	7.213	.05
Conditions	71.81	2	35.91	2.370	N.S.
Sessions x Conditions	3.52	4	.88	1	
Sessions x Subjects	70.96	10	7.10		
Conditions x Subjects	151.52	10	15.15		
Sessions x Conditions x Subjects	157.15	20	7.86		
TOTAL	643.86	53			

SIGNIFICANCE OF DIFFERENCES BETWEEN INDIVIDUAL PAIRS OF SESSION MEANS

		SESSION	
	MEAN	II	III
MEAN		45.06	41.72
SESSION			
I	42.94	2.12*	-1.22
II	45.06		-3.34*

Critical value of d = .198 at the .05 level.

* Significant at the .05 level.

CORRELATIONS OF SUBJECTS' CERTAINTIES

WITH THOSE OF THE NORMATIVE MODEL (r_a)

Table 12a Condition I (0.0,0.0)

Subject	Session		
	I	II	III
1	.898	.923	.842
2	.841	.937	.860
3	.911	.942	.935
4	.883	.884	.855
5	.855	.914	.864
6	.811	.927	.922

Table 12b Condition II (0.7,0.0)

Subject	Session		
	I	II	III
1	.748	.706	.758
2	.818	.838	.768
3	.922	.901	.915
4	.574	.816	.757
5	.902	.923	.904
6	.862	.879	.687

Table 12c Condition III (0.9,0.0)

Subject	Session		
	I	II	III
1	.869	.816	.795
2	.823	.909	.804
3	.779	.811	.888
4	.930	.896	.852
5	.813	.883	.915
6	.708	.674	.827

Table 12d

ANALYSIS OF VARIANCE ON THE CORRELATIONS OF SUBJECTS' CERTAINTIES WITH THOSE OF THE NORMATIVE MODEL
(Fisher's Z Transformation)

Source	S.S.	d.f.	M.S.	F	P
Subjects	.750	5			
Sessions	.240	2	.120		
Conditions	.550	2	.275	3.012	N.S.
Sessions x Conditions	.147	4	.037	3.230	N.S.
Sessions x Subjects	.398	10	.040	<1	
Conditions x Subjects	.851	10	.085		
Sessions x Conditions x Subjects	1.148	20	.057		
TOTAL	4.084	53			

Table 12d

ANALYSIS OF VARIANCE ON THE CORRELATIONS OF SUBJECTS' CERTAINTIES WITH THOSE OF THE NORMATIVE MODEL

(Fisher's Z Transformation)

Source	S.S.	d.f.	M.S.	F	P
Subjects	.750	5			
Sessions	.240	2	.120		
Conditions	.550	2	.275	3.012	N.S.
Sessions x Conditions	.147	4	.037	3.230	N.S.
Sessions x Subjects	.398	10	.040	<1	
Conditions x Subjects	.851	10	.085		
Sessions x Conditions x Subjects	1.148	20	.057		
TOTAL	4.084	53			

Multiple regression analyses were carried out upon the subjects' certainties using the two cues X_1 and X_2 as predictor variables. The resulting multiple correlation coefficients (R_s's) are shown in Tables 13 (a, b and c). An analysis of variance on these measures (Table 13 d) showed that subjects' certainties were less well predicted by such a linear combination in Condition III (0.9,0.0) than in the other two conditions. There was also a significant effect of sessions, the coefficients being significantly lower in session III than session II.

The correlation of subjects' certainties with the sum of the two cues i.e. with $(X_1 + X_2)$ are shown in Tables 14 (a, b and c). An analysis of variance on these coefficients (Table 14 d) showed a significantly lower correlation in Condition III (0.9,0.0) than in the other conditions as well as a lower correlation in session III than in the two previous sessions. When the possible covariance effects of the term ABS $(X_1 - X_2)$ were removed from these correlation coefficients, an analysis of variance on the resulting partial correlations (Table 14 e) showed no effect of sessions though these partial correlations were also lower in Condition III (0.9,0.0) than in the other conditions. F ratios to test the significance of the better fit of the linear combination of the two cues X_1 and X_2 with estimated weights over that of the term $(X_1 + X_2)$ were calculated and are shown in Tables 15 (a, b and c).

The Lens Model indices for a linear combination of the two cues (with estimated weights) were computed for each subject. G is shown in Tables 16 (a (i), b (i) and c (i)), GR_eR_s in Tables 16 (a (ii), b (ii) and c (ii)), C in Tables 16 (a (iii), b (iii) and c (iii)) and $C\sqrt{(1 - R_e^2)(1 - R_s^2)}$ in Tables 16 (a (iv), b (iv) and c (iv)). Analyses of variance were carried out on both the G and C indices. Though G was not significantly affected by either condition or session (Table 16d), C was found to be significantly higher in Condition III (0.9,0.0) than in the other two conditions, and was significantly lower in session I than in sessions II and III (Table 16e).

PREDICTABILITY OF SUBJECT'S CERTAINTIES FROM A LINEAR COMBINATION OF THE TWO CUES WITH ESTIMATED WEIGHTS (R_s)

Table 13a Condition I (0.0,0.0)

Subject	Session		
	I	II	III
1	.847	.840	.739
2	.845	.918	.862
3	.908	.917	.917
4	.853	.862	.805
5	.844	.868	.836
6	.813	.919	.933
(Model)	(.985)	(.965)	(.968)

Table 13b Condition II (0.7,0.0)

Subject	Session		
	I	II	III
1	.824	.850	.699
2	.903	.903	.926
3	.880	.890	.881
4	.668	.847	.848
5	.893	.860	.839
6	.894	.916	.905
(Model)	(.868)	(.862)	(.772)

Table 13c Condition III (0.9,0.0)

Subject	Session		
	I	II	III
1	.743	.820	.756
2	.781	.852	.838
3	.890	.892	.728
4	.754	.755	.674
5	.848	.778	.646
6	.959	.929	.495
(Model)	(.709)	(.711)	(.570)

Table 13d

ANALYSIS OF VARIANCE ON THE R_ss OF A LINEAR

COMBINATION OF THE CUES

(Fisher's Z Transformation)

Source	S.S.	d.f.	M.S.	F	P
Subjects	1.161	5	.232		
Sessions	.336	2	.168	7.229	<.01
Conditions	.470	2	.235	13.143	<.01
Sessions x Conditions	.365	4	.091	1.692	N.S.
Sessions x Subjects	.233	10	.023		
Conditions x Subjects	.179	10	.018		
Sessions x Conditions x Subjects	1.078	20	.054		
TOTAL	3.822	53			

SIGNIFICANCE OF DIFFERENCES BETWEEN

INDIVIDUAL PAIRS OF CONDITION MEANS

(Fisher's Z Transformation)

	MEAN	CONDITION II (0.7,0.0)	CONDITION III (0.9,0.0)
MEAN		1.323	1.132
CONDITION			
I (0.0,0.0)	1.335	-.012	-.203*
II(0.7,0.0)	1.323		-.191*

Critical value of d = .099 at the .05 level.

SIGNIFICANCE OF DIFFERENCES BETWEEN

INDIVIDUAL PAIRS OF SESSION MEANS

(Fisher's Z Transformation)

	MEAN	SESSION II	SESSION III
MEAN		1.354	1.162
SESSION			
I	1.275	.079	-.113
II	1.354		-.194*

Critical value of d = .160 at the .05 level.

* Significant at the .05 level.

THE CORRELATIONS OF SUBJECTS' CERTAINTIES WITH THE TERM ($X_1 + X_2$)

Table 14a Condition I (0.0,0.0)

Subject	Session		
	I	II	III
1	-.847	-.838	-.738
2	-.837	-.912	-.851
3	-.905	-.910	-.914
4	-.852	-.861	-.797
5	-.834	-.867	-.824
6	-.811	-.918	-.933
(Model)	(-.985)	(-.964)	(-.968)

Table 14b Condition II (0.7,0.0)

Subject	Session		
	I	II	III
1	-.815	-.760	-.581
2	-.867	-.865	-.921
3	-.879	-.883	-.865
4	-.662	-.839	-.846
5	-.892	-.848	-.834
6	-.892	-.913	-.904
(Model)	(-.868)	(-.855)	(-.750)

Table 14c Condition III (0.9,0.0)

Subject	Session		
	I	II	III
1	-.699	-.817	-.756
2	-.781	-.847	-.836
3	-.889	-.889	-.701
4	-.752	-.733	-.646
5	-.847	-.763	-.636
6	-.959	-.917	-.488
(Model)	(-.709)	(-.687)	(-.556)

Table 14d

ANALYSIS OF VARIANCE ON THE CORRELATIONS OF SUBJECTS' CERTAINTIES WITH THE TERM ($X_1 + X_2$)

(Fisher's Z Transformation)

Source	S.S.	d.f.	M.S.	F	P
Subjects	1.260	5			
Sessions	.295	2	.148	5.704	<.05
Conditions	.431	2	.215	11.249	<.01
Sessions x Conditions	.375	4	.094	1.564	N.S.
Sessions x Subjects	.259	10	.026		
Conditions x Subjects	.191	10	.019		
Sessions x Conditions x Subjects	1.199	20	.060		
TOTAL	4.010	53			

SIGNIFICANCE OF DIFFERENCES BETWEEN INDIVIDUAL PAIRS OF CONDITION MEANS (Fisher's Z Transformation)

	MEAN	CONDITION II (0.7,0.0)	CONDITION III (0.9,0.0)
MEAN		1.262	1.107
CONDITION			
I (0.0,0.0)	1.318	-.054	-.211*
II (0.7,0.0)	1.262		-.155*

Critical value of d = .103 at the .05 level.

SIGNIFICANCE OF DIFFERENCES BETWEEN INDIVIDUAL PAIRS OF SESSION MEANS (Fisher's Z Transformation)

	MEAN	SESSION II	SESSION III
MEAN		1.306	1.129
SESSION			
I	1.252	.054	-.123*
II	1.306		-.177*

Critical value of d = .119 at the .05 level.

* Significant at the .05 level.

Table 14e

ANALYSIS OF VARIANCE ON THE CORRELATIONS OF SUBJECTS' CERTAINTIES WITH THE TERM ($X_1 + X_2$) WHEN THE EFFECT OF THE TERM ABS($X_1 - X_2$) IS PARTIALLED OUT (Fisher's Z Transformation)

Source	S.S.	d.f.	M.S.	F	P
Subjects	1.319	5			
Sessions	.130	2	.065	2.451	N.S.
Conditions	.298	2	.149	6.324	<.05
Sessions x Conditions	.302	4	.076	1.292	N.S.
Sessions x Subjects	.265	10	.027		
Conditions x Subjects	.236	10	.024		
Sessions x Conditions x Subjects	1.168	20	.058		
TOTAL	3.718	53			

SIGNIFICANCE OF DIFFERENCES BETWEEN INDIVIDUAL PAIRS OF CONDITION MEANS (Fisher's Z Transformation)

	MEAN	CONDITION II (0.7,0.0)	CONDITION III (0.9,0.0)
MEAN		1.330	1.191
CONDITION			
I (0.0,0.0)	1.362	-.032	-.171*
II(0.7,0.0)	1.330		-.139*

Critical value of d = .114 at the .05 level.

* Significant at the .05 level.

F RATIOS TO TEST THE SIGNIFICANCE OF THE BETTER FIT OF THE LINEAR COMBINATION OF THE CUES WITH ESTIMATED WEIGHTS OVER THAT OF THE TERM $(X_1 + X_2)$

Table 15a Condition I (0.0,0.0)

Subject	Session		
	I	II	III
1	.14	.46	.08
2	2.41	3.12	3.34
3	1.65	4.01	1.45
4	.32	.28	1.82
5	2.66	.26	3.03
6	.41	.39	.27
(Model)	(.62)	(1.19)	(.00)

Table 15b Condition II (0.7,0.0)

Subject	Session		
	I	II	III
1	2.25	24.25*	13.90*
2	15.92*	17.54*	2.99
3	.29	2.72	5.59*
4	.71	2.19	.63
5	.08	3.72	1.19
6	.88	1.50	.71
(Model)	(.10)	(2.39)	(3.93)

Table 15c Condition III (0.9,0.0)

Subject	Session		
	I	II	III
1	6.58*	.89	.03
2	.00	1.48	.42
3	.52	.94	3.79
4	.36	3.60	3.18
5	.23	2.76	1.00
6	.00	7.94*	.38
(Model)	(.07)	(3.25)	(1.11)

* Significant at the .05 level.

THE LENS MODEL INDICES FOR A

LINEAR COMBINATION OF THE TWO CUES

Table 16a (i) Condition I (0.0,0.0)

THE MATCHING INDEX G

Subject	Session		
	I	II	III
1	1.000	.994	1.000
2	.986	.988	.988
3	.998	.996	.997
4	.997	.996	.989
5	.985	1.000	.985
6	.999	1.000	.999

Table 16a (ii) Condition I (0.0,0.0)

GR_eR_s

Subject	Session		
	I	II	III
1	.834	.806	.714
2	.821	.876	.824
3	.893	.882	.884
4	.838	.829	.770
5	.819	.837	.797
6	.800	.887	.902

THE LENS MODEL INDICES FOR A LINEAR COMBINATION OF THE TWO CUES

Table 16a (iii) Condition I (0.0,0.0)

THE MATCHING INDEX C

Subject	Session		
	I	II	III
1	.694	.820	.751
2	.219	.594	.279
3	.258	.574	.506
4	.502	.419	.562
5	.394	.589	.486
6	.103	.393	.212

Table 16a (iv) Condition I (0.0,0.0)

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

Subject	Session		
	I	II	III
1	.064	.117	.128
2	.020	.062	.036
3	.019	.060	.051
4	.045	.056	.084
5	.037	.077	.067
6	.010	.041	.019

THE LENS MODEL INDICES FOR A

LINEAR COMBINATION OF THE TWO CUES

Table 16b (i) Condition II (0.7,0.0)

THE MATCHING INDEX G

Subject	Session		
	I	II	III
1	.985	.828	.940
2	.967	.987	.942
3	1.000	1.000	.999
4	.987	1.000	.986
5	1.000	.999	.991
6	.996	.978	.983

Table 16b (ii) Condition II (0.7,0.0)

GR_eR_s

Subject	Session		
	I	II	III
1	.704	.607	.507
2	.757	.769	.673
3	.764	.767	.679
4	.572	.731	.646
5	.775	.741	.641
6	.773	.772	.687

THE LENS MODEL INDICES FOR A

LINEAR COMBINATION OF THE TWO CUES

Table 16b (iii) Condition II (0.7,0.0)

THE MATCHING INDEX C

Subject	Session		
	I	II	III
1	.157	.370	.552
2	.285	.317	.396
3	.669	.579	.786
4	.006	.318	.330
5	.568	.705	.760
6	.400	.525	-.001

Table 16b (iv) Condition II (0.7,0.0)

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

Subject	Session		
	I	II	III
1	.044	.099	.251
2	.061	.069	.095
3	.158	.134	.095
4	.002	.086	.111
5	.127	.182	.263
6	.089	.107	-.000

THE LENS MODEL INDICES FOR A

LINEAR COMBINATION OF THE TWO CUES

Table 16c (i) Condition III (0.9,0.0)

THE MATCHING INDEX G

Subject	Session		
	I	II	III
1	.955	.986	.971
2	.999	.988	.987
3	1.000	.982	.999
4	1.000	1.000	.998
5	1.000	.998	.999
6	.999	.910	.998

Table 16c (ii) Condition III (0.9,0.0)

GR_eR_s

Subject	Session		
	I	II	III
1	.503	.576	.419
2	.553	.599	.471
3	.631	.623	.414
4	.534	.537	.383
5	.601	.552	.368
6	.680	.602	.281

THE LENS MODEL INDICES FOR A

LINEAR COMBINATION OF THE TWO CUES

Table 16c (iii) Condition III (0.9,0.0)

THE MATCHING INDEX C

Subject	Session		
	I	II	III
1	.776	.597	.699
2	.612	.841	.741
3	.459	.591	.841
4	.854	.778	.772
5	.567	.749	.873
6	.140	.280	.763

Table 16c (iv) Condition III (0.9,0.0)

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

Subject	Session		
	I	II	III
1	.366	.240	.376
2	.269	.310	.332
3	.147	.188	.474
4	.395	.358	.469
5	.212	.331	.547
6	.028	.073	.545

Table 16d

ANALYSIS OF VARIANCE ON THE MATCHING INDEX G FOR THE LINEAR COMBINATION OF X_1 AND X_2 (Fisher's Z Transformation)

Source	S.S.	d.f.	M.S.	F	P
Subjects	9.168	5			
Sessions	1.532	2	.766	1.897	N.S.
Conditions	1.658	2	.829	<1	
Sessions x Conditions	2.304	4	.576	1.732	N.S.
Sessions x Subjects	4.039	10	.404		
Conditions x Subjects	12.284	10	1.228		
Sessions x Conditions x Subjects	6.652	20	.333		
TOTAL	37.637	53			

Table 16e

ANALYSIS OF VARIANCE ON THE MATCHING INDEX C FOR THE LINEAR COMBINATION OF X_1 AND X_2 (Fisher's Z Transformation)

Source	S.S.	d.f.	M.S.	F	P
Subjects	1.142	5			
Sessions	.487	2	.243	7.677	<.01
Conditions	1.489	2	.745	4.531	<.05
Sessions x Conditions	.240	4	.060	1.181	N.S.
Sessions x Subjects	.317	10	.032		
Conditions x Subjects	1.644	10	.164		
Sessions x Conditions x Subjects	1.017	20	.051		
TOTAL	6.336	53			

SIGNIFICANCE OF DIFFERENCES BETWEEN INDIVIDUAL PAIRS OF CONDITION MEANS (Fisher's Z Transformation)

MEAN	CONDITION	
	II (0.7,0.0)	III (0.9,0.0)
MEAN	.497	.867
CONDITION		
I (0.0,0.0) .536	-.039	.331*
II(0.7,0.0) .497		.370*

Critical value of d = .301 at the .05 level.

SIGNIFICANCE OF DIFFERENCES BETWEEN INDIVIDUAL PAIRS OF SESSION MEANS (Fisher's Z Transformation)

MEAN	SESSION	
	II	III
MEAN	.672	.725
SESSION		
I	.169*	.222*
II		.053

Critical value of d = .132 at the .05 level.

* Significant at the .05 level.

In view of the claims of subjects in a pilot study (which resembled the tasks forming Conditions II and III) that they considered not only the magnitude of the two cues (they generally reported that they summed or averaged them) but also the magnitude of the difference between them regardless of the sign of the difference i.e. ABS $(X_1 - X_2)$, it was decided to regress the certainties of the present subjects against both these transformations.

i.e. $\hat{Y}_s = \beta_1(X_1 + X_2) + \beta_2(ABS(X_1 - X_2))$

The coefficients of multiple correlation resulting from this analysis are displayed in Tables 17 (a, b and c). Neither conditions nor sessions were found to have a significant effect on this index (Table 17 d).

The correlations between the subjects' certainties and the ABS $(X_1 - X_2)$ term alone are shown in Tables 18 (a, b and c). An analysis of variance on these coefficients (Table 18 d) showed no effect of sessions but a significantly lesser degree of correlation in Condition I (0.0,0.0) than in Conditions II (0.7,0.0) and III (0.9,0.0), Condition II (0.7,0.0) also having a significantly lower correlation than Condition III (0.9,0.0). When the possible covariance effects of the $(X_1 + X_2)$ term were removed, analysis of variance on the resulting partial correlations no longer showed a significant difference between Conditions II (0.7,0.0) and III (0.9,0.0) but showed a significantly higher correlation in session III than in earlier sessions. (Table 18 e).

F ratios to test the significant of the individual terms, ABS $(X_1 - X_2)$ and $(X_1 + X_2)$, to the fit of the above model were computed and are presented in Tables 19 (a, b and c) for $(X_1 + X_2)$ and Tables 20 (a, b and c) for ABS $(X_1 - X_2)$.

PREDICTABILITY OF THE SUBJECTS' CERTAINTIES FROM A LINEAR COMBINATION OF THE TERMS $(X_1 + X_2)$ AND $ABS(X_1 - X_2)$ WITH ESTIMATED WEIGHTS

Table 17a Condition I (0.0,0.0)

Subject	Session		
	I	II	III
1	.849	.838	.771
2	.891	.942	.892
3	.905	.912	.920
4	.874	.865	.845
5	.852	.868	.825
6	.811	.918	.933
(Model)	(.986)	(.964)	(.971)

Table 17b Condition II (0.7,0.0)

Subject	Session		
	I	II	III
1	.815	.762	.740
2	.868	.867	.922
3	.917	.908	.947
4	.676	.856	.861
5	.922	.891	.912
6	.908	.923	.904
(Model)	(.969)	(.957)	(.963)

Table 17c Condition III (0.9,0.0)

Subject	Session		
	I	II	III
1	.815	.864	.841
2	.846	.921	.907
3	.911	.918	.897
4	.886	.866	.856
5	.889	.855	.841
6	.959	.917	.775
(Model)	(.942)	(.922)	(.910)

Table 17d

ANALYSIS OF VARIANCE ON THE PREDICTABILITY OF SUBJECTS' CERTAINTIES FROM A LINEAR COMBINATION OF THE TERMS $(X_1 + X_2)$ AND ABS$(X_1 - X_2)$

(Fisher's Z Transformation)

Source	S.S.	d.f.	M.S.	F	P
Subjects	1.1785	5			
Sessions	.0341	2	.0170	<1	
Conditions	.0022	2	.0011	<1	
Sessions x Conditions	.1647	4	.0412	1.131	N.S.
Sessions x Subjects	.1824	10	.0182		
Conditions x Subjects	.3009	10	.0301		
Sessions x Conditions x Subjects	.7278	20	.0364		
TOTAL	2.5906	53			

THE CORRELATION OF SUBJECTS' CERTAINTIES WITH THE TERM ABS(X_1-X_2)

Table 18a Condition I (0.0,0.0)

Subject	Session		
	I	II	III
1	-.065 n.s.	-.043 n.s.	-.091 n.s.
2	.308	.198 n.s.	.411
3	.035 n.s.	-.101 n.s.	.057 n.s.
4	-.193 n.s.	-.120 n.s.	-.139 n.s.
5	.176 n.s.	.002 n.s.	.113 n.s.
6	-.002 n.s.	-.015 n.s.	.158 n.s.
(Model)	(-.044)	(-.020)	(.089)

Table 18b Condition II (0.7,0.0)

Subject	Session		
	I	II	III
1	-.150 n.s.	-.004 n.s.	-.471
2	-.171 n.s.	-.135 n.s.	-.060 n.s.
3	-.394	-.279	-.403
4	-.240 n.s.	-.233 n.s.	-.180 n.s.
5	-.365	-.338	-.388
6	-.302	-.201 n.s.	-.035 n.s.
(Model)	(-.560)	(-.493)	(-.621)

Table 18c Condition III (0.9,0.0)

Subject	Session		
	I	II	III
1	-.571	-.420	-.472
2	-.498	-.507	-.467
3	-.403	-.380	-.654
4	-.633	-.583	-.648
5	-.460	-.513	-.634
6	-.222 n.s.	-.166 n.s.	-.665
(Model)	(-.769)	(-.726)	(-.791)

THE CORRELATION OF SUBJECTS' CERTAINTIES WITH THE TERM ABS(X_1-X_2)

Table 18a Condition I (0.0,0.0)

Subject	Session		
	I	II	III
1	-.065 n.s.	-.043 n.s.	-.091 n.s.
2	.308	.198 n.s.	.411
3	.035 n.s.	-.101 n.s.	.057 n.s.
4	-.193 n.s.	-.120 n.s.	-.139 n.s.
5	.176 n.s.	.002 n.s.	.113 n.s.
6	-.002 n.s.	-.015 n.s.	.158 n.s.
(Model)	(-.044)	(-.020)	(.089)

Table 18b Condition II (0.7,0.0)

Subject	Session		
	I	II	III
1	-.150 n.s.	-.004 n.s.	-.471
2	-.171 n.s.	-.135 n.s.	-.060 n.s.
3	-.394	-.279	-.403
4	-.240 n.s.	-.233 n.s.	-.180 n.s.
5	-.365	-.338	-.388
6	-.302	-.201 n.s.	-.035 n.s.
(Model)	(-.560)	(-.493)	(-.621)

Table 18c Condition III (0.9,0.0)

Subject	Session		
	I	II	III
1	-.571	-.420	-.472
2	-.498	-.507	-.467
3	-.403	-.380	-.654
4	-.633	-.583	-.648
5	-.460	-.513	-.634
6	-.222 n.s.	-.166 n.s.	-.665
(Model)	(-.769)	(-.726)	(-.791)

Table 18d

ANALYSIS OF VARIANCE ON THE CORRELATION OF SUBJECTS' CERTAINTIES WITH THE TERM ABS(X_1 - X_2) (Fisher's Z Transformation)

Source	S.S.	d.f.	M.S.	F	P
Subjects	.3875	5			
Sessions	.0350	2	.0175	1.455	N.S.
Conditions	3.2105	2	1.6053	46.375	<.01
Sessions x Conditions	.1771	4	.0443	2.559	N.S.
Sessions x Subjects	.1203	10	.0120		
Conditions x Subjects	.3461	10	.0346		
Sessions x Conditions x Subjects	.3460	20	.0173		
TOTAL	4.6225	53			

SIGNIFICANCE OF DIFFERENCES BETWEEN INDIVIDUAL PAIRS OF CONDITION MEANS (Fisher's Z Transformation)

		CONDITION	
	MEAN	II (0.7,0.0)	III (0.9,0.0)
MEAN		.251	.557
CONDITION			
I (0.0,0.0)	-.040	.291*	.598*
II(0.7,0.0)	.251		.306*

Critical value of d = .138 at the .05 level.

* Significant at the .05 level.

Table 18e

ANALYSIS OF VARIANCE ON THE CORRELATIONS OF SUBJECTS' CERTAINTIES WITH THE TERM ABS(X_1 - X_2) WHEN THE EFFECT OF THE TERM (X_1 + X_2) IS PARTIALLED OUT (Fisher's Z Transformation)

Source	S.S.	d.f.	M.S.	F	P
Subjects	1.442	5			
Sessions	.632	2	.316	12.705	<.01
Conditions	4.069	2	2.035	11.395	<.01
Sessions x Conditions	.050	4	.012	<1	
Sessions x Subjects	.249	10	.025		
Conditions x Subjects	1.785	10	.179		
Sessions x Conditions x Subjects	.721	20	.036		
TOTAL	8.948	53			

SIGNIFICANCE OF DIFFERENCES BETWEEN INDIVIDUAL PAIRS OF CONDITION MEANS (Fisher's Z Transformation)

	MEAN	CONDITION II (0.7,0.0)	CONDITION III (0.9,0.0)
MEAN		.364	.645
CONDITION			
I (0.0,0.0)	-.025	.389*	.670*
II (0.7,0.0)	.364		.281

Critical value of d = .314 at the .05 level.

SIGNIFICANCE OF DIFFERENCES BETWEEN INDIVIDUAL PAIRS OF SESSION MEANS (Fisher's Z Transformation)

	MEAN	SESSION II	SESSION III
MEAN		.252	.481
SESSION			
I	.251	.001	.230*
II	.252		.229*

Critical value of d = .117 at the .05 level.

*Significant at the .05 level.

F RATIOS TO TEST THE SIGNIFICANCE OF THE CONTRIBUTION OF THE $(X_1 + X_2)$ TERM TO THE FIT OF A LINEAR COMBINATION OF THE TERMS $(X_1 + X_2)$ AND ABS$(X_1 - X_2)$

Table 19a Condition I (0.0,0.0)

Subject	Session		
	I	II	III
1	120.93	110.73	67.97
2	159.05	355.92	143.75
3	213.16	230.10	256.92
4	144.62	137.00	114.37
5	119.39	142.94	97.98
6	90.58	252.82	305.83
(Model)	(1628.25)	(624.81)	(763.21)

Table 19b Condition II (0.7,0.0)

Subject	Session		
	I	II	III
1	89.74	65.16	33.76
2	137.78	139.23	265.38
3	203.21	199.98	332.26
4	34.60	119.55	129.27
5	223.39	154.53	191.10
6	195.03	256.48	209.07
(Model)	(479.94)	(371.80)	(348.86)

Table 19c Condition III (0.9,0.0)

Subject	Session		
	I	II	III
1	47.20	105.34	77.66
2	76.94	184.09	160.95
3	183.92	209.49	90.94
4	84.39	77.47	54.99
5	129.21	81.60	48.83
6	509.30	239.31	18.62
(Model)	(122.69)	(101.54)	(54.63)

F RATIOS TO TEST THE SIGNIFICANCE OF THE CONTRIBUTION OF THE ABS(X_1 - X_2) TERM TO THE FIT OF A LINEAR COMBINATION OF THE TERMS (X_1 + X_2) AND ABS(X_1-X_2)

Table 20a Condition I (0.0,0.0)

Subject	Session		
	I	II	III
1	.74 n.s.	.00 n.s.	5.74
2	21.31	23.26	16.31
3	.28 n.s.	1.12 n.s.	3.30 n.s.
4	7.59	1.32 n.s.	13.08
5	5.18	.26 n.s.	.15 n.s.
6	.00 n.s.	.17 n.s.	.00 n.s.
(Model)	(3.64)	(.26)	(5.22)

Table 20b Condition II (0.7,0.0)

Subject	Session		
	I	II	III
1	.09 n.s.	.32 n.s.	21.79
2	.26 n.s.	.89 n.s.	.52 n.s.
3	20.16	11.99	66.89
4	1.68 n.s.	5.05	4.73
5	16.53	17.04	38.27
6	7.28	5.56	.05 n.s.
(Model)	(142.93)	(102.06)	(234.91)

Table 20c Condition III (0.9,0.0)

Subject	Session		
	I	II	III
1	24.50	14.60	21.77
2	17.29	41.10	32.86
3	11.02	15.67	75.60
4	48.46	40.13	55.43
5	16.22	25.82	48.45
6	.00 n.s.	.00 n.s.	42.59
(Model)	(160.00)	(118.99)	(140.90)

The Lens Model indices were computed for this new model and are displayed in Tables 21 (a (i), b (i) and c (i)) for G, Tables 21 (a (ii), b (ii) and c (ii)) for GR_eR_s, Tables 21 (a (iii), b (iii) and c (iii)) for C and Tables 21 (a (iv), b (iv) and c (iv)) for $C\sqrt{(1 - R_e^2)(1 - R_s^2)}$. An analysis of variance on the G index showed significantly better matching in Condition I (0.0,0.0) than in the other two conditions, but no effect of sessions (Table 21d). The C index was not affected significantly by conditions but was significantly lower in session I than in sessions II and III (Table 21e).

THE LENS MODEL INDICES FOR A LINEAR COMBINATION

OF THE TWO TERMS $(X_1 + X_2)$ AND $ABS(X_1 - X_2)$

Table 21a (i) Condition I (0.0,0.0)

THE MATCHING INDEX G

Subject	Session		
	I	II	III
1	1.000	1.000	.978
2	.922	.973	.926
3	.997	.996	1.000
4	.984	.993	.967
5	.968	1.000	.999
6	.999	1.000	.997

Table 21a (ii) Condition I (0.0,0.0)

GR_eR_s

Subject	Session		
	I	II	III
1	.837	.808	.732
2	.810	.884	.802
3	.889	.876	.893
4	.848	.828	.794
5	.814	.836	.800
6	.799	.886	.903

THE LENS MODEL INDICES FOR A LINEAR COMBINATION OF THE TWO TERMS $(X_1 + X_2)$ AND $ABS(X_1 - X_2)$

Table 21a (iii) Condition I (0.0,0.0)

THE MATCHING INDEX C

Subject	Session I	Session II	Session III
1	.690	.796	.722
2	.416	.599	.533
3	.308	.607	.451
4	.433	.421	.477
5	.475	.590	.474
6	.116	.398	.218

Table 21a (iv) Condition I (0.0,0.0)

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

Subject	Session I	Session II	Session III
1	.061	.115	.110
2	.032	.053	.058
3	.022	.066	.043
4	.035	.056	.061
5	.042	.078	.064
6	.011	.042	.019

THE LENS MODEL INDICES FOR A LINEAR COMBINATION

OF THE TWO TERMS $(X_1 + X_2)$ AND $ABS(X_1 - X_2)$

Table 21b (i) Condition II (0.7,0.0)

THE MATCHING INDEX G

Subject	Session		
	I	II	III
1	.909	.860	1.000
2	.914	.927	.805
3	.985	.974	.966
4	.968	.965	.882
5	.978	.988	.966
6	.962	.949	.789

Table 21b (ii) Condition II (0.7,0.0)

GR_eR_s

Subject	Session		
	I	II	III
1	.718	.627	.712
2	.768	.769	.715
3	.876	.846	.881
4	.634	.790	.732
5	.874	.842	.849
6	.846	.837	.686

THE LENS MODEL INDICES FOR A LINEAR COMBINATION
OF THE TWO TERMS $(X_1 + X_2)$ AND $ABS(X_1 - X_2)$

Table 21b (iii) Condition II (0.7,0.0)

THE MATCHING INDEX C

Subject	Session		
	I	II	III
1	.216	.417	.251
2	.403	.475	.514
3	.466	.454	.398
4	-.328	.172	.180
5	.294	.613	.503
6	.153	.372	.002

Table 21b (iv) Condition II (0.7,0.0)

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

Subject	Session		
	I	II	III
1	.031	.079	.046
2	.050	.069	.054
3	.046	.055	.035
4	-.060	.026	.025
5	.028	.081	.056
6	.016	.042	.000

THE LENS MODEL INDICES FOR A LINEAR COMBINATION
OF THE TWO TERMS $(X_1 + X_2)$ AND $ABS(X_1 - X_2)$

Table 21c (i) Condition III (0.9,0.0)

THE MATCHING INDEX G

Subject	Session		
	I	II	III
1	.984	.921	.896
2	.947	.948	.870
3	.879	.887	.972
4	.987	.986	.980
5	.917	.965	.980
6	.750	.749	1.000

Table 21c (ii) Condition III (0.9,0.0)

GR_eR_s

Subject	Session		
	I	II	III
1	.755	.734	.685
2	.754	.805	.718
3	.754	.751	.793
4	.824	.788	.763
5	.767	.761	.749
6	.678	.633	.705

THE LENS MODEL INDICES FOR A LINEAR COMBINATION OF THE TWO TERMS $(X_1 - X_2)$ AND $ABS(X_1 - X_2)$

Table 21c (iii) Condition III (0.9,0.0)

THE MATCHING INDEX C

Subject	Session		
	I	II	III
1	.585	.419	.485
2	.380	.688	.491
3	.178	.387	.520
4	.679	.559	.412
5	.298	.607	.737
6	.315	.269	.464

Table 21c (iv) Condition III (0.9,0.0)

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

Subject	Session		
	I	II	III
1	.114	.082	.109
2	.068	.103	.086
3	.025	.059	.095
4	.106	.108	.089
5	.046	.122	.166
6	.030	.042	.122

Table 21d

ANALYSIS OF VARIANCE ON THE MATCHING INDEX G FOR THE LINEAR COMBINATION OF $(X_1 + X_2)$ AND ABS$(X_1 - X_2)$ (Fisher's Z Transformation)

Source	S.S.	d.f.	M.S.	F	P
Subjects	5.603	5			
Sessions	.276	2	.138	<1	
Conditions	14.944	2	7.472	11.210	<.01
Sessions x Conditions	2.127	4	.532	<1	
Sessions x Subjects	3.331	10	.333		
Conditions x Subjects	6.665	10	.667		
Sessions x Conditions x Subjects	14.933	20	.747		
TOTAL	47.879	53			

SIGNIFICANCE OF DIFFERENCES BETWEEN INDIVIDUAL PAIRS OF CONDITION MEANS (Fisher's Z Transformation)

	MEAN	CONDITION II (0.7,0.0)	CONDITION III (0.9,0.0)
MEAN		1.947	1.922
CONDITION			
I (0.0,0.0)	3.050	-1.103*	-1.128*
II (0.7,0.0)	1.947		-0.025

Critical value of d = .606 at the .05 level.

* Significant at the .05 level.

Table 21e

ANALYSIS OF VARIANCE ON THE MATCHING INDEX C FOR THE LINEAR COMBINATION OF $(X_1 + X_2)$ AND ABS$(X_1 - X_2)$ (Fisher's Z Transformation)

Source	S.S.	d.f.	M.S.	F	P
Subjects	.795	5			
Sessions	.324	2	.162	14.448	<.01
Conditions	.536	2	.268	3.161	N.S.
Sessions x Conditions	.062	4	.015	<1	
Sessions x Subjects	.112	10	.011		
Conditions x Subjects	.848	10	.085		
Sessions x Conditions x Subjects	.612	20	.031		
TOTAL	3.289	53			

SIGNIFICANCE OF DIFFERENCES BETWEEN INDIVIDUAL PAIRS OF SESSION MEANS (Fisher's Z Transformation)

	MEAN	SESSION II	SESSION III
MEAN		.559	.487
SESSION I	.371	.188*	.116*
SESSION II	.559		-.072

Critical value of d = .079 at the .05 level.

* Significant at the .05 level.

EXPERIMENT II

Conclusions and Discussion

The level of performance of subjects in this experiment was generally higher than that of subjects in experiment I. All subjects performed at a level unlikely to be obtained if they were responding randomly and did so in all sessions of all conditions. The ease of interpretation of results however, is reduced by the fact that all subjects performed under all conditions, whereas in the previous experiment, subjects performed under one of the conditions only. Nonetheless, it is unlikely that a medical clinician, for example, is required to make only one decision of the type described in his day to day work, but probably makes such decisions within a number of different sets of disease populations.

No significant effect of conditions was found on either of the indices of accuracy i.e. total number correct and correlation of subjects' certainties with those of the normative model. In view of the more predictable nature of task forming Condition III (0.9,0.0) we might have expected a greater accuracy in this condition since the feedback given to the subjects on the training trials was so much more reliable. However, on examining the various indices open to us, it becomes clear that this accuracy is not due to the same processes in all conditions.

In all conditions the certainties of all subjects were significantly correlated with a linear combination of the two cues, whether the weight for each cue was estimated by multiple regression or whether the weights were set equal to each other i.e. (the correlation with $\beta_1 X_1 + \beta_2 X_2$ or with $\beta_1(X_1 + X_2)$). In general the fit of the linear combination with estimated weights was not significantly better than that of the linear combination with weights set equal, confirming our belief, and the verbal reports of the subjects, that they tended to consider the sum or mean of the two cues when making their decisions. However, the degree of correlation of these linear combinations with the certainties of the subjects was significantly less in Condition III (0.9,0.0) than in the other conditions. Since G, the correlation between the best linear representation of the subject and the best linear representation of the task, did not differ from condition to condition and since R_e, the linear predictability of the task is lower in Condition III (0.9,0.0) than in the other conditions we would expect GR_eR_s, that part of subjects' accuracy which could be attributed to their correct use of the cues in a linear manner, to be lower in Condition III (0.9,0.0) than in the other two conditions. (The argument about R_e also applies to Condition II (0.7,0.0)). Yet there are no differences between conditions in terms of accuracy, and we are led to suspect that subjects in Condition III (0.9,0.0)(and possibly in Condition II (0.7,0.0)) were gaining accuracy by using the cues in a manner which was non linear. The fact that the index of non-linear matching, C, is greater in Condition III (0.9,0.0) than in the other conditions, serves to strengthen this suspicion as does the apparently greater magnitude of the $C\sqrt{(1 - R_e^2)(1 - R_s^2)}$ index for Condition III (0.9,0.0), (though this last index is effected by R_e and is not strictly comparable over conditions).

On examining the correlation of subjects' certainties with the term ABS $(X_1 - X_2)$ only one subject's certainties have significant correlations with this term in Condition I (0.0,0.0), four subjects in Condition II (0.7,0.0) (though only two with consistency), and all six subjects in Condition III (0.9,0.0). The analysis of variance shows these correlations to be higher in Condition III (0.9,0.0) than in Condition II (0.7,0.0) while in turn they are higher in the latter condition than in Condition I (0.0,0.0). In an attempt to discover to what extent these differences might be due to a covariance effect of the term $(X_1 + X_2)$, the possible effects of this term were removed, the analysis of variance on the resulting partial correlations showed the correlation in Condition III (0.9,0.0) still to be greater than in the other two conditions, which were now found not to differ significantly from one another. Thus it is possible that any apparently greater reliance upon the terms ABS $(X_1 - X_2)$ in Condition II (0.7,0.0) than in Condition I (0.0,0.0) may simply be a covariance effect of $(X_1 + X_2)$, but such an effect cannot explain the greater reliance upon this term in Condition III (0.9,0.0). Subjects in this condition seem to have appreciated the diagnostic value of this term, and to have used it to help them assign their certainty estimates. (A parallel procedure was applied to the correlations of subjects' certainties with the term $(X_1 + X_2)$ to see if any of the significant differences found with these coefficients could be explained by a covariance effect of the term ABS $(X_1 - X_2)$. The pattern of significant differences was not altered by such a process).

The multiple correlation coefficients resulting from a multiple regression of the subjects' certainties against both the $(X_1 + X_2)$ and ABS $(X_1 - X_2)$ terms were not found to differ significantly over conditions. Thus the inclusion of the ABS $(X_1 - X_2)$ term has made up for the lack of predictive power of the $(X_1 + X_2)$ for the certainties of subjects in Condition III (0.9,0.0).

F ratios to test the significance of the contribution of the $(X_1 + X_2)$ term to the fit of the above linear combination of $(X_1 + X_2)$ and ABS $(X_1 - X_2)$ were significant for all subjects in all sessions of all conditions. Thus, though subjects in Condition III (0.9,0.0) were using the cues non-linearly, they were also using them in a linear manner. The ABS $(X_1 - X_2)$ term contributed significantly for all subjects in Condition III (0.9,0.0), (though for subject 6 only in the third session), for most subjects in Condition II (0.7,0.0) and for one or two subjects in Condition I (0.0,0.0). Particularly in Condition I (0.0,0.0) it can be seen that though this term is often not significantly correlated with the certainties, when taken by itself, the correlations of the certainties with the $(X_1 + X_2)$

term is so large that even these very small amounts of extra explicable variance (i.e. that variance explained by the ABS $(X_1 - X_2)$ term which is not explained by $(X_1 + X_2)$) are shown by the F ratio to be significant. Thus though the term is regarded as contributing significantly to the fit of the above model, it is in some cases capable of explaining only 2% or 3% of the variance in the subjects' certainties, it just happens that this 2% or 3% is not explicable by the $(X_1 + X_2)$ term.

The interpretation of the Lens Model indices is even more difficult in this experiment than the previous one. Whether the underlying model being tested is the linear combination of X_1 and X_2 or the linear combination of the terms $(X_1 + X_2)$ and ABS $(X_1 - X_2)$, the tasks forming the three conditions are not of equal predictability. Unlike the task in experiment I, the present tasks are not even equally predictable to a rank order criterion. The rank order indices were computed but are not presented since they did not differ in any marked way from their product moment equivalents.

Some of the indices of the model with a linear combination of X_1 and X_2 give an indication of those subjects within a condition whose accuracy may be regarded as being due almost totally to the utilization of the cues in a linear manner and those for whom some non-linear process must be postulated to explain the accuracy of their predictions. For example, in Condition II (0.7,0.0) and III (0.9,0.0), the index $C\sqrt{(1-R_e^2)(1-R_s^2)}$

shows a markedly higher value when the corresponding correlation between the ABS $(X_1 - X_2)$ term and subjects' certainties is significant than when it is not so. However the task is so highly linearly predictable in Condition I (0.0,0.0) that this similarity is no longer obvious, all values of the index being very small. (This task is of course completely linear to a rank order criterion).

The indices of the model with the linear combination of $(X_1 + X_2)$ and ABS $(X_1 - X_2)$ do not add a great deal to our knowledge of the processes involved. The high values of GR_eR_s in all conditions may be taken as an indication that the accuracy of all subjects in all conditions is reasonably well explained, if we assume that they are using these terms appropriately and very little is added by the use of terms other than these. However, much of this information may be obtained from simple multiple regression and the lack of any error theory for the Lens Model indices makes their interpretation difficult when the tasks are of unequal predictability and the response scales are possibly unreliable.

It would appear then, that the decision processes of subjects in Condition I (0.0,0.0), where the cue intercorrelations are the same in both disease populations, can be reasonably well described by a model which linearly combines the two cues (apparently giving equal weight to both) and bases its certainty estimates on the resulting value.

In Condition III (0.9,0.0) where, in one disease population the cues are highly correlated but in the other are independent, it is necessary to postulate some non-linear utilization of the cues as well as this linear utilization in order to gain a reasonable representation of subjects' decision processes. In this condition all subjects claimed to consider the differences between the two cues (regardless of sign) as well as the magnitude of the cues when making their decisions, and it was found that a linear combination of the two terms ABS $(X_1 - X_2)$ and $(X_1 + X_2)$ could make quite accurate predictions of the subjects' certainties and that this accuracy suffered if either term was removed.

In Condition II (0.7,0.0) where the two cues are moderately correlated in one disease population but independent in the other, the situation is not so clear. On average, there is no more sign of the utilization of the ABS $(X_1 - X_2)$ term in this condition than in Condition I (0.0,0.0) when the possible covariance effects of the term $(X_1 + X_2)$ were removed. However, this partial correlation process removes all the variance that could possibly be due to $(X_1 + X_2)$ not just that variance which is due to this term and, since the analysis of variance on the unaltered correlation coefficients does show these coefficients to be greater in Condition II (0.7,0.0) than in Condition I (0.0,0.0), we are left with a dilemma. Turning to the subjects' verbal reports, we find that only Subject 4 reported considering the similarity of the two cues in Condition II (0.7,0.0). Yet neither the correlation of her certainties with the ABS $(X_1 - X_2)$

term nor the resulting F ratios are particularly high. The F ratios are significant for a number of subjects in this condition, but not for all of them. It seems then that some subjects may have been incorporating this ABS $(X_1 - X_2)$ term into their decision processes, whereas others appear to have relied solely on the linear combination of X_1 and X_2. The overall effect however, is not significantly different from behaviour in Condition I (0.0,0.0) when possible covariance effects are removed.

It appears then, that human subjects are able to modify their behaviour to take into account the differing relationships which can exist between cues from one diagnostic population to another, and that in these sorts of tasks, at least, this non-linear use of cues can be fairly well represented by the inclusion of the term ABS $(X_1 - X_2)$ in a linear combination with the sum or average of X_1 and X_2. In order for all subjects to use such a cue however, the difference in correlation must be quite large, e.g. + 0.9 and + 0.0, with populations less different in correlation, not all subjects use this cue.

EXPERIMENT III

FURTHER INVESTIGATION OF DECISIONS BETWEEN BIVARIATE NORMAL POPULATIONS WITH UNEQUAL MEAN VECTORS AND UNEQUAL COVARIANCE MATRICES

INTRODUCTION TO EXPERIMENT III

In the second experiment it was demonstrated that when subjects are required to make discriminations between the members of two bivariate normal distributions differing not only in mean vector but also in dispersion matrix, they tend to pay more attention to the pattern or configuration of scales than if the dispersion matrices are equal. In the condition with the largest difference in cue intercorrelation between the populations, the non-linear transformation ABS $(X_1 - X_2)$ was highly correlated with the certainty estimates of all subjects, and accounted for sufficient of the variance not explicable by a simple linear combination to justify its inclusion in a multiple regression equation.

The subjects performed well in all conditions, their certainty estimates being highly correlated with those generated by the normative model (with its square and cross product terms). Under different circumstances this may have led to the conclusion that the subjects were actually behaving in the manner prescribed by the model, but, from the evidence of verbal reports and of the regression analysis it appears that subjects were actually achieving this high level of performance by incorporating the ABS $(X_1 - X_2)$ term into their decision processes. The consideration of the ABS $(X_1 - X_2)$ transformation by the subjects, is very much an "ad hoc" process, being valid as it is, in only a limited class of situations of this type. This is in marked contrast to the normative model which fits perfectly all situations of this kind, changing only the values of the weights of the various terms to suit the parameters of the populations.

If this inclusion of the ABS $(X_1 - X_2)$ term is as "ad hoc" a process as is felt then it should be possible to design a situation where other non-linear terms than this will be utilized by human subjects. The present experiment was designed to attempt to discover to what extent the inclusion of non-linear terms of this kind is modified by certain parameters of the task, which have an all but trivial effect on the normative model.

The utilization of the ABS $(X_1 - X_2)$ term by the subjects implies some sort of realization on their part, that the cue intercorrelation differs from one population to another, that the magnitude of the two cues are more similar in one population than in the other. Put another way, they realised that if the two cues had similar magnitude then this increases the probability of this patient having Disease 1. In fact in Conditions II and III of experiment II, if X_1 had a value x then the most likely value of X_2 was x if the patient had Disease 1, but in Disease 2 the expected value of the cue X_2 was independent of the value of the cue X_1. So for Disease 1 the best estimate of X_1 is given by:-

$$\hat{X}_1 = X_2$$

Conversely the best estimate of X_2 i.e. X_2 is given by

$$\hat{X}_2 = X_1$$

What would be the effect of making the relationships say

$$\hat{X}_1 = X_2 + C$$

and $$\hat{X}_2 = X_1 - C \quad ?$$

This question is obviously related to the effect of the parameter "a" in the learning of linear functions of the kind.

$$Y = a + b\,X$$

The relationship between X_1 and X_2 is however only probabilistic, and the effect of the parameter in probabilistic situations has not been investigated. However, in the learning of linear functions the parameter "a" was found to have no significant effect so we might expect a similar finding if the relationship between Y and X was prone to error. Our task, though, does not require the subject to predict X_1 from X_2 or vice versa and it is possible that the relationship existing between them may go unnoticed and hence unused by the subjects if it is of the form

$$\hat{X}_1 = X_2 + c$$

where c has a value other than zero.

A transformation of the scores of all patients (both Disease 1 and Disease 2) on one of the predictor variables, by the addition of a constant would create such a relationship between X_1 and X_2 in Disease 1 without altering the relationship in Disease 2 (.i.e independence). It would also invalidate the use of the ABS $(X_1 - X_2)$ transformation while in no way altering the efficiency of the normative model, which accommodates this change with a simple alteration of parameter values. It remains to be seen whether subjects will now find some other non-linear transformation of the cues or whether the information contained in cue intercorrelations will go unnoticed and unused.

The research into both the learning of linear functions and learning correlations has shown that performance is consistently worse in tasks with negative rather than positive relationships between the variable i.e. functions of the form

$$Y = a + b X$$

where b is a negative number. Supposing we made the relationship between X_1 and X_2 of the form

$$\hat{X}_1 = c - X_2$$

(The c parameter being set so as to keep the actual values of both X_1 and X_2 in a similar range to their range before transformation).

There is even stronger reason here to believe that the relationship between X_1 and X_2 will go unnoticed, since it is so difficult to perceive and use accurately when it is the main aim of the task. It would seem likely then that subjects put in a situation of this kind might miss the relationship altogether and fall back on a simple linear combination of cues.

Such a relationship may be introduced by the transformation of one variable such that what was previously the highest score takes the value of the lowest and vice versa and all scores in between are accordingly transformed. Again, this transformation has no effect on the efficiency of the normative model, the changes being accomodated by changes in the value of certain weighting terms.

With these thoughts in mind, three tasks were generated.

Condition I (Untransformed)

The first task was simply that forming the third condition in Experiment II, i.e. Disease 1 had mean score 7.5 on both variables, standard deviations of 2.0 and cue intercorrelation of +.9. Disease 2 had mean score 9.5 on both variables, standard deviation 2.0 and zero cue intercorrelation. The samples were drawn from populations with these parameters in the manner outlined in Appendix B, and besides forming the stimuli in this condition they were also suitably transformed to form the stimuli of the next two conditions to be described. In this way the conditions were made exactly equal in terms of difficulty (or discriminability) from a statistical point of view, and were not open to differences through sampling.

Condition II (X_2 + 2)

The scores of all patients, in the original sample, on variable X_2 were transformed by adding a constant C = 2.0. This in no way affected the standard deviations and intercorrelations of the cues within each population. The population means, however, were increased by 2.0 on the X_2 variable, making them 9.5 for Disease 1 and 11.5 for Disease 2. Mean scores on X_1 were not affected.

Condition III (X_2 Reversed)

The scores of all subjects, in the original sample, on variable X_2 were transformed by rotating them about 8.5 (thus 7.5 became 9.5, 5.0 became 12.0 etc.) The range of scores remained unchanged, the standard deviations for each population remaining 2.0. The correlation in Disease 1 became -.9, though the correlation in Disease 2 remained 0.0. The mean scores on variable X_2 were reversed, Disease 1 having now a mean of 9.5, Disease 2 one of 7.5.

(All these conditions are graphically displayed in Appendix C.)

SUBJECTS

Again, it was felt that subjects whose motivation could be assured, should be used. Twelve male students, average age approximately 25 years, all known to the experimenter and all honours or postgraduate students in psychology acted as subjects. They each participated for three one-hour sessions in each of the three conditions, all orders of presentation of the conditions occurring an equal number of times. They were paid at a rate of 30p an hour for their participation.

APPARATUS

Again, the same equipment as was used in the previous two experiments.

PROCEDURE

The same procedure was followed here as in the second experiment, the subjects being asked to give a few sentences about their decision processes after the third session of each condition.

EXPERIMENT III

Results

Again only the data from the 50 test trials of each session were analysed, the certainties of the subjects being transformed to yield a scale ranging from 1 to 100. The total number of correct responses (i.e. agreements with the normative model) are shown in Tables 22 (a, b and c) and an analysis of variance showed a significant interaction effect of conditions and sessions on this index (Table 22 d). Which seems to be due to performance improving over sessions in Condition I (Untransformed) and Condition II (X_2 + 2), but, if anything, deteriorating over sessions in Condition III (X_2 Reversed). The other measure of accuracy, the correlations of the subjects' certainties with those of the normative model, are shown in Tables 23 (a, b, c), only sessions were found to affect this measure significantly, with a lower correlation in the first session than in sessions II and III.

NUMBER OF CORRECT RESPONSES (i.e. AGREEMENTS WITH THE NORMATIVE MODEL).

Table 22a Condition I (Untransformed)

Subject	Session		
	I	II	III
1	39	38	37
2	42	44	40
3	27 n.s.	33	43
4	39	44	46
5	24 n.s.	36	34
6	24 n.s.	28 n.s.	28 n.s.
7	39	41	47
8	41	40	45
9	42	41	43
10	42	42	44
11	41	45	44
12	42	47	47

Table 22b Condition II (X_2+2)

Subject	Session		
	I	II	III
1	40	33	36
2	42	41	39
3	43	38	44
4	37	37	43
5	44	41	44
6	26 n.s.	24 n.s.	31
7	26 n.s.	37	49
8	41	41	43
9	24 n.s.	35	31
10	40	39	42
11	41	42	42
12	38	36	41

NUMBER OF CORRECT RESPONSES (i.e. AGREEMENTS WITH THE NORMATIVE MODEL).

Table 22c Condition III (X_2 Reversed)

Subject	Session		
	I	II	III
1	39	36	34
2	40	37	36
3	42	34	37
4	38	41	34
5	39	37	32
6	33	37	41
7	38	41	38
8	33	37	26 n.s.
9	33	26 n.s.	19 n.s.
10	39	37	38
11	40	36	37
12	43	37	37

Table 22d

ANALYSIS OF VARIANCE ON THE NUMBER OF CORRECT RESPONSES

Source	S.S.	d.f.	M.S.	F	P
Subjects	1096.63	11			
Sessions	23.46	2	11.76	1	
Conditions	190.13	2	95.06	1.64	N.S.
Sessions x Conditions	265.26	4	66.31	7.07	.01
Sessions x Subjects	261.43	22	11.88		
Conditions x Subjects	1274.09	22	57.91		
Sessions x Conditions x Subjects	412.52	44	9.38		
TOTAL	2523.52	107			

SIGNIFICANCE OF DIFFERENCES BETWEEN INDIVIDUAL PAIRS OF (CONDITION x SESSION) MEANS

CONDITION			I (Untrans)		II (X_2 + 2)			III (Reversed)		
	SESSION		II	III	I	II	III	I	II	III
		MEAN	39.92	41.50	36.83	37.00	39.58	38.08	36.33	34.08
I (Untrans)	I	36.83	3.09*	4.67*	0.00	0.17	2.75*	1.25	-0.50	-1.75
	II	39.92		1.58	-3.09*	-2.92*	-0.34	-1.84	-3.59*	-5.84*
	III	41.50			-4.67*	-4.50*	-1.92	-3.42*	-5.17*	-7.42*
II (X_2 + 2)	I	36.83				0.17	2.65*	1.25	-0.50	-2.75*
	II	37.00					2.58*	1.08	-0.67	-2.92*
	III	39.58						-1.50	-3.25*	-5.50*
III (X_2 Rev)	I	38.08							-1.75	-4.00*
	II	36.33								-2.25

Critical value of d = 2.53 at the .05 level.

* Significant at the .05 level.

Table 22d

ANALYSIS OF VARIANCE ON THE NUMBER OF CORRECT RESPONSES

Source	S.S.	d.f.	M.S.	F	P
Subjects	1096.63	11			
Sessions	23.46	2	11.76	1	
Conditions	190.13	2	95.06	1.64	N.S.
Sessions x Conditions	265.26	4	66.31	7.07	.01
Sessions x Subjects	261.43	22	11.88		
Conditions x Subjects	1274.09	22	57.91		
Sessions x Conditions x Subjects	412.52	44	9.38		
TOTAL	2523.52	107			

SIGNIFICANCE OF DIFFERENCES BETWEEN INDIVIDUAL PAIRS OF (CONDITION x SESSION) MEANS

CONDITION			I (Untrans)		II (X_2 + 2)			III (Reversed)		
	SESSION		II	III	I	II	III	I	II	III
		MEAN	39.92	41.50	36.83	37.00	39.58	38.08	36.33	34.08
I (Untrans)	I	36.83	3.09*	4.67*	0.00	0.17	2.75*	1.25	-0.50	-1.75
	II	39.92		1.58	-3.09*	-2.92*	-0.34	-1.84	-3.59*	-5.84*
	III	41.50			-4.67*	-4.50*	-1.92	-3.42*	-5.17*	-7.42*
II (X_2 + 2)	I	36.83				0.17	2.65*	1.25	-0.50	-2.75*
	II	37.00					2.58*	1.08	-0.67	-2.92*
	III	39.58						-1.50	-3.25*	-5.50*
III (X_2 Rev)	I	38.08							-1.75	-4.00*
	II	36.33								-2.25

Critical value of d = 2.53 at the .05 level.

* Significant at the .05 level.

CORRELATIONS OF THE SUBJECTS' CERTAINTIES WITH THOSE OF THE NORMATIVE MODEL (r_a)

Table 23a Condition I (Untransformed)

Subject	Session		
	I	II	III
1	.662	.674	.773
2	.709	.857	.822
3	.169n.s.	.438	.662
4	.746	.842	.920
5	-.067n.s.	.430	.435
6	-.037n.s.	-.010n.s.	.197n.s.
7	.735	.800	.833
8	.800	.834	.849
9	.814	.690	.686
10	.870	.828	.789
11	.783	.852	.883
12	.794	.912	.905

Table 23b Condition II (X_2+2)

Subject	Session		
	I	II	III
1	.750	.604	.698
2	.829	.919	.883
3	.647	.669	.709
4	.697	.744	.875
5	.794	.775	.757
6	.072 n.s.	.026 n.s.	.133 n.s.
7	.050 n.s.	.652	.650
8	.784	.819	.881
9	.192 n.s.	.417	.367
10	.779	.685	.762
11.	.835	.823	.785
12	.735	.728	.800

CORRELATIONS OF THE SUBJECTS' CERTAINTIES WITH THOSE OF THE NORMATIVE MODEL (r_a)

Table 23c Condition III (X_2 Reversed)

Subject	Session		
	I	II	III
1	.567	.627	.703
2	.648	.644	.703
3	.642	.548	.687
4	.714	.795	.495
5	.684	.625	.480
6	.560	.567	.799
7	.641	.725	.653
8	.604	.625	.247 n.s.
9	.307	.218 n.s.	-.148 n.s.
10	.665	.779	.815
11	.662	.598	.764
12	.719	.649	.760

Table 23d

ANALYSIS OF VARIANCE ON THE CORRELATIONS OF SUBJECTS' CERTAINTIES WITH THOSE OF THE NORMATIVE MODEL (Fisher's Z Transformation)

Source	S.S.	d.f.	M.S.	F	P
Subjects	6.929	11			
Sessions	.334	2	.167	5.781	<.01
Conditions	.669	2	.334	1.394	N.S.
Sessions x Conditions	.209	4	.052	1.577	N.S.
Sessions x Subjects	.636	22	.029		
Conditions x Subjects	5.278	22	.240		
Sessions x Conditions x Subjects	1.457	44	.033		
TOTAL	15.452	107			

SIGNIFICANCE OF DIFFERENCES BETWEEN INDIVIDUAL PAIRS OF SESSION MEANS (Fisher's Z Transformation)

	MEAN	SESSION II	SESSION III
MEAN		.854	.904
SESSION			
I	.769	.085*	.235*
II	.854		.050

Critical value of d = .083 at the .05 level.

* Significant at the .05 level.

The multiple correlation between subjects' certainties and the two cues X_1 and X_2 are shown in Tables 24 (a, b and c). These correlations were significantly higher in Condition III (X_2 Reversed) than in the other two conditions (Table 24d).

The Lens Model indices for the linear combination of X_1 and X_2 were computed and are shown :-

G in Tables 25 (a (i), b (i) and c (i)),

GR_eR_s in Tables 25 (a (ii), b (ii) and c (ii)),

C in Tables 25 (a (iii), b (iii) and c (iii)) and

$C\sqrt{(1 - R_e^2)(1 - R_s^2)}$ in Tables 25 (a (iv), b (iv) and c(iv)).

Neither sessions nor conditions significantly affected G (Table 25d) but both had significant effects on C, the index being lower in Condition III (X_2 Reversed) than in the other conditions and being higher in session III than in previous sessions (Table 25e).

THE PREDICTABILITY OF SUBJECTS' CERTAINTIES BY A LINEAR COMBINATION OF THE TWO CUES WITH ESTIMATED WEIGHTS (R_s).

Table 24a Condition I (Untransformed)

Subject	Session I	Session II	Session III
1	.847	.888	.878
2	.627	.571	.464
3	.382	.700	.705
4	.856	.872	.799
5	.385	.762	.686
6	.149n.s.	.452	.299n.s.
7	.688	.595	.609
8	.745	.587	.579
9	.482	.252n.s.	.370
10	.747	.699	.702
11	.531	.759	.685
12	.577	.642	.629
Model	(.722)	(.744)	(.767)

Table 24b Condition II (X_2+2)

Subject	Session I	Session II	Session III
1	.927	.883	.833
2	.665	.766	.758
3	.594	.687	.711
4	.712	.811	.520
5	.787	.827	.721
6	.130n.s.	.043n.s.	.188n.s.
7	.322n.s.	.674	.659
8	.622	.676	.704
9	.142n.s.	.131n.s.	.259n.s.
10	.787	.507	.840
11	.651	.832	.546
12	.846	.828	.818
Model	(.722)	(.744)	(.767)

THE PREDICTABILITY OF SUBJECTS' CERTAINTIES BY A LINEAR COMBINATION OF THE TWO CUES WITH ESTIMATED WEIGHTS (R_s)

Table 24c Condition III (X_2 Reversed)

Subject	Session		
	I	II	III
1	.654	.814	.892
2	.954	.948	.951
3	.737	.840	.742
4	.849	.790	.643
5	.837	.891	.725
6	.654	.826	.867
7	.645	.641	.744
8	.655	.785	.428
9	.402	.186n.s.	.195n.s.
10	.893	.900	.935
11	.748	.719	.896
12	.915	.927	.965
Model	(.722)	(.744)	(.767)

Table 24d

ANALYSIS OF VARIANCE ON THE R_ss FOR A LINEAR COMBINATION OF THE TWO CUES (Fisher's Z Transformation)

Source	S.S.	d.f.	M.S.	F	P
Subjects	8.695	11			
Sessions	.136	2	.068	1.485	N.S.
Conditions	2.675	2	1.338	6.276	.01
Sessions x Conditions	.034	4	.008	<1	
Sessions x Subjects	1.011	22	.046		
Conditions x Subjects	4.690	22	.213		
Sessions x Conditions x Subjects	1.685	44	.038		
TOTAL	18.926	107			

SIGNIFICANCE OF DIFFERENCES BETWEEN INDIVIDUAL PAIRS OF CONDITION MEANS (Fisher's Z Transformation)

	MEAN	CONDITION II (X_2 + 2)	CONDITION III (X_2 Reversed)
MEAN		.818	1.129
CONDITION			
I (Untransformed)	.776	.032	.353*
II (X_2 + 2)	.818		.311*

Critical value of d = .226 at the .05 level.

* Significant at the .05 level.

THE LENS MODEL INDICES FOR A LINEAR COMBINATION OF THE TWO CUES

Table 25a(i) Condition I (Untransformed)

THE MATCHING INDEX G

Subject	Session		
	I	II	III
1	.991	1.000	.987
2	.897	.925	.985
3	.150	.892	1.000
4	.995	.999	.992
5	.005	.819	.602
6	.956	.089	.319
7	.967	.988	.989
8	.990	1.000	.994
9	.995	1.000	.998
10	.994	.981	.995
11	.988	1.000	.988
12	1.000	1.000	.995

Table 25a(ii) Condition I (Untransformed)

GR_eR_s

Subject	Session		
	I	II	III
1	.606	.660	.665
2	.405	.393	.350
3	.041	.464	.540
4	.614	.648	.608
5	.001	.464	.317
6	.103	.030	.073
7	.480	.437	.463
8	.533	.437	.442
9	.346	.188	.283
10	.536	.509	.536
11	.379	.564	.519
12	.416	.477	.480

THE LENS MODEL INDICES FOR A LINEAR COMBINATION OF THE TWO CUES

Table 25a(iii) Condition I (Untransformed)

THE MATCHING INDEX C

Subject	Session		
	I	II	III
1	.153	.044	.350
2	.563	.846	.831
3	.200	-.054	.266
4	.477	.593	.810
5	-.107	-.079	.254
6	-.205	-.067	.202
7	.509	.673	.728
8	.579	.735	.780
9	.771	.776	.676
10	.726	.666	.555
11	.690	.665	.779
12	.669	.849	.853

Table 25a(iv) Condition I (Untransformed)

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

Subject	Session		
	I	II	III
1	.056	.014	.107
2	.304	.465	.472
3	.128	-.026	.121
4	.171	.194	.313
5	-.068	-.034	.118
6	-.140	-.040	.123
7	.256	.362	.370
8	.267	.398	.407
9	.468	.502	.403
10	.334	.319	.254
11	.405	.290	.364
12	.378	.435	.425

THE LENS MODEL INDICES FOR A LINEAR COMBINATION OF THE TWO CUES

Table 25b(i) Condition II (X_2+2)

THE MATCHING INDEX G

Subject	Session		
	I	II	III
1	1.000	.988	.972
2	.994	.997	1.000
3	1.000	.988	1.000
4	.976	.993	.996
5	.994	.997	.942
6	-.311	.277	.208
7	-.376	.974	.985
8	.930	.998	.993
9	.680	.862	.413
10	.980	.950	.988
11	.995	.993	.917
12	.985	.963	.984

Table 25b(ii) Condition II (X_2+2)

GR_cR_s

Subject	Session		
	I	II	III
1	.669	.648	.622
2	.477	.568	.582
3	.429	.505	.545
4	.501	.599	.397
5	.565	.613	.521
6	-.029	.009	.030
7	-.087	.488	.498
8	.417	.501	.537
9	.070	.084	.082
10	.557	.582	.637
11	.467	.374	.384
12	.602	.593	.618

THE LENS MODEL INDICES FOR A LINEAR COMBINATION OF THE TWO CUES.

Table 25b (iii) Condition II ($X_2 + 2$)

THE MATCHING INDEX C

Subject	Session		
	I	II	III
1	.311	-.140	.205
2	.681	.817	.720
3	.392	.339	.363
4	.402	.371	.872
5	.537	.430	.532
6	.147	.025	.163
7	.212	.333	.315
8	.676	.644	.756
9	.163	.503	.461
10	.520	.278	.359
11	.700	.778	.746
12	.361	.360	.494

Table 25b(iv) Condition II ($X_2 + 2$)

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

Subject	Session		
	I	II	III
1	.081	-.044	.073
2	.352	.352	.310
3	.218	.165	.164
4	.196	.145	.478
5	.229	.162	.237
6	.101	.017	.103
7	.139	.165	.152
8	.367	.318	.344
9	.112	.333	.285
10	.222	.103	.125
11	.368	.449	.401
12	.133	.135	.182

THE LENS MODEL INDICES FOR A LINEAR COMBINATION OF THE TWO CUES

Table 25c(i) Condition III (X_2 Reversed)

THE MATCHING INDEX G

Subject	Session I	Session II	Session III
1	.862	.997	.979
2	.983	.994	.949
3	1.000	.989	.993
4	1.000	.994	.892
5	.995	.996	.999
6	1.000	.992	.999
7	.925	.998	1.000
8	.994	.934	.936
9	.915	.993	-.580
10	.991	.998	.982
11	.997	.996	.995
12	.996	.997	.992

Table 25 c(ii) Condition II (X_2 Reversed)

GR_eR_s

Subject	Session I	Session II	Session III
1	.407	.604	.670
2	.676	.700	.693
3	.532	.618	.566
4	.612	.584	.440
5	.610	.660	.556
6	.472	.609	.665
7	.431	.480	.571
8	.469	.545	.308
9	.266	.137	-.087
10	.638	.668	.705
11	.538	.532	.684
12	.657	.687	.735

THE LENS MODEL INDICES FOR A LINEAR COMBINATION OF THE TWO CUES

Table 25c(iii) Condition III (X_2 Reversed)

THE MATCHING INDEX C

Subject	Session		
	I	II	III
1	.305	.059	.114
2	-.137	-.261	-.004
3	.235	-.193	.283
4	.277	.514	.111
5	.219	-.113	-.172
6	.169	-.111	.420
7	.398	.487	.191
8	.258	.193	-.102
9	.065	.123	-.098
10	.085	.382	.485
11	.270	.141	.282
12	.220	-.150	.148

Table 25c(iv) Condition III (X_2 Reversed)

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

Subject	Session		
	I	II	III
1	.160	.023	.033
2	-.029	-.056	-.001
3	.110	-.070	.122
4	.101	.211	.055
5	.083	-.034	-.076
6	.088	-.042	.134
7	.210	.250	.082
8	.135	.080	-.059
9	.041	.081	-.062
10	.027	.111	.110
11	.124	.065	.080
12	.062	-.038	.025

Table 25d

ANALYSIS OF VARIANCE ON THE MATCHING INDEX G FOR A LINEAR COMBINATION OF THE TWO CUES (Fisher's Z Transformation)

Source	S.S.	d.f.	M.S.	F	P
Subjects	23.201	11			
Sessions	2.411	2	1.206	1.003	N.S.
Conditions	3.384	2	1.692	<1	
Sessions x Conditions	3.739	4	.935	1.134	N.S.
Sessions x Subjects	26.452	22	1.202		
Conditions x Subjects	73.587	22	3.345		
Sessions x Conditions x Subjects	36.281	44	.825		
TOTAL	149.055	107			

Table 25e

ANALYSIS OF VARIANCE ON THE MATCHING INDEX C FOR A LINEAR COMBINATION OF THE TWO CUES (Fisher's Z Transformation)

Source	S.S.	d.f.	M.S.	F	P
Subjects	4.447	11			
Sessions	.293	2	.146	3.884	.05
Conditions	4.524	2	2.262	9.991	.01
Sessions x Conditions	.325	4	.081	2.301	N.S.
Sessions x Subjects	.829	22	.038		
Conditions x Subjects	4.981	22	.226		
Sessions x Conditions x Subjects	1.555	44	.035		
TOTAL	16.954	107			

SIGNIFICANCE OF DIFFERENCES BETWEEN INDIVIDUAL PAIRS OF CONDITION MEANS (Fisher's Z Transformation)

	MEAN	CONDITION II (X_2 + 2)	CONDITION III (X_2 Reversed)
MEAN		.518	.149
CONDITION			
I (Untransformed)	.627	-.109	-.478*
II (X_2 + 2)	.518		-.369*

Critical value of d = .233 at the .05 level

SIGNIFICANCE OF DIFFERENCES BETWEEN INDIVIDUAL PAIRS OF SESSION MEANS (Fisher's Z Transformation)

	MEAN	SESSION II	SESSION III
MEAN		.394	.505
SESSION			
I	.395	-.001	.110*
II	.394		.111*

Critical value of d = .095 at the .05 level.

* Significant at the .05 level

The multiple correlation coefficients and the Lens Model indices were also computed for the linear combination of $(X_1 + X_2)$ and $ABS(X_1 - X_2)$ for Condition I (Untransformed) and for the equivalent terms $(X_1 + X_2 - 2)$ and $ABS(X_1 - (X_2 - 2))$ for Condition II $(X_2 + 2)$ and $(X_1 - X_2)$ and $ABS((X_1 + X_2) - 17)$ for Condition III (X_2 Reversed). The multiple correlation coefficients (Tables 26 (a, b and c)) were found to be affected only by sessions, being significantly lower in session I than in sessions II and III (Table 26d). The matching index G (Tables 27 (a (i), b (i) and c (i)))was significantly lower in Condition III (X_2 Reversed) than in Condition I (Untransformed) but was unaffected by sessions (Table 27d). The index of "non model" matching, C (Tables 27 (a (iii), b (iii) and c (iii)))was affected by neither sessions nor conditions (Table 27e). GR_eR_s, (Tables 27 (a (ii), b (ii) and c (ii))), and $C\sqrt{(1 - R_e^2)(1 - R_s^2)}$ (Tables 27 (a (iv), b (iv) and c (iv))), were not subjected to analyses of variance as they are simply made up of terms already tested and no such analyses would be independent of these previous analyses.

THE PREDICTABILITY OF SUBJECTS' CERTAINTIES BY A LINEAR COMBINATION OF THE TERMS (X_1+X_2) AND ABS(X_1-X_2) OR THEIR EQUIVALENTS.

Table 26a Condition I (Untransformed)

Subject	Session		
	I	II	III
1	.847	.884	.890
2	.638	.863	.935
3	.337	.604	.737
4	.845	.884	.889
5	.063n.s.	.684	.521
6	.135n.s.	.093n.s.	.285n.s.
7	.715	.761	.793
8	.781	.812	.776
9	.740	.691	.719
10	.814	.754	.789
11	.743	.790	.849
12	.715	.811	.862
Model	(.876)	(.900)	(.947)

Table 26b Condition II (X_2+2)

Subject	Session		
	I	II	III
1	.914	.896	.833
2	.782	.867	.863
3	.656	.732	.758
4	.735	.819	.834
5	.801	.832	.746
6	.070n.s.	.080n.s.	.093n.s.
7	.211n.s.	.712	.699
8	.743	.742	.818
9	.143n.s.	.484	.422
10	.811	.730	.821
11	.744	.813	.759
12	.804	.779	.852
Model	(.876)	(.900)	(.947)

THE PREDICTABILITY OF SUBJECTS' CERTAINTIES BY A LINEAR COMBINATION OF THE TERMS (X_1+X_2) AND ABS(X_1-X_2) OR THEIR EQUIVALENTS.

Table 26c Condition III (X_2 Reversed)

Subject	Session		
	I	II	III
1	.516	.815	.892
2	.953	.905	.942
3	.739	.857	.744
4	.841	.814	.618
5	.820	.895	.705
6	.651	.829	.876
7	.606	.708	.740
8	.647	.764	.380
9	.366	.199	.146
10	.893	.896	.940
11	.741	.726	.895
12	.914	.927	.965
Model	(.876)	(.900)	(.947)

Table 26d

ANALYSIS OF VARIANCE ON THE CORRELATIONS OF SUBJECTS' CERTAINTIES WITH THE LINEAR COMBINATION OF THE TERMS $(X_1 + X_2)$ AND $ABS(X_1 - X_2)$ OR THEIR EQUIVALENTS (Fisher's Z Transformation)

Source	S.S.	d.f.	M.S.	F	P
Subjects	9.380	11			
Sessions	.675	2	.338	9.057	<.01
Conditions	.770	2	.385	1.522	N.S.
Sessions x Conditions	.094	4	.024	<1	
Sessions x Subjects	.820	22	.037		
Conditions x Subjects	5.566	22	.253		
Sessions x Conditions x Subjects	1.796	44	.041		
TOTAL	19.095	107			

SIGNIFICANCE OF DIFFERENCES BETWEEN INDIVIDUAL PAIRS OF SESSION MEANS (Fisher's Z Transformation)

	MEAN	SESSION II	SESSION III
MEAN		1.045	1.067
SESSION			
I	.889	.156*	.178*
II	1.045		.022

Critical value of d = .086 at the .05 level.

* Significant at the .05 level.

THE LENS MODEL INDICES FOR A LINEAR COMBINATION OF THE TWO TERMS (X_1+X_2) AND $ABS(X_1-X_2)$ OR THEIR EQUIVALENTS.

Table 27a (i) Condition I (Untransformed)

THE MATCHING INDEX G

Subject	Session I	Session II	Session III
1	.811	.796	.884
2	1.000	.940	.905
3	.583	.950	.952
4	.909	.931	.982
5	.978	.717	.920
6	.868	.403	.926
7	.991	.994	.994
8	.979	.987	.997
9	.970	.831	.921
10	.991	.994	.995
11	.988	.952	1.000
12	.999	.999	.989

Table 27a(ii) Condition I (Untransformed)

GR_eR_s

Subject	Session I	Session II	Session III
1	.602	.634	.745
2	.559	.730	.801
3	.172	.516	.664
4	.673	.741	.826
5	.054	.441	.454
6	.103	.034	.250
7	.621	.681	.746
8	.670	.722	.732
9	.629	.517	.626
10	.707	.674	.743
11	.643	.677	.803
12	.626	.729	.807

THE LENS MODEL INDICES FOR A LINEAR COMBINATION OF THE TWO TERMS (X_1+X_2) AND $ABS(X_1-X_2)$ OR THEIR EQUIVALENTS.

Table 27a(iii) Condition I (Untransformed)

THE MATCHING INDEX C

Subject	Session		
	I	II	III
1	.236	.198	.190
2	.406	.577	.188
3	-.007	.225	-.010
4	.436	.495	.641
5	-.253	-.037	-.067
6	-.294	-.100	-.172
7	.341	.418	.441
8	.431	.443	.578
9	.570	.549	.264
10	.582	.538	.231
11	.434	.659	.469
12	.499	.720	.601

Table 27a(iv) Condition I (Untransformed)

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

Subject	Session		
	I	II	III
1	.060	.040	.028
2	.151	.127	.021
3	-.003	-.078	-.002
4	.112	.101	.095
5	-.122	-.012	-.019
6	-.140	-.043	-.053
7	.115	.118	.087
8	.130	.113	.118
9	.185	.173	.059
10	.163	.154	.046
11	.140	.176	.080
12	.168	.184	.099

THE LENS MODEL INDICES FOR A LINEAR COMBINATION OF THE TWO TERMS (X_1+X_2) AND ABS(X_1-X_2) OR THEIR EQUIVALENTS.

Table 27b(i) Condition II (X_2+2)

THE MATCHING INDEX G

Subject	Session I	Session II	Session III
1	.816	.709	.823
2	.998	.996	.993
3	.986	.967	.971
4	.937	.890	.962
5	.954	.883	.943
6	-.999	.637	.646
7	.248	.992	.953
8	.990	.990	.994
9	.938	.759	.715
10	.994	.810	.928
11	.932	.982	.960
12	.835	.921	.934

Table 27b(ii) Condition II (X_2+2)

GR_eR_s

Subject	Session I	Session II	Session III
1	.654	.571	.649
2	.684	.777	.812
3	.567	.638	.697
4	.603	.656	.759
5	.669	.661	.666
6	-.061	.046	.057
7	.046	.636	.630
8	.645	.661	.768
9	.118	.331	.296
10	.662	.592	.721
11	.648	.645	.690
12	.588	.645	.753

THE LENS MODEL INDICES FOR A LINEAR COMBINATION OF THE TWO TERMS (X_1+X_2) AND ABS(X_1-X_2) OR THEIR EQUIVALENTS.

Table 27b(iii) Condition II (X_2+2)

THE MATCHING INDEX C

Subject	Session		
	I	II	III
1	.490	.169	.255
2	.484	.652	.439
3.	.221	.106	.058
4	.288	.352	.652
5	.433	.471	.426
6	.276	-.047	.237
7	.011	.054	.086
8	.431	.542	.600
9	.134	.226	.279
10	.413	.365	.222
11	.582	.598	.454
12	.512	.301	.276

Table 27b(iv) Condition II (X_2+2)

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

Subject	Session		
	I	II	III
1	.096	.033	.046
2	.145	.142	.071
3	.080	.031	.012
4	.094	.088	.116
5	.125	.114	.091
6	.133	-.021	.076
7	.005	.017	.020
8	.139	.158	.111
9	.064	.086	.082
10	.117	.093	.041
11	.187	.178	.095
12	.147	.082	.047

THE LENS MODEL INDICES FOR A LINEAR COMBINATION OF THE TWO TERMS (X_1+X_2) AND $ABS(X_1-X_2)$ OR THEIR EQUIVALENTS.

Table 27c(i) Condition III (X_2 Reversed)

THE MATCHING INDEX G

Subject	Session		
	I	II	III
1	.896	.782	.775
2	.813	.773	.787
3	.885	.688	.850
4	.872	.963	.831
5	.887	.757	.805
6	.760	.774	.902
7	.979	.986	.890
8	.912	.790	.927
9	.540	.560	-.587
10	.820	.904	.855
11	.906	.894	.841
12	.824	.794	.823

Table 27c(ii) Condition III (X_2 Reversed)

GR_eR_s

Subject	Session		
	I	II	III
1	.405	.573	.654
2	.679	.661	.702
3	.573	.530	.598
4	.642	.705	.486
5	.638	.610	.537
6	.433	.577	.747
7	.520	.628	.624
8	.519	.543	.333
9	.174	.100	-.081
10	.642	.729	.760
11	.588	.584	.713
12	.660	.662	.751

THE LENS MODEL INDICES FOR A LINEAR COMBINATION OF THE TWO TERMS (X_1+X_2) AND ABS(X_1-X_2) OR THEIR EQUIVALENTS.

Table 27c(iii) Condition III (X_2 Reversed)

THE MATCHING INDEX C

Subject	Session I	Session II	Session III
1	.392	.212	.333
2	-.212	-.125	-.091
3	.212	.077	.414
4	.274	.352	.034
5	.168	.081	-.252
6	.346	-.039	.332
7	.318	.314	.134
8	.231	.291	-.284
9	.297	.275	-.210
10	.106	.259	.495
11	.228	.046	.360
12	.302	-.080	.101

Table 27c(iv) Condition III (X_2 Reversed)

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

Subject	Session I	Session II	Session III
1	.162	.054	.049
2	-.031	-.017	-.010
3	.069	.017	.089
4	.071	.089	.009
5	.046	.016	-.058
6	.126	-.010	.052
7	.122	.097	.029
8	.085	.082	-.085
9	.133	.188	-.067
10	.023	.050	.055
11	.074	.014	.052
12	.059	-.013	.009

Table 27d

ANALYSIS OF VARIANCE ON THE MATCHING INDEX G FOR THE LINEAR COMBINATION OF THE TERMS $(X_1 + X_2)$ AND $ABS(X_1 - X_2)$ OR THEIR EQUIVALENTS (Fisher's Z Transformation)

Source	S.S.	d.f.	M.S.	F	P
Subjects	30.830	11			
Sessions	.175	2	.087	<1	
Conditions	15.571	2	7.785	6.767	<.01
Sessions x Conditions	2.317	4	.579	1.249	N.S.
Sessions x Subjects	12.309	22	.560		
Conditions x Subjects	25.310	22	1.150		
Sessions x Conditions x Subjects	20.401	44	.464		
TOTAL	106.913	107			

SIGNIFICANCE OF DIFFERENCES BETWEEN INDIVIDUAL PAIRS OF CONDITION MEANS (Fisher's Z Transformation)

	MEAN	CONDITION II $(X_2 + 2)$	CONDITION III (X_2 Reversed)
MEAN		1.663	1.231
CONDITION			
I (Untransformed)	2.161	-.498	-.930*
II $(X_2 + 2)$	1.663		-.432

Critical value of d = .524 at the .05 level.

* Significant at the .05 level.

Table 27d

ANALYSIS OF VARIANCE ON THE MATCHING INDEX G FOR THE LINEAR COMBINATION OF THE TERMS $(X_1 + X_2)$ AND $ABS(X_1 - X_2)$ OR THEIR EQUIVALENTS (Fisher's Z Transformation)

Source	S.S.	d.f.	M.S.	F	P
Subjects	30.830	11			
Sessions	.175	2	.087	<1	
Conditions	15.571	2	7.785	6.767	<.01
Sessions x Conditions	2.317	4	.579	1.249	N.S.
Sessions x Subjects	12.309	22	.560		
Conditions x Subjects	25.310	22	1.150		
Sessions x Conditions x Subjects	20.401	44	.464		
TOTAL	106.913	107			

SIGNIFICANCE OF DIFFERENCES BETWEEN INDIVIDUAL PAIRS OF CONDITION MEANS (Fisher's Z Transformation)

	MEAN	CONDITION II $(X_2 + 2)$	CONDITION III (X_2 Reversed)
MEAN		1.663	1.231
CONDITION			
I (Untransformed)	2.161	-.498	-.930*
II $(X_2 + 2)$	1.663		-.432

Critical value of d = .524 at the .05 level.

* Significant at the .05 level.

Table 27e

ANALYSIS OF VARIANCE ON THE MATCHING INDEX C FOR THE LINEAR COMBINATION OF THE TERMS $(X_1 + X_2)$ AND ABS$(X_1 - X_2)$ OR THEIR EQUIVALENTS (Fisher's Z Transformation)

Source	S.S.	d.f.	M.S.	F	P
Subjects	1.723	11			
Sessions	.048	2	.024	1.159	N.S.
Conditions	.912	2	.456	3.351	N.S.
Sessions x Conditions	.181	4	.045	1.428	N.S.
Sessions x Subjects	.456	22	.021		
Conditions x Subjects	2.993	22	.136		
Sessions x Conditions x Subjects	1.393	44	.032		
TOTAL	7.706	107			

The correlations of the $(X_1 + X_2)$ term, or its equivalent, with the certainties of the subjects are shown in Tables 28 (a (i), b (i) and c (i)). The corresponding F ratios to test the significance of the correlations of these terms to the fit of the linear combination of $(X_1 + X_2)$ and ABS $(X_1 - X_2)$ are shown in Tables 28 (a (ii), b (ii) and c (ii)). An analysis of variance on the correlation coefficients showed them to be significantly higher in condition III (X_2 Reversed) than in the other two conditions (Table 28 d). This was not affected by the removal of any covariance effect of the ABS $(X_1 - X_2)$ term (or its equivalents) (Table 28 e).

Similarly, the correlations of subjects' certainties with the ABS $(X_1 - X_2)$ term or its equivalents are shown in Tables 29 (a (i), b (i) and c (i)), and the corresponding F ratios to test the contribution of this term to the above model are shown in Tables 29 (a (ii), b (ii) and c (ii)). The analysis of variance on these coefficients showed them to be significantly smaller in condition III (X_2 Reversed) than in the other two conditions. They were also significantly smaller in session I than in session II and III. The removal of possible covariance effects of the $(X_1 + X_2)$ term did not affect the pattern of significant differences between conditions, but the coefficients were now found to be significantly larger in session III than in sessions I and II.

Finally Tables 30 (a, b and c) show the F ratios computed to test the significance of the better fit of the linear combination of $X_1 + X_2$ with estimated weights over that of the term $(X_1 + X_2)$ or its equivalents.

THE CORRELATION OF SUBJECTS' CERTAINTIES WITH THE TERM (X_1+X_2) OR IT'S EQUIVALENT.

Table 28a(i) Condition I (Untransformed)

Subject	Session		
	I	II	III
1	-.847	-.884	-.878
2	-.515	-.495	-.434
3	-.001n.s.	-.578	-.695
4	-.830	-.859	-.798
5	-.058n.s.	-.676	-.506
6	-.135n.s.	.019n.s.	-.145
7	-.632	-.572	-.577
8	-.715	-.582	-.577
9	-.481	-.250n.s.	-.358
10	-.722	-.662	-.673
11	-.531	-.755	-.685
12	-.568	-.638	-.604
Model	(-.714)	(-.737)	(-.755)

F RATIOS TO TEST THE SIGNIFICANCE OF THE COMBINATION OF THE (X_1+X_2) TERM (OR IT'S EQUIVALENT) TO THE FIT OF THE LINEAR COMBINATION OF THE TERMS (X_1+X_2) AND ABS (X_1-X_2) OR THEIR EQUIVALENTS.

Table 28a(ii) Condition I (Untransformed)

Subject	Session		
	I	II	III
1	109.58	150.85	141.72
2	11.49	7.75	9.96
3	.52n.s.	16.66	35.69
4	91.21	112.78	92.09
5	.11n.s.	39.59	12.69
6	.74n.s.	.12n.s.	.21n.s.
7	24.45	13.72	18.80
8	41.93	15.71	18.30
9	8.94	.00n.s.	2.27n.s.
10	46.62	25.66	33.34
11	13.09	47.86	42.20
12	16.46	23.36	27.72
Model	(57.18)	(61.88)	(135.05)

THE CORRELATION OF SUBJECTS' CERTAINTIES WITH THE TERM (X_1+X_2) OR IT'S EQUIVALENT.

Table 28b(i) Condition II (X_2+2)

Subjects	Session		
	I	II	III
1	-.914	-.882	-.832
2	-.665	-.749	-.745
3	-.590	-.687	-.696
4	-.710	-.811	-.510
5	-.762	-.826	-.711
6	.058n.s.	-.006n.s.	-.005n.s.
7	.076n.s.	-.631	-.659
8	-.539	-.663	-.703
9	-.080n.s.	-.120n.s.	-.062n.s.
10	-.786	-.507	-.793
11	-.650	-.812	-.452
12	-.803	-.762	-.818
Model	(-.714)	(-.737)	(-.755)

F RATIOS TO TEST THE SIGNIFICANCE OF THE CONTRIBUTION OF THE (X_1+X_2) TERM (OR IT'S EQUIVALENT) TO THE FIT OF THE LINEAR COMBINATION OF THE TERMS (X_1+X_2) AND $ABS(X_1-X_2)$ OR THEIR EQIVALENTS.

Table 28b(ii) Condition II (X_2+2)

Subjects	Session		
	I	II	III
1	216.07	182.84	93.67
2	31.89	55.16	61.71
3	18.89	30.44	36.15
4	39.53	72.88	11.75
5	56.14	81.88	39.44
6	.09n.s.	.02n.s.	.03n.s.
7	.84n.s.	20.85	28.64
8	13.82	25.79	41.98
9	.09n.s.	.21n.s.	.25n.s.
10	65.53	7.97	68.83
11	27.79	80.29	6.77
12	.76.30	50.52	85.85
Model	(57.18)	(61.88)	(135.05)

THE CORRELATION OF SUBJECTS' CERTAINTIES WITH THE TERM (X_1+X_2) OR IT'S EQUIVALENT

Table 28c(i) Condition III(X_2 Reversed)

Subject	Session		
	I	II	III
1	-.509	-.813	-.892
2	-.953	-.948	-.942
3	-.732	-.840	-.741
4	-.837	-.767	-.617
5	-.812	-.890	-.705
6	-.648	-.826	-.858
7	-.555	-.639	-.729
8	-.633	-.763	-.367
9	-.340	-.186n.s.	-.140n.s.
10	-.893	-.883	-.935
11	-.729	-.718	-.893
12	-.914	-.926	-.964
Model	(-.714)	(-.737)	(-.755)

F RATIOS TO TEST THE SIGNIFICANCE OF THE CONTRIBUTION OF THE (X_1+X_2) TERM (OR IT'S EQUIVALENT) TO THE FIT OF THE LINEAR COMBINATION OF THE TERMS (X_1+X_2) AND ABS(X_1-X_2) OR THEIR EQUIVALENTS.

Table 28c(ii) ConditionIII (X_2 Reversed)

Subject	Session		
	I	II	III
1	13.68	84.16	170.66
2	424.23	400.82	340.59
3	46.46	126.30	49.36
4	95.96	52.84	25.34
5	79.17	175.39	41.90
6	33.11	94.21	117.91
7	15.58	22.20	44.62
8	25.77	59.25	5.61
9	7.24	1.94n.s.	1.03n.s.
10	167.78	140.94	300.24
11	44.89	39.60	163.27
12	214.52	256.10	552.87
Model	(57.17)	(61.88)	(135.05)

Table 28d

ANALYSIS OF VARIANCE ON THE CORRELATIONS OF SUBJECTS' CERTAINTIES WITH THE TERM ($X_1 + X_2$) OR ITS EQUIVALENT

(Fisher's Z Transformation)

Source	S.S.	d.f.	M.S.	F	P
Subjects	9.880	11			
Sessions	.260	2	.130	2.198	N.S.
Conditions	3.245	2	1.623	5.978	$<.01$
Sessions x Conditions	.017	4	.004	<1	
Sessions x Subjects	1.293	22	.059		
Conditions x Subjects	5.972	22	.271		
Sessions x Conditions x Subjects	2.169	44	.049		
TOTAL	22.836	107			

SIGNIFICANCE OF DIFFERENCES BETWEEN INDIVIDUAL PAIRS OF CONDITION MEANS

(Fisher's Z Transformation)

	MEAN	CONDITION II (X_2 + 2)	CONDITION III (X_2 Reversed)
MEAN		.755	1.086
CONDITION			
I (Untransformed)	.690	.065	.396*
II (X_2 + 2)	.755		.331*

Critical value of d = .255 at the .05 level.

*Significant at the .05 level.

THE CORRELATION OF SUBJECTS' CERTAINTIES WITH THE TERM ABS(X_1-X_2) OR IT'S EQUIVALENT.

Table 29a(i) ConditionI (Untransformed)

Subject	Session		
	I	II	III
1	-.245n.s.	-.288n.s.	-.403
2	-.511	-.838	-.921
3	-.323n.s.	-.373	-.442
4	-.398	-.506	-.614
5	-.041n.s.	-.143n.s.	-.274n.s.
6	-.053n.s.	-.078n.s.	-.278n.s.
7	-.506	-.676	-.693
8	-.512	-.739	-.668
9	-.679	-.691	-.703
10	-.574	-.576	-.596
11	-.653	-.492	-.685
12	-.583	-.698	-.768
Model	(-.697)	(-.749)	(-.773)

F RATIOS TO TEST THE SIGNIFICANCE OF THE CONTRIBUTION OF THE ABS (X_1-X_2) TERM (OR IT'S EQUIVALENT) TO THE FIT OF THE LINEAR COMBINATION OF THE TERMS (X_1+X_2) AND ABS(X_1-X_2) OR THEIR EQUIVALENTS.

Table 29a(ii) Condition I (Untransformed)

Subject	Session		
	I	II	III
1	.00n.s.	.27n.s.	4.70
2	11.18	91.71	257.62
3	6.03	2.26n.s.	6.08
4	4.15	9.35	34.24
5	.03n.s.	1.07n.s.	1.03n.s.
6	.01n.s.	.39n.s.	3.09n.s.
7	10.72	28.28	37.37
8	11.93	44.33	31.60
9	32.78	37.31	37.81
10	19.85	14.18	21.14
11	28.31	6.88	42.31
12	18.14	34.47	68.77
Model	(52.32)	(66.00)	(147.69)

THE CORRELATION OF SUBJECTS' CERTAINTIES WITH THE TERM ABS (X_1-X_2) OR IT'S EQUIVALENT.

Table 29b(i) Condition II (X_2+2)

Subject	Session		
	I	II	III
1	-.273n.s.	-.177n.s.	-.287
2	-.590	-.678	-.641
3	-.448	-.486	-.497
4	-.391	-.399	-.786
5	-.461	-.393	-.430
6	-.054n.s.	-.077n.s.	-.090n.s.
7	-.166n.s.	-.537	-.421
8	-.649	-.551	-.612
9	-.137n.s.	-.480	-.417
10	-.423	-.674	-.445
11	-.538	-.282	-.717
12	-.265n.s.	-.428	-.474
Model	(-.697)	(-.749)	(-.773)

F RATIOS TO TEST THE SIGNIFICANCE OF THE CONTRIBUTION OF THE ABS(X_1-X_2) TERM (OR IT'S EQUIVALENT) TO THE FIT OF THE LINEAR COMBINATION OF THE TERMS (X_1+X_2) AND ABS(X_1-X_2) OR THEIR EQUIVALENTS.

Table 29b(ii) Condition II (X_2+2)

Subject	Session		
	I	II	III
1	.00n.s.	5.57	.20n.s.
2	20.50	35.84	35.16
3	6.76	6.55	9.94
4	3.66n.s.	1.81n.s.	68.32
5	7.96	1.54n.s.	5.41
6	.07n.s.	.30n.s.	.41n.s.
7	1.91.n.s.	10.45	4.97
8	27.52	11.59	24.94
9	.67n.s.	13.48	9.98
10	5.54	27.80	6.69
11	13.78	.02n.s.	41.13
12	.11n.s.	3.14n.s.	9.64
Model	(52.32)	(66.00)	(147.49)

THE CORRELATION OF SUBJECTS' CERTAINTIES WITH THE TERM ABS (X_1-X_2) OR IT'S EQUIVALENT.

Table 29c(i) Condition III (X_2 Reversed)

Subject	Session I	Session II	Session III
1	-.229n.s.	-.247n.s.	-.239n.s.
2	-.279	-.277n.s.	-.270n.s.
3	-.311	-.144n.s.	-.289
4	-.333	-.531	-.221n.s.
5	-.349	-.239n.s.	-.222n.s.
6	-.136n.s.	-.242n.s.	-.426
7	-.396	-.515	-.343
8	-.314	-.241n.s.	-.205n.s.
9	-.030n.s.	-.001n.s.	-.002n.s.
10	-.273n.s.	-.461	-.374
11	-.343	-.359	-.335
12	-.285	-.299	-.332
Model	(-.697)	(-.749)	(-.773)

F RATIOS TO TEST THE SIGNIFICANCE OF THE CONTRIBUTION OF THE (X_1+X_2) TERM (OR IT'S EQUIVALENT) TO THE FIT OF THE LINEAR COMBINATION OF THE TERMS (X_1+X_2) AND ABS (X_1-X_2) OR THEIR EQUIVALENTS.

Table 29c(ii) Condition III (X_2 Reversed)

Subject	Session I	Session II	Session III
1	.40n.s.	.36n.s.	.25n.s.
2	.00n.s.	2.48n.s.	.16n.s.
3	1.00n.s.	5.22	.50n.s.
4	1.27n.s.	10.19n.s.	.09n.s.
5	1.85n.s.	1.94n.s.	.00n.s.
6	.28n.s.	.57n.s.	6.11
7	4.38	8.72	1.72n.s.
8	1.43n.s.	.16n.s.	.54n.s.
9	1.01n.s.	.25n.s.	.09n.s.
10	.00n.s.	5.39	3.71n.s.
11	1.85n.s.	1.11n.s.	1.10n.s.
12	.05n.s.	.55n.s.	1.17n.s.
Model	(52.32)	(66.00)	(147.69)

Table 29d

ANALYSIS OF VARIANCE ON THE CORRELATIONS OF SUBJECTS' CERTAINTIES WITH THE TERM ABS ($X_1 - X_2$) OR IT'S EQUIVALENT

(Fisher's Z Transformation)

Source	S.S.	d.f.	M.S.	F	P
Subjects	2.492	11			
Sessions	.371	2	.185	8.434	<.01
Conditions	1.821	2	.910	9.008	<.01
Sessions x Conditions	.217	4	.054	2.761	N.S.
Sessions x Subjects	.484	22	.022		
Conditions x Subjects	2.224	22	.101		
Sessions x Conditions x Subjects	.863	44	.020		
TOTAL	8.672	107			

SIGNIFICANCE OF DIFFERENCES BETWEEN INDIVIDUAL PAIRS OF CONDITION MEANS

(Fisher's Z Transformation)

	MEAN	CONDITION II (X_2 + 2)	CONDITION III (Reversed)
MEAN		.477	.293
CONDITION			
I (Untransformed)	.609	-.132	-.316*
II (X_2 + 2)	.477		-.184*

Critical value of d = .155 at the .05 level.

SIGNIFICANCE OF DIFFERENCES BETWEEN INDIVIDUAL PAIRS OF SESSION MEANS

(Fisher's Z Transformation)

	MEAN	SESSION II	SESSION III
MEAN		.469	.526
SESSION			
I	.384	.085*	.142*
II	.469		.057

Critical value of d = .073 at the .05 level.

* Significant at the .05 level.

Table 29e

ANALYSIS OF VARIANCE ON THE CORRELATIONS OF SUBJECTS' CERTAINTIES WITH THE TERM ABS(X_1 - X_2) OR IT'S EQUIVALENT WHEN THE EFFECT OF THE TERM (X_1 + X_2) OR IT'S EQUIVALENT IS PARTIALLED OUT

(Fisher's Z Transformation)

Source	S.S.	d.f.	M.S.	F	P
Subjects	3.154	11			
Sessions	.624	2	.312	7.517	<.01
Conditions	3.673	2	1.837	15.734	<.01
Sessions x Conditions	.241	4	.060	1.953	N.S.
Sessions x Subjects	.913	22	.042		
Conditions x Subjects	2.568	22	.117		
Sessions x Conditions x Subjects	1.357	44	.031		
TOTAL	12.530	102			

SIGNIFICANCE OF DIFFERENCES BETWEEN INDIVIDUAL PAIRS OF CONDITION MEANS

(Fisher's Z Transformation)

	MEAN	CONDITION II (X_2 + 2)	CONDITION III (X_2 Reversed)
MEAN		.365	.077
CONDITION			
I (Untransformed)	.523	-.158	-.446*
II (X_2 + 2)	.365		-.288*

Critical value of d = .236 at the .05 level.

SIGNIFICANCE OF DIFFERENCES BETWEEN INDIVIDUAL PAIRS OF SESSION MEANS

(Fisher's Z Transformation)

	MEAN	SESSION II	SESSION III
MEAN		.272	.429
SESSION			
I	.263	.009	.166*
II	.272		.157*

Critical value of d = .141 at the .05 level.

* Significant at the .05 level.

Table 29e

ANALYSIS OF VARIANCE ON THE CORRELATIONS OF SUBJECTS' CERTAINTIES WITH THE TERM ABS(X_1 - X_2) OR IT'S EQUIVALENT WHEN THE EFFECT OF THE TERM (X_1 + X_2) OR IT'S EQUIVALENT IS PARTIALLED OUT

(Fisher's Z Transformation)

Source	S.S.	d.f.	M.S.	F	P
Subjects	3.154	11			
Sessions	.624	2	.312	7.517	<.01
Conditions	3.673	2	1.837	15.734	<.01
Sessions x Conditions	.241	4	.060	1.953	N.S.
Sessions x Subjects	.913	22	.042		
Conditions x Subjects	2.568	22	.117		
Sessions x Conditions x Subjects	1.357	44	.031		
TOTAL	12.530	102			

SIGNIFICANCE OF DIFFERENCES BETWEEN INDIVIDUAL PAIRS OF CONDITION MEANS

(Fisher's Z Transformation)

		CONDITION	
	MEAN	II (X_2 + 2)	III (X_2 Reversed)
MEAN		.365	.077
CONDITION			
I (Untransformed)	.523	-.158	-.446*
II (X_2 + 2)	.365		-.288*

Critical value of d = .236 at the .05 level.

SIGNIFICANCE OF DIFFERENCES BETWEEN INDIVIDUAL PAIRS OF SESSION MEANS

(Fisher's Z Transformation)

		SESSION	
	MEAN	II	III
MEAN		.272	.429
SESSION			
I	.263	.009	.166*
II	.272		.157*

Critical value of d = .141 at the .05 level.

* Significant at the .05 level.

Table 28e

ANALYSIS OF VARIANCE ON THE CORRELATIONS OF SUBJECTS' CERTAINIES WITH THE TERM $(X_1 + X_2)$ OR IT'S EQUIVALENT WHEN THE EFFECT OF THE TERM ABS$(X_1 - X_2)$ OR IT'S EQUIVALENT IS PARTIALLED OUT (Fisher's Z Transformation)

Source	S.S.	d.f.	M.S.	F	P
Subjects	10.461	11			
Sessions	.200	2	.100	1.505	N.S.
Conditions	3.686	2	1.843	6.947	<.01
Sessions x Conditions	.036	4	.009	<1	
Sessions x Subjects	1.464	22	.067		
Conditions x Subjects	5.837	22	.265		
Sessions x Conditions x Subjects	2.343	44	.053		
TOTAL	23.027	107			

SIGNIFICANCE OF DIFFERENCES BETWEEN INDIVIDUAL PAIRS OF CONDITION MEANS (Fisher's Z Transformation)

	MEAN	CONDITION II $(X_2 + 2)$	CONDITION III (X_2 Reversed)
MEAN		.690	1.044
CONDITION			
I (Untransformed)	.622	.068	.422*
II $(X_2 + 2)$	.690		.354

Critical value of d = .356 at the .05 level.

* Significant at the .05 level.

F RATIOS TO TEST THE SIGNIFICANCE OF THE BETTER FIT OF THE LINEAR COMBINATION OF X_1 AND X_2 WITH ESTIMATED WEIGHTS OVER THAT OF (X_1+X_2) OR IT'S EQUIVALENT

Table 30a Condition I (Untransformed)

Subject	Session		
	I	II	III
1	.00n.s.	1.86	.11
2	9.86	5.59	1.59
3	8.04	14.32	1.19n.s.
4	7.62	4.52	.23n.s.
5	7.98	14.00	19.05
6	.19n.s.	12.07	3.54n.s.
7	6.60	2.02n.s.	2.89n.s.
8	4.66	.43n.s.	.14n.s.
9	.05n.s.	.06n.s.	.46n.s.
10	3.95n.s.	4.63	3.63n.s.
11	.00n.s.	.65n.s.	.02n.s.
12	.76n.s.	.46n.s.	2.34n.s.
Model	(1.10)	(1.00)	(2.21)

Table 30b COndition II (X_2+2)

Subject	Session		
	I	II	III
1	7.95	.11n.s.	.31n.s.
2	..06n.s.	2.79	2.19n.s.
3	.37n.s.	.04n.s.	1.98n.s.
4	.24n.s.	.02n.s.	1.23n.s.
5	4.91	.32n.s.	1.29n.s.
6	.65n.s.	.08n.s.	1.72n.s.
7	5.15	4.85	.00n.s.
8	7.37	1.51n.s.	.20n.s.
9	.66n.s.	.13n.s.	3.19n.s.
10	.19n.s.	.01n.s.	12.46
11	.06n.s.	5.03	6.25
12	11.79	15.80	.00n.s.
Model	(1.10)	(1.00)	(2.00)

F RATIOS TO TEST THE SIGNIFICANCE OF THE BETTER FIT OF THE LINEAR COMBINATION OF X_1 AND X_2 WITH THE ESTIMATED WEIGHTS OVER THAT OF (X_1+X_2) OR IT'S EQUIVALENT.

Table 30c Condition III (X_2 Reversed)

Subject	Session		
	I	II	III
1	13.88	.27n.s.	.08n.s.
2	.69n.s.	.09n.s.	8.47
3	.74n.s.	.05n.s.	.25n.s.
4	3.46n.s.	4.48	2.63n.s.
5	6.46	.24n.s.	2.79n.s.
6	.52n.s.	.00n.s.	2.94n.s.
7	8.75	.13n.s.	2.40n.s.
8	2.32n.s.	4.13	2.82n.s.
9	2.59n.s.	.00n.s.	.90n.s.
10	.00n.s.	7.53	.00n.s.
11	3.02n.s.	.08n.s.	1.36n.s.
12	.69n.s.	.62n.s.	2.26n.s.
Model	(1.10)	(1.00)	(2.00)

Conclusions and Discussions

As in experiment I, we find that some subjects did not learn to use the cues in a valid manner throughout the three sessions of some conditions. Subject six and nine are worthy of note. Subject six performed markedly above chance level only in Condition III (X_2 Reversed) and subject six only in Condition I (Untransformed). It seems that this poor performance was probably a result of a number of factors, lack of motivation, a possible misunderstanding of the task (though this was tested by questioning, and these subjects appeared to show no lower degree of comprehension than others), and possibly a difficulty in discarding the rules generated in one condition on moving to a condition where these rules were no longer appropriate.

The significant interaction effect of conditions and sessions on the total number of correct classifications by the subjects seems to be due to an improvement in accuracy over sessions taking place in Condition I (Untransformed) and II (X_2 + 2) but no such improvement taking place in Condition III (X_2 Reversed) where, if anything accuracy deteriorated over sessions. By the third session, these processes resulted in a significantly worse performance in Condition III (X_2 Reversed) than in the other two conditions. These trends were also shown in the other accuracy index, the correlations of subjects' certainties with those of the normative model, though here they did not reach significance. It would appear then, that subjects did find the condition in which a negative correlation existed between cues in one disease population more difficult that ones in which a positive correlation existed. In view of the research showing that human subjects find negative relationships more difficult to deal with, we might expect that this lower accuracy may be due to a lack of utilization of the fact that the two populations to be discriminated have different levels of cue intercorrelation.

In all three conditions the certainties of all but one or two subjects are significantly correlated with a linear combination of X_1 and X_2 with estimated weights, or weights set equal in magnitude. By the third session in each condition the fit of the model with estimated weights is only significantly better than that of the model with "a priori" weights for about the same number of subjects. The subjects, in Condition I (Untransformed) and II ($X_2 + 2$) had again claimed that they based their decisions on the sum or mean of the two cues but in Condition III (X_2 Reversed) reported basing their decisions on $(X_1 - X_2)$ which has in this condition the same validity as $(X_1 + X_2)$ in the other conditions. The above findings confirm these verbal reports.

The certainties of subjects in Condition III (X_2 Reversed) were generally better predicted by a linear combination of the two cues (with either estimated or "a priori" weights) than their certainties in the other conditions. Since the degree of matching, G, was not found to differ over conditions and the linear dependencies R_e of all conditions were equal, we might between cues in one disease population more difficult that ones in which a positive correlation existed. In view of the research showing that human subjects find negative relationships more difficult to deal with, we might expect that this lower accuracy may be due to a lack of utilization of the fact that the two populations to be discriminated have different levels of cue intercorrelation.

In all three conditions the certainties of all but one or two subjects are significantly correlated with a linear combination of X_1 and X_2 with estimated weights, or weights set equal in magnitude. By the third session in each condition the fit of the model with estimated weights is only significantly better than that of the model with "a priori" weights for about the same number of subjects. The subjects, in Condition I (Untransformed) and II ($X_2 + 2$) had again claimed that they based their decisions on the sum or mean of the two cues but in Condition III (X_2 Reversed) reported basing their decisions on $(X_1 - X_2)$ which has in this condition the same validity as $(X_1 + X_2)$ in the other conditions. The above findings confirm these verbal reports.

The certainties of subjects in Condition III (X_2 Reversed) were generally better predicted by a linear combination of the two cues (with either estimated or "a priori" weights) than their certainties in the other conditions. Since the degree of matching, G, was not found to differ over conditions and the linear dependencies R_e of all conditions were equal, we might expect that part of subjects' accuracy which could be put down to their use of cues in an appropriate linear manner i.e. GR_eR_s to be higher in Condition III (X_2 Reversed) than the other conditions. In fact, performance is worse in Condition III (X_2 Reversed), (though not significantly so in terms of the correlation of subjects' certainties with those of the normative model), than in the other two conditions. We are led to suspect then, that subjects in Conditions I (Untransformed) and II ($X_2 + 2$) are using the cues in a non-linear way and thereby improving their performance. Indeed C, the index of non-linear matching is significantly lower in Condition III (Reversed) than in the other two conditions.

Many subjects in Condition I (Untransformed) reported that they considered the absolute difference between the cues i.e. ABS ($X_1 - X_2$) when making their decisions. Similarly in Condition II ($X_2 + 2$) many reported considering the nearness of $X_2 - X_1$ to 2, i.e. ABS ($X_1 - (X_2 - 2)$). No subject reported using such non-linear cues in Condition III (X_2 Reversed) but, since ABS ($X_1 - X_2$) and ABS ($X_1 - (X_2 - 2)$) are of equal predictive power, it was decided that the term ABS (($X_1 + X_2$) - 17) i.e. the nearness of the sum of the cues of 17, would be considered in this condition as it is equivalent to the above two terms in its predictive power. On considering the correlation of subjects' certainties with these non-linear terms it is apparent that these are small and insignificant in Condition III (X_2 Reversed). They are in fact significantly smaller in this condition, whether or not the possible covariance effects of the linear terms $(X_1 + X_2)$, $(X_1 + X_2 - 2)$ and $(X_1 - X_2)$ are removed.

When the appropriate linear and non-linear terms were linearly combined in an attempt to predict the certainties of subjects in each condition, it was found that the resulting multiple correlations did not differ over conditions. Thus the more predictable nature of subjects' certainties in Condition III (X_2 Reversed) by an "a priori" linear combination of X_1 and X_2 is compensated for by the inclusion of these non-linear terms. Though the linear terms contribute significantly to the fit of these models for all but one or two subjects in each condition, the non-linear terms though contributing to the fit of the model for most subjects in Conditions I (Untransformed) and II ($X_2 + 2$) do not do so with consistency for any of the subjects in Condition III (X_2 Reversed). The Lens Model index G showing the matching of subjects' utilization of these linear and non-linear terms with that of the normative model, show such matching to be significantly less in Condition III (X_2 Reversed) than in the other conditions, presumably as a result of the lack of utilization of the non-linear term in this condition. The index of non model matching C did not vary significantly over conditions.

Rank order equivalents of the foregoing indices were not computed for the data collected in this experiment. Since the tasks forming the three conditions were essentially identical, for example, in R_e and the extent to which non-linear terms were valid, it is not likely that the possible non-linearity in the certainty values would have a differential effect on the Lens Model indices for each condition. It seems unlikely then that such an analysis could tell us much more about the processes involved than the indices already discussed, using least squares regression theory.

DISCUSSION AND CONCLUSIONS

The optimal statistical approach to situations requiring the discrimination of the members of two or more multivariate normal populations was formulated by Fisher (1936) and was used throughout the present study as a normative model against which the behaviour of human subjects was compared. In cases where the covariance matrices are equal i.e. the shapes of the multivariate normal population to be distinguished are identical, the optimal procedure is to linearly combine the scores on the cues (with suitably derived weights) and assign those observations (patients), for which the value of this combination is greater than some rationally derived constant, to one population and all other observations to the other population (in the two population case).

The behaviour of many subjects in conditions of this sort seemed to be very well predicted by a model which weighted the two cues equally in a linear combination and based its responses on the value of the result. The verbal reports of many subjects of the way their decisions were made, were exactly parallel to the linear discriminant function:- they added the two cues together and compared the result with a constant, if it was greater than that constant they assigned the patient to one disease population otherwise he was assigned to the other.

The degree of cue intercorrelation (equal for both populations making up any one condition, but varying over conditions) did not appear to have much effect on either the accuracy of subjects' responses or on the degree to which their certainty in their responses could be predicted by a linear combination of the cues. Such an effect had been found by Naylor and Schenck (1968) in a task with a continuous criterion and one explanation of this effect which was put forward was that in tasks with high intercorrelations between the predictor cues it is possible to disregard one or more cue as much of the information contained in it is also included in other cues. In the present experiments, however, there was no evidence of subjects relying only on one of the two cues and if one of the cues was significantly related to their responses then the other one generally was also. The large intersubject differences in learning which were evident in the experiments of the present study may, however, be obscuring many effects which would show clearly with less inter subject variance.

Many subjects did not learn to use the predictor cues in a valid way throughout the length of the experiment possibly due to a low level of motivation since they were not in general volunteers, but students taking part in order to fulfil a course requirement. Some of these subjects, though they were not using the cues in a valid way, were quite consistent in the way the decisions were made.

It is possible that this consistent but invalid behaviour was due to some kind of "superstitious" learning (Skinner, 1948); that the feedback provided, reinforced some "hypothesis", which a subject was considering, for a sufficiently large number of trials for the subject to accept it as valid. A similar effect has been reported previously by Azuma and Cronbach (1966) in a one population task and it would seem that such behaviour is particularly likely to result from tasks in which even the optimal decision policy does not result in perfect accuracy thus diminishing the power of feedback.

We see then that in these tasks as in that of Vlek and Van der Heijden (1970) who used spatially presented cues, the responses of subjects who learn to use the cues in a way which is at all valid can be reasonably well predicted by a linear combination of the two cues in a manner very similar to the statistical technique Linear Discriminant Function analysis.

When the populations whose members are to be distinguished do not have equal covariance matrices i.e. patterns or relations exist in the scores of the members of one population which do not exist in the scores of the members of the other, a linear combination of the cues no longer provides the optimal decision policy. The normative model in such situations includes curvilinear terms in the squares of the cues, configural terms in their cross products as well as the cues themselves, all suitably weighted in a linear combination, and bases its decision on the value of the result. Such differences in covariance matrices seem particularly likely to occur in medical diagnosis situations, in view of the often made claim that the "pattern" of the scores of a patient on a series of tests is important in assigning him to one diagnostic category rather than another. It can be seen, then, that tasks which require judges to discriminate the members of one population where the predictor cues are highly intercorrelated from the members of another population where no such intercorrelations exist, are intrinsically non-linear. Unlike previous experiments on the learning of non-linear cue utilization (c.f. Hammond and Summers 1968) experiments based on the situation outlined above, would not be adding non-linearly to a basically linear task. Their non-linearity would follow quite logically from the different parameter values of the populations to be discriminated.

A great deal of effort was put into finding "real life" situations which were intrinsically non-linear in an attempt to discover to what extent humans process information in such situations in a non-linear manner. Results from such studies have, in general been disappointing, very little evidence of a strong dependence on non-linear processes was found. However, these tasks were only "felt" to be non-linear, no statistical investigation of the tasks took place and the possibility remains that assumption of non-linearity may not have been justified, in which case it would be hardly surprising that the judges did not seem to be processing the information in a non-linear way. We see now, however, that it is possible to generate tasks in the laboratory, which do not differ too greatly from real life tasks, yet which are known not only to be non-linear, but non-linear in a way which can be specified exactly.

It was found that human subjects did appear to learn to use cues in a non-linear manner in discriminating between the members of a population in which the cues were highly positively correlated and the members of a population in which the cues were independent. Rather than replace the linear combination of the two cues however, this non-linearity seemed to take the form of the inclusion of a non-linear transformation of the cues into the decision processes of the subjects and supplemented rather than replaced the utilization of the cues in a linear manner. Subjects reported that they consider the absolute difference between the two cues, and a multiple regression of the subjects' responses on this term and a term representing the sum (or average) of the two cues yielded high values of the multiple correlation coefficient which were, in general, reduced significantly by the removal of either term.

When the cues in one population were less highly positively correlated, (though still correlated to a degree which could have been perceived by subjects if this had been their major task (c.f. Beach and Scopp, 1966, Erlick and Mills, 1967)),and the cues in the other population were independent, it appeared that some of the subjects did not perceive the non-linear nature of the task and did not use the cues in a non-linear manner. The inclusion of the absolute difference term in a multiple regression with the subjects' responses as predicted variables, did not, for these subjects significantly increase the value of the multiple correlation coefficient above its value with only the linear term, the sum (or average) of the two cues. Some subjects did, however, appear to be weighting the non-linear term significantly but on average there was no evidence of greater dependence on this cue in this condition, than in a condition with equal covariance matrices, at least when possible covariance effects of the linear term had been removed. Also, since some of the subjects had previously had experience of the condition with a high correlation between the cues in one population, it could be that this led them to look for possible differences in the pattern of scores of the member of the two populations, which otherwise would have gone unnoticed.

The weight given to the absolute difference between the two cues in this study would seem to imply that the subjects had perceived that the two scores of a member of one population tended to be of a similar magniture whereas no similarity was apparent in the magnitude of the two scores of members of the other population. Unlike the cross product and square terms of the normative model this absolute difference term is only valid in a limited number of the sorts of situations we are considering. Particularly, the cues must have similar distributions in terms of mean and standard deviations. By transforming the cues in various (rather trivial at a statistical level) ways, it is possible to invalidate this particular non-linear term, and to discover to what extent the use of the cues in a non-linear way is dependent upon for example the particular ranges of the cues involved.

The addition of a constant to the scores of the members of both populations on one of the cues caused no apparent change in the accuracy of subjects' judgements, nor did they appear to be relying more upon a simple linear combination of the two cues than in a condition with untransformed cues. Many subjects reported considering the extent to which one cue minus the other came close to some constant value, and on the inclusion of a term expressed as ABS ($(X_1 - X_2)$ - Constant), equal in predictive power (with regard to the normative model) as was the absolute difference term in the condition with untransformed cues, in a regression analysis with the mean or sum of the cues, these terms predicted the responses of subjects as well as the equivalent terms in the condition with untransformed cues. It would appear then that subjects are still able to perceive and use differences in the pattern of scores when the cues do not have exactly the same distribution, a result which is directly parallel to the lack of effect of the additive constant in research into learning linear functions. (c.f. Eisler and Spolander (1970); De Klerk, Oppe and Truijens, 1972).

The situation was found to be somewhat different when one cue was transformed so that what were previously high scores on this cue were now made low and vice versa. Though the range of scores on each cue remained constant, what had been a positive correlation between the cues in one population now became negative though the cues retained their previous range. The accuracy of performance in this situation was significantly lower than in the untransformed condition in later sessions, but the dependence of subjects upon a linear combination of the two cues was significantly higher. No evidence could be found of non-linear cue utilization which is not altogether surprising in view of the problems subjects have with negatively related cues in other situations (c.f., Brehmer, 1971; Eisler and Spolander, 1970; Erlick and Mills, 1967; Naylor and Clark, 1968). It would appear, then, that subjects did not perceive that the correlation between the scores of members of the two populations differed, or, if they did perceive it, they were unable to make use of it.

The approach of human judges to the type of task discussed above would appear to differ quite fundamentally from that of the statistical model. The statistical model can accommodate all the above tasks by suitable changes in the value of the individual parameter (these values are logically derived from the statistical properties of the populations involved). The approach of a human subject however, would appear to be far more "ad hoc". He would appear to formulate one "rule" for one situation and another "rule" for another situation. His decision as to what terms should be considered is highly situation dependent and the "rule" he develops for one task may be completely invalid in another task which does not differ from it in any great degree. We have seen how this situation dependent approach leads subjects to use the absolute difference between the cues, a non-linear term, in one condition but to base their responses only on a linear combination of the cues in a condition which differs from the first only at a trivial level, from a statistical point of view. The statistical model however, performs equally well in both these tasks, and does so, not by the inclusion of some terms and the exclusion of others, but simply by altering its parameter values.

It is unlikely that judges in a "real-life" situation are ever going to have conditions so condusive to the learning of valid decision processes as were provided to the subjects in the present study. Yet, even in these ideal situations, the non-linear nature of tasks with negatively correlated cues or cues with only moderate degrees of correlation, was often not noticed by the subjects. Indeed even in tasks which were linear the performance of some subjects was no better than if they were responding randomly. Most diagnostic decisions in "real life" are not made on the basis of only two cues as in the present tasks nor are there usually only two possible diagnoses. These increases in number of variables and number of decision categories increase the possible location of differences in the relations between cues disproportionately. In fact when a judge has, say, sixteen cues to consider in determining which of possibly ten or twenty or more diseases a patient has, he would probably do as much as possible to simplify his decision processes rather than make them more complex by the inclusion of non-linear terms. The subjects in the present experiment all used simple linear and non-linear transformations of the cues, at no time did anyone report taking squares or roots or any other such complex term in making his decisions. It would be interesting to study learning tasks with more cues and decision categories than the present ones, but it is doubtful whether practical difficulties, particularly the length of time it would take to learn to make predictions in an accurate way, would be outweighed by the benefits derived from such studies.

In summary then, it is hoped that this study has gone some way to fill what was a glaring gap in the psychological research into human decision making i.e. the ability of subjects to learn to use probabilistic cues in identifying the members of more than one multivariate normal population. It has been shown that in some situations of this sort, subjects clearly combine the information from cues in a non-linear way (without the non-linearity of the situations being pointed out to the subject as in the study of Hammond and Summers (1965) strengthening our conviction that such non-linear processes could be developed in real life given suitable conditions. However, in order for such situations to be identified, a great deal more work must be carried out on the ecological side of Brunswick's Lens Model.

BIBLIOGRAPHY

Anderson, N.H. A Simple Model for Information Integration. in Abelson, R.P. et al. (eds). Theories of Cognitive Consistency. Chicago: Rand McNally, 1968.

Anderson, N.H. Looking for Configurality in Clinical Judgement. Psychological Bulletin. 1972, 78, 93 - 102.

Anderson,T.W. An introduction to Multivariate Statistical Analysis. New York: John Wiley & Sons Inc., 1958.

Azuma, H. & Cronbach, L.J. Cue Response Correlations in the Attainment of a Scalar Concept. American Journal of Psychology. 1966, 79, 38 - 49.

Beach, L.R. & Scopp, T.S. Inferences about Correlations. Psychonomic Science. 1966, 6, 253 - 254.

Bischof, L.S. Interpreting Personality Theories. Japan: Harper and Row, and John Weatherhill, Inc. 1964.

Bjorkman, M. Learning of Linear Functions : Comparison between a Positive and Negative Slope. Rep. Psychol. Lab., Univ. Stockholm, No. 153. 1965.

BIBLIOGRAPHY

Anderson, N.H. A Simple Model for Information Integration. in Abelson, R.P. et al. (eds). Theories of Cognitive Consistency. Chicago: Rand McNally, 1968.

Anderson, N.H. Looking for Configurality in Clinical Judgement. Psychological Bulletin. 1972, 78, 93 - 102.

Anderson, T.W. An introduction to Multivariate Statistical Analysis. New York: John Wiley & Sons Inc., 1958.

Azuma, H. & Cronbach, L.J. Cue Response Correlations in the Attainment of a Scalar Concept. American Journal of Psychology. 1966, 79, 38 - 49.

Beach, L.R. & Scopp, T.S. Inferences about Correlations. Psychonomic Science. 1966, 6, 253 - 254.

Bischof, L.S. Interpreting Personality Theories. Japan: Harper and Row, and John Weatherhill, Inc. 1964.

Bjorkman, M. Learning of Linear Functions : Comparison between a Positive and Negative Slope. Rep. Psychol. Lab., Univ. Stockholm, No. 153. 1965.

Brehmer, B. Subjects Ability to use Functional Rules.
Umea Psychological Reports. 43. 1971.

Brunswick, E. Perception and Representative Design of
Psychological Experiments.
Berkeley: University of California Press. 1956.

Brunswick, E. & Herma, H. Probability Learning of Perceptual
Cues in the Establishment of a Weight Illusion.
Journal of Experimental Psychology. 1951, 41, 281 - 290.

Brunswick, E. & Reiter, L. Eindruckscharaktere
Schematisierter Gesichter.
Z. Psychol., 1934, 142, 67 - 134.

Carroll, J.D. Functional Learning : The Learning of Continuous
Functional Mappings Relating Stimulus and Response Continua.
Educational Teaching Service, Princeton, Research Bulletin
RB - 63 - 26. 1963.

Dawes, R.M. & Corrigan, B. Linear Models in Decision Making.
Psychological Bulletin. 1974, 81, 95 - 106.

De Klerk, L.F.W., Oppe, S. & Truijens, C.L. Learning of Positive
Linear Functions under Two Different Feedback Conditions.
Acta Psychologica. 1972, 36, 421 - 430.

Dudycha, L.W. & Naylor, J.C. Characteristics of the Human Inference Process in Complex Choice Behavior Situations.
Organizational Behavior and Human Performance. 1966, 1, 110 - 128.

Einhorn, H.J. The use of Nonlinear, Noncompensatory Models in Decision Making.
Psychological Bulletin. 1970, 73, 221 - 230.

Eisenbeis, R.A. & Avery, R.B. Discriminant Analysis and Classification Procedures - Theory and Applications.
Massachusetts: D.C. Heath & Co. 1972.

Eisler, H. & Spolander, K. On the Sign of Slope in the Learning of Linear Functions.
Scandinavian Journal of Psychology. 1970, 11, 176 - 184.

Erlick, D.E. Human Estimates of Statistical Relatedness.
Psychonomic Science. 1966, 5, 365 - 366.

Erlick, D.E. & Mills, R.G. Perceptual Quantifications of Conditional Dependency.
Journal of Experimental Psychology. 1967, 73, 9 - 14.

Fisher, R.A. The Use of Multiple Measurements in Taxonomic Problems.
Annals of Eugenics. 1936, 7, 179 - 188.

Goldberg, L.R. The Effectiveness of Clinicians' Judgements: The Diagnosis of Organic Brain Damage from the Bender-Gestalt Test. Journal of Consulting Psychology. 1959, 23, 25 - 33.

Goldberg, L.R. Diagnosticians versus Diagnostic Signs : The Diagnosis of Psychosis versus Neurosis from the M.M.P.I. Psychological Monographs. 1965, 79, (a, Whole No. 602).

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

Goldberg, L.R. Five Models of Clinical Judgement: The Empirical Comparison between Linear and Nonlinear Representation of the Human Inference Process.
Organizational Behavior and Human Performance. 1971, 6, 458 - 479.

Gray, C.W., Barnes, C.B. & Wilkinson, E.F. The Process of Prediction as a Function of the Correlation between Two Scaled Variables.
Psychonomic Science. 1965, 3, 231 - 232.

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 and Sons Inc., 1968.

Green, D.M. Detection of Multiple Signals in Noise.
In Swets, J.A. (ed.) Signal Detection and Recognition by Human Observers
Contemporary Readings.
New York: John Wiley & Sons, Inc., 1964.

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 Nonlinear cues.
Psychological Review. 1965, 72, 215 - 224.

Hoffman, P.J. The Paramorphic Representation of Clinical Judgement.
Psychological Buttetin. 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 and Sons Inc., 1968.

Hoffman, P.S., 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.

Jenkins, H.M. & Ward, W.C. Judgment of Contingency between Responses and Outcomes. Psychological Monograph. 1965, 79, 1 - 17.

Johnston, R. & MacNeal, B.F. Statistical versus Clinical Prediction: Length of Neuropsychiatric Hospital Stay. Journal of Abnormal Psychology. 1967, 72, 335 - 340.

Kleinmuntz, B. The Processing of Clinical Information by Man and Machine. In Kleinmuntz, B. (ed.) Formal Representation of Human Judgment. New York: John Wiley and Sons, Inc., 1968.

Lichtenstein, S. & Feeney, G.J. The Importance of the Data Generating Model in Probability Estimation. Organizational Behavior and Human Performance. 1968, 3, 62 - 67.

Luft, J. Implicit Hypotheses and Clinical Predictions. Journal of Abnormal and Social Psychology. 1950, 15, 756 - 760.

McArthur, C. Analyzing the Clinical Process. Journal of Counselling Psychology. 1954, 1, 203 - 208.

Meehl, P.E. Clinical versus Actuarial Prediction: A Theoretical Analysis and a Review of the Evidence.
Mineapolis: University of Minnesota Press. 1954.

Meehl, P.E. A Comparison of Clinicians with Five Statistical Methods of Identifying Psychotic M.M.P.I. Profiles.
Journal of Counselling Psychology. 1959, 6, 102 - 109.

Meehl, P.E. The Cognitive Activity of the Clinician.
American Psychologist. 1960, 15, 19 - 27.

Mood, A.M. & Graybill, F.A. Introduction to the Theory of Statistics (Second Edition)
Japan: McGraw - Hill and Kogakusha. 1963.

Naylor, J.C. & Clark, R.D. Intuitive Inference Strategies in Learning Tasks as a Function of Validity, Magnitude & Sign.
Organizational Behavior and Human Performance. 1968, 3, 378 - 399.

Naylor, J.C. & Schenck, E.A. The Influence of Cue Redundancy upon the Human Inference Process for Tasks of Varying Degrees of Predictability.
Organizational Behavior and Human Performance. 1968, 3, 47 - 61.

Noble, B. Applied Linear Algebra.
New Jersey: Prentice Hall. 1969.

Oskamp, S. The Relationship of Clinical Experience and Training Methods to Several Criteria of Clinical Prediction. Psychological Monographs. 76, (28, whole No. 547). 1962.

Parker, C.A. As a Clinician thinks .....
Journal of Counselling Psychology. 1958, 5, 253 - 262.

Rodwan, A.S. & Hake, H.W. The Discriminant Function as a Model for Perception.
American Journal of Psychology. 1964, 77, 380 - 392.

Rorer, L.G., Hoffman, P.S., Dickman, H.D. & Slovic,P. Configural Judgments Revealed.
Proceedings of the 75th Annual Convention of The American Psychological Association. 1967, 2, 196 - 196.

Schenck, E. and Naylor, J.C. A Cautionary Note Concerning the Use of Regression Analysis for Capturing the Strategies of People.
Educational and Psychological Measurement. 1968, 28, 3 - 7.

Schum, D.A. Inferences on the basis of Conditionally Nonindependent Data.
Journal of Experimental Psychology. 1966, 72, 401 - 409.

Skinner, B.F. "Superstition" in the pigeon.
Journal of Experimental Psychology. 1948, 38, 168 - 172.

Slovic, P. Cue Consistency and Cue Utilization in Judgement. American Journal of Psychology. 1966, 79, 427 - 434.

Smedslund, J. Multiple - Probability Learning. Oslo: Akademisk Forlag. 1955.

Smedslund, J. The Concept of Correlation in Adults. Scandinavian Journal of Psychology. 1963, 4, 165 - 173.

Summers, S.A. The Learning of Responses to Multiple Weighted Cues. Journal of Experimental Psychology. 1962, 64, 29 - 34.

Swets, J.A., Tanner, W.P. & Birdsall, T.G. Decision Processes in Perception. Psychological Review. 1961, 68, 301 - 340.

Tanner, W.P. and Swets, J.A. A Decision-Making Theory of Visual Detection. Psychological Review. 1954, 61, 401 - 409.

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

Tversky, A. Assessing Uncertainty. Journal of the Royal Statistical Society. 1974, 36, Series B, 148 - 159.

Uhl, C.N. Learning Interval Concepts I.
Journal of Experimental Psychology. 1963, 66, 264 - 273.

Vlek, C.A.J. & Van Der Heijden, L.H.C. Alternative Data Generating Models for Probabilistic Information Processing.
Organizational Behaviour and Human Performance. 1969, 4, 43 - 55.

Vlek, C.A.J. & Van Der Heijden, L.H.C. Aspects of Suboptimality in a Multidimensional Probabilistic Information Processing Task.
Acta Psychologica. 1970, 34, 300 - 310.

Wherry, R.J., Naylor, J.C., Wherry, R.J. & Fallis, R.F. Generating Multiple Samples of Multivariate Data with Arbitrary Population Parameters.
Psychometrika. 1965, 30, 303 - 313.

Wiggins, N. & Hoffman, P.S. Three Models of Clinical Judgment.
Journal of Abnormal Psychology. 1968, 73, 70 - 77.

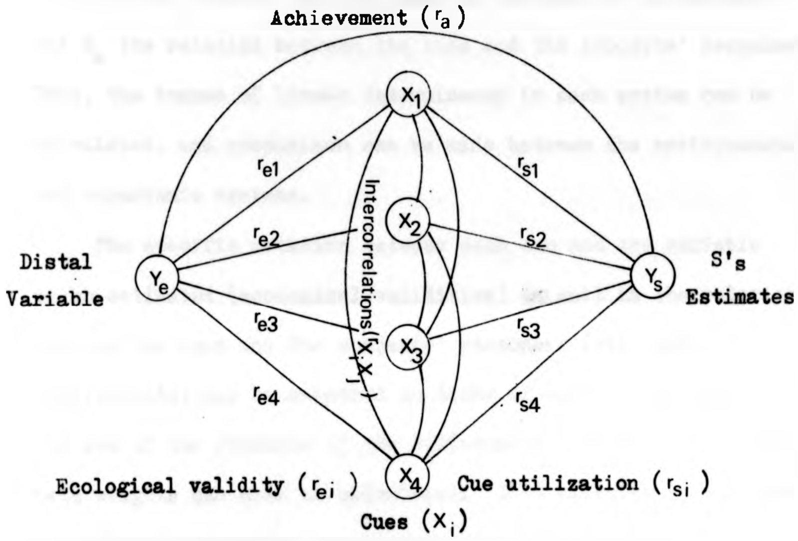
Yntema, D.B. & Torgerson, W.S. Man - Machine Co-operation in Decisions Requiring Common Sense.
IRE Transactions on Human Factors in Electronics, 1961,
HFE - Z, 20 - 26.

APPENDIX A

THE HURSCH, HAMMOND AND HURSCH (1974)

AND HAMMOND, HURSCH AND TODD (1964)

Statistical Formulation of Brunswick's Lens Model.



Brunswick used the analogy of a convex lens to describe situations of the kind in which inferences about an uncertain, probabilistic environment must be made on the basis of probabilistic data. Because both ecological and organismic systems involve a criterion (distal variable and response, respectively) and because several variables (cues) are correlated in various amounts with both distal variable and response, multiple correlation methods may be applied to the analysis of the interactions of the two systems.

When the process of clinical inference is examined in terms of the Lens Model, the application of multiple regression analysis is straight forward. Each half of the Lens can be described in terms of multiple correlation coefficient (R_e for the environment, and R_s for the subject). R_e describes the relationship between the cues and the variable to be estimated and R_s the relation between the cues and the subjects' responses. Thus, the degree of linear determinancy in each system can be calculated, and comparison can be made between the environmental and organismic systems.

The specific relation between each cue and the variable to be estimated (ecological validities) as well as the relation between the cues and the subjects' responses (utilization coefficients) may be measured in terms of correlation coefficients. The sum of the products of the differences between the respective beta weights can also be calculated. This calculation provides an index of the extent to which the subjects use each cue relative to the validity of the cue.

A partial correlation between the subjects' judgements and the variable to be estimated (with the linear variance of each system eliminated) may be calculated.

The relation between overall achievement (correlation between the subjects' estimates and the variable estimated) and the statistical components described above is set forth in the following equation:-

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

where r_a = correlation between subjects' judgements and the variable estimated.

R_e = the multiple correlation between the cues and the variable estimated.

R_s = the multiple correlations between the cues and the subject's judgements.

Σd = the sum of the products $(r_{ei} - r_{si})(\beta_{ei} - \beta_{si})$ where r_{ei} = the correlation between cue i and the variable estimated, r_{si} = the correlation between cue i and the subject's judgements, β_{ei} = the beta weight for the correlation between cue i and the variable estimated and β_{si} = the beta weight between cue i and the subject's judgements.

C = the correlation between the variance unaccounted for by the multiple correlation in the ecology and the variance unaccounted for by the multiple correlation in the subject's response system.

Tucker (1964) showed that this relationship could be equally well expressed as

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

where G is the correlation between the predictions of the criterion using the weights β_{ei} and the predictions of the subject's judgements using the weights β_{si}.

G, then, may be regarded as an index of the extent to which the subject uses the cues linearly and in an appropriate manner. Similarly, C may be regarded as an index of the extent to which the subject uses the cues in an appropriate non-linear way.

The above developments of the Lens Model were made on the assumption that the model in which we are interested is the simple linear one.

i.e. $Y = \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots + \beta_N X_N + \epsilon$

G therefore represents the correlation between the linear models which best fit the environmental criterion and the subject's responses i.e. between

$$Y_e = \beta_{e1} X_1 + \beta_{e2} X_2 + \ldots\ldots + \beta_{eN} X_N$$

and

$$Y_s = \beta_{s1} X_1 + \beta_{s2} X_2 + \ldots\ldots + \beta_{sN} X_N$$

C represents the correlation between the remaining, non-linearly predictable, variance in the environmental and judgemental system.

Theoretically, there is no reason why this method of analysis should not be applied to psychologically non-linear models (though these must be statistically linear).

e.g. $Y = \beta_1 X_1 + \beta_2 X_2^{\,2} + \ldots\ldots + \beta_N X_N^{\,N} + \epsilon$

G then represents the correlation between those models of the above form which best predict the environmental criterion and the subject's responses, i.e. between

$$\hat{Y}_e = \beta_{e1} X_1 + \beta_{e2} X_2^{\,2} + \ldots\ldots + \beta_{eN} X_N^{\,N}$$

and

$$\hat{Y}_s = \beta_{s1} X_1 + \beta_{s2} X_2^{\,2} + \ldots\ldots + \beta_{sN} X_N^{\,N}$$

C is now the correlation between the remaining variance, not predictable from the model, in the environmental and judgemental systems.

The terms GR_eR_s and $C\sqrt{(1 - R_e^2)(1 - R_s^2)}$ which were formerly regarded as indices of the degree to which the subjects' accuracy was due to appropriate use of the cues in a linear and non-linear manner respectively, may now be regarded as indices of the degree to which the subject's accuracy is due to appropriate use of the cues in a manner prescribed by the model and appropriate use of cues in manners not so prescribed.

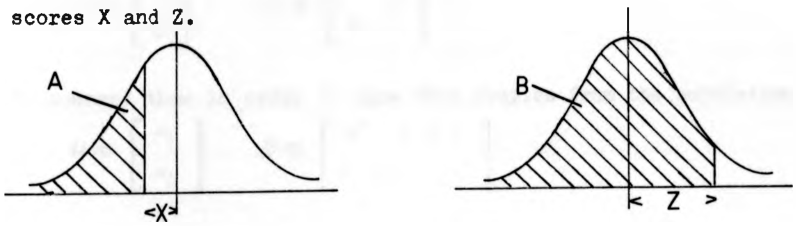
APPENDIX B

THE GENERATION OF SAMPLES FROM BIVARIATE NORMAL POPULATIONS WITH KNOWN PARAMETERS.

The following method was used to generate random samples from a bivariate normal populate defined by the parameters.

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{bmatrix}$$

Two random numbers A and B (between zero and 1) were obtained from a random number generating procedure. Each was fed into an inverse normal procedure giving two random z scores. i.e. A and B were treated as the area under the normal curve below same z score, the inverse normal procedure computed the appropriate z scores X and Z.



X and Z are therefore random samples from the standardized normal distribution (N(0,1)).

A new variable Y was formed by the following operation:-

$$Y = \rho X + \sqrt{1 - \rho^2}\, Z$$

Since the mean of the sum of two independent normal variables is equal to the sum of their means,

and since ρX is distributed as $N(0, \rho^2)$

and $\sqrt{1-\rho^2}\, Z$ is distributed as $N(0, 1 - \rho^2)$

then the variable Y has mean zero.

Since the variance of the sum of two independent normal variables, is equal to the sum of their variances Y is distributed variance $\rho^2 + 1 - \rho^2 = 1$.

Since the sum of two normal variables is itself a normally distributed variate Y is distributed as N (0,1).

The correlation between the variables X and Y is equal to the square root of the proportion of the variance in Y which can be explained by X i.e. $\sqrt{\rho^2} = \rho$

X and Y may be transformed so as to correspond to any bivariate normal population. At present, they form a random observation from the population

$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

To convert them in order to make them samples from the population

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_1 \sigma_2 \rho \\ \sigma_1 \sigma_2 \rho & \sigma_2^2 \end{bmatrix}$$

where Y is from variable 1 and X from variable 2, the following transformations may be applied.

$$Y' = \sigma_1 Y + \mu_1$$
$$X' = \sigma_2 X + \mu_2$$

For n - variate populations where n is greater than two, this method is not easily generalized and it is suggested that a method such as that of Wherry, Naylor, Wherry and Fallis (1965) be used.

APPENDIX C

STATISTICAL APPENDIX

The "normal distribution" is the name applied to the familiar bell-shaped curve which so frequently results when a large number of events play a part in determining the score of an item on some continuous variable. It is a function of such distributions that they can be uniquely described by only two parameters, the mean μ and variance σ^2. The probability density function is given by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

It is possible to generalize this concept of normal distribution to more than one variable. The resulting multivariate normal distribution being characterized in terms of a vector of mean values μ and a covariance matrix Σ. The former of these is simply a vector of the mean values of the population on each of the variables. Σ is analagous to the variance σ^2 in the univariate case, its entries in the diagonals being the variances of the population on each variable, the off-diagonal entries being the covariances between each pair of variables.

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}$$

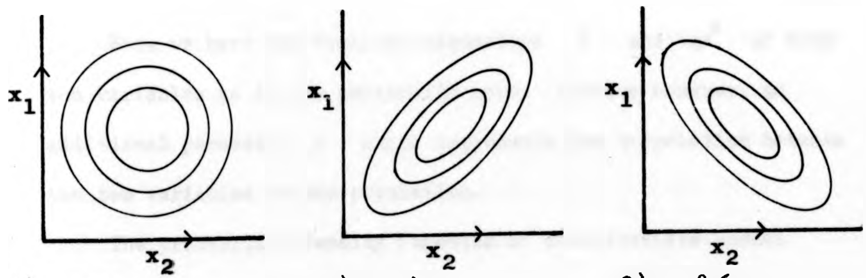
$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{12} & \cdots & \sigma_1\sigma_n\rho_{1n} \\ \sigma_1\sigma_2\rho_{12} & \sigma_2^2 & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_1\sigma_n\rho_{1n} & \cdots & \cdots & \sigma_n^2 \end{bmatrix}$$

Letting $\sigma_1 = \sigma_2$ for simplicity, if points of equal density (i.e. points (x_1, x_2) such that $F(x_1, x_2)$ equals a constant) are plotted, it can be seen that:-

(1) If $\rho = 0$ these points form circles concentric about the mean.

(2) If $\rho > 0$ the points form ellipses, whose major axis has positive slope.

(3) If $\rho < 0$ the points form ellipses, whose major axis has negative slope.



1) $\rho = 0$ 2) $\rho > 0$ 3) $\rho < 0$

The probability density function of the multivariate normal distribution is

$$f(x) = \frac{|\Sigma|^{-1/2}}{(2\pi)^{n/2}} e^{-1/2 (x-\mu)' \Sigma^{-1} (x-\mu)}$$

Where X is a vector of an individual's scores on the n variables.

We shall be particularly concerned with the bivariate normal distribution, characterized by

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} \sigma_1' & \sigma_1 \sigma_2 \rho \\ \sigma_1 \sigma_2 \rho & \sigma_2 \end{bmatrix}$$

Here we have the familiar parameters μ and σ^2 of both the variables as in the univariate case. We have however, an additional parameter ρ which represents the correlation between the two variables in the population.

The probability density function of the bivariate normal distribution is

$$F(x) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \quad e^{-\frac{1}{2}\left[\frac{1}{1-\rho^2}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)\right]}$$

where $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

DISCRIMINANT ANALYSIS.

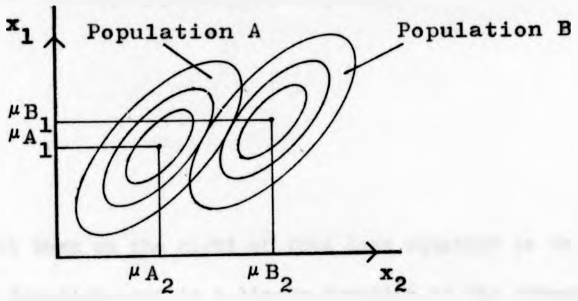
Situations frequently arise in which it is necessary to classify items, people etc., as belonging to one of two or more multivariate normal populations on the basis of measurements carried out on those items or people. R.A. Fisher (1936) developed a method which allowed optimal classification of objects into multivariate normal populations. The technique was later named Discriminant Function Analysis.

For ease of explanation, the two population case with equal probabilities will be illustrated. Generalisation to more than two populations or unequal prior probabilities is not difficult.

Assume that we have measures on n variables for an individual, $X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$, and we wish to classify this individual as belonging to one of two n - variate normal populations whose parameters μ and Σ are known or can be estimated. The covariance matrices of the two populations are assumed to be equal, so the populations may be viewed as two n - dimensional ellipsoids of the same shape and orientation, occupying different positions in an n - dimensional space.

The bivariate situation is illustrated overleaf.

(It should be noted that throughout this appendix the two populations are labelled A and B rather than 1 and 2 as in the experiments. This is to facilitate indexing.)



Clearly at a point $X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ in the space, both populations have a certain density - $F_A(X)$ and $F_B(X)$. The optimum strategy is, clearly, to assign an individual at X to that population whose density is greater at that point.

i.e. If $F_A(X) > F_B(X)$ assign to population A.

If $F_A(X) < F_B(X)$ assign to population B.

This may be phrased in terms of the likelihood ratio.

$$L(X) = \frac{F_A(X)}{F_B(X)}$$

If $L(X) > 1$ assign to population A.

If $L(X) < 1$ assign to population B.

Clearly if $L(X) = 1$ both populations are equally likely.

$$L(X) = \frac{F_A(X)}{F_B(X)} = \frac{\left[(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}\right]^{-1} e^{-\frac{1}{2}\left[(X-\mu_A)' \Sigma^{-1} (X-\mu_A)\right]}}{\left[(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}\right]^{-1} e^{-\frac{1}{2}\left[(X-\mu_B)' \Sigma^{-1} (X-\mu_B)\right]}}$$

taking logs

$$\mathrm{Ln}\, L(X) = -\tfrac{1}{2}\left[(X-\mu_A)'\Sigma^{-1}(X-\mu_A) - (X-\mu_B)'\Sigma^{-1}(X-\mu_B)\right]$$

$$\mathrm{Ln}\, L(X) = X'\Sigma^{-1}(\mu_A \ \mu_B) - \tfrac{1}{2}(\mu_A+\mu_B)'\Sigma^{-1}(\mu_A+\mu_B)$$

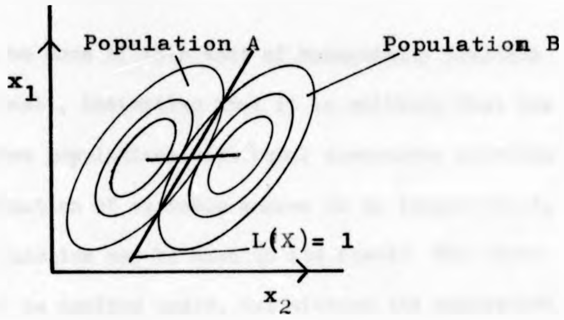
The first term on the right of this last equation is the discriminant function, and is a linear function of the components of the observation vector X. The function may be evaluated for any X and if the resulting values are greater than zero (i.e. $\mathrm{Ln}\, L(X) > 0$ and therefore $L(X) > 1$ then population A is the more likely population at that point.

In the bivariate situation the above expression takes the form:-

$$\frac{1}{1-\rho^2}\left(\frac{\mu_{A1}}{\sigma_{A1}^2} - \frac{\rho\mu_{A2}}{\sigma_{A1}\sigma_{A2}} - \frac{\mu_{B1}}{\sigma_{B1}^2} + \frac{\rho\mu_{B2}}{\sigma_{B1}\sigma_{B2}}\right)x_1$$
$$+\frac{1}{1-\rho^2}\left(\frac{\mu_{A2}}{\sigma_{A2}} - \frac{\rho\mu_{A1}}{\sigma_{A1}\sigma_{A2}} - \frac{\mu_{B2}}{\sigma_{B2}^2} + \frac{\rho\mu_{B1}}{\sigma_{B1}\sigma_{B2}}\right)x_2$$
$$-\frac{1}{2(1-\rho^2)}\left(\frac{\mu^2_{A1}}{\sigma^2_{A1}} - \frac{2\mu_{A1}\mu_{A2}\rho}{\sigma_{A1}\sigma_{A2}} + \frac{\mu_{A2}}{\sigma_{A2}^2} - \frac{\mu^2_{B1}}{\sigma_{B2}^2} + \frac{2\mu_{B1}\mu_{B2}\rho}{\sigma_{B1}\sigma_{B2}} - \frac{\mu^2_{B2}}{\sigma^2_{B2}}\right)$$
$$= \mathrm{Ln}\, L(X)$$

This is clearly a linear combination of the scores of the individual on the two variables.

For a given value of L (X) (or Ln L(X)) it can be shown that those points in the space with this value of L(X) lie on a straight line which cut the line joining the two population terms.



The line for L (X) = 1, i.e. the line of points which are equally likely to be from populations A or B, in fact bisects the line joining the population means.

Thus, any problem involving the classification of items or people as belonging to one of two or more populations differing only in mean vectors may be reduced to the comparison of a weighted combination of scores on the n - variables with some constant, usually LnL(X) = 0 but sometimes other values (for instance when payoffs or prior probabilities are unequal).

UNEQUAL COVARIANCE MATRICES.

We might question the likelihood of the assumption of equality of the covariance matrices. In some instances there is little reason to suppose that the relationship between the n - variables will be the same in both populations. The claims of physicians, for instance, that the pattern of scores is often important in making differential diagnoses, would lead us to suspect that the relationship between scores in one of the disease populations differs from that in another. The assumption of equality may be tested by taking samples from the populations and carrying out a test of homogeneity (c.f. Anderson, 1958).

What can be done if this test of homogeneity yields a significant result, indicating that it is unlikely that the samples are from populations with equal covariance matrices? A linear combination of variable scores is no longer valid, what sort of function may be used in its stead? The above derivation may be applied again, but without the assumption that $\Sigma_A = \Sigma_B$.

(c.f. Eisenbeis and Avery, 1972):-

$$\mathrm{Ln}L(X) = \frac{F_A(X)}{F_B(X)} = \mathrm{Ln}\left\{\frac{\left[(2\pi)^{\frac{n}{2}}|\Sigma_A|^{\frac{1}{2}}\right]^{-1} e^{-\frac{1}{2}\left[(X-\mu_A)'\Sigma_A^{-1}(X-\mu_A)\right]}}{\left[(2\pi)^{\frac{n}{2}}|\Sigma_B|^{\frac{1}{2}}\right]^{-1} e^{-\frac{1}{2}\left[(X-\mu_B)'\Sigma_B^{-1}(X-\mu_B)\right]}}\right\}$$

$$= \frac{1}{2}\mathrm{Ln}|\Sigma_B.\Sigma_A^{-1}| - \frac{1}{2}\left[(X-\mu_A)'\Sigma_A^{-1}(X-\mu_A) - (X-\mu_B)'\Sigma_B^{-1}(X-\mu_B)\right]$$

If the value of this expression is greater than the critical value of LnL(X) (zero in the unbiased case) then an individual at point X is assigned to population A.

This rule may be expressed in terms of a quadratic function:-

Assign to population A if

$$X'(\Sigma_A^{-1} - \Sigma_B^{-1})X - 2(\mu_A'\Sigma_A^{-1} - \mu_B'\Sigma_B^{-1})X + \mu_A'\Sigma_A^{-1}\mu_A - \mu_B'\Sigma_B^{-1}\mu_B$$

$$< \mathrm{Ln}|\Sigma_B.\Sigma_A^{-1}| - 2\,\mathrm{Ln}L(X)$$

or when LnL(X) = 0

$$< \mathrm{Ln}|\Sigma_B.\Sigma_A^{-1}|$$

Clearly this is a far more complex function than was obtained when Σ_A and Σ_B were assumed equal. In the bivariate case with

$$\Sigma_A = \begin{bmatrix} \sigma_{A1}^2 & \sigma_{A1}\sigma_{A2}\rho_A \\ \sigma_{A1}\sigma_{A2}\rho_A & \sigma_{A2}^2 \end{bmatrix} \quad \text{and} \quad \Sigma_B = \begin{bmatrix} \sigma_{B1}^2 & \sigma_{B1}\sigma_{B2}\rho_B \\ \sigma_{B1}\sigma_{B2}\rho_B & \sigma_{B2}^2 \end{bmatrix}$$

the function becomes:-

$$\frac{1}{2}\left(\frac{1}{(1-\rho_B^2)\,\sigma_{B1}^2} - \frac{1}{(1-\rho_A^2)\,\sigma_{A1}^2}\right)x_1^2 + \frac{1}{2}\left(\frac{1}{(1-\rho_B^2)\,\sigma_{B2}^2} - \frac{1}{(1-\rho_A^2)\,\sigma_{A2}^2}\right)x_2^2$$

$$+\left(\frac{\mu_{A1}}{\sigma_{A1}^2(1-\rho_A^2)} - \frac{\rho_A\,\mu_{A2}}{\sigma_{A1}\sigma_{A2}(1-\rho_A^2)} - \frac{\mu_{B1}}{\sigma_{B1}^2(1-\rho_B^2)} + \frac{\rho_B\,\mu_{B2}}{\sigma_{B1}\sigma_{B2}(1-\rho_B^2)}\right)x_1$$

$$+\left(\frac{\mu_{A2}}{\sigma_{A2}^2(1-\rho_A^2)} - \frac{\rho_A\,\mu_{A1}}{\sigma_{A1}\sigma_{A2}(1-\rho_A^2)} - \frac{\mu_{B2}}{\sigma_{B2}^2(1-\rho_B^2)} + \frac{\rho_B\,\mu_{B1}}{\sigma_{B1}\sigma_{B2}(1-\rho_B^2)}\right)x_2$$

$$+\left(\frac{\rho_A}{\sigma_{A1}\sigma_{A2}(1-\rho_A^2)} - \frac{\rho_B}{\sigma_{B1}\sigma_{B2}(1-\rho_B^2)}\right)x_1 x_2$$

$$-\frac{1}{2}\left[\frac{\mu_{A1}^2}{\sigma_{A1}^2(1-\rho_A^2)} - \frac{2\mu_{A1}\mu_{A2}\rho_A}{\sigma_{A1}\sigma_{A2}(1-\rho_A^2)} + \frac{\mu_{A2}^2}{\sigma_{A2}^2(1-\rho_A^2)} - \frac{\mu_{B1}^2}{\sigma_{B1}^2(1-\rho_B^2)} + \frac{2\mu_{B1}\mu_{B2}\rho_B}{\sigma_{B1}\sigma_{B2}(1-\rho_B^2)} - \frac{\mu_{B2}^2}{\sigma_{B2}^2(1-\rho_B^2)}\right]$$

$$< \mathrm{Ln}\left(\sigma_{A1}\,\sigma_{A2}\sqrt{1-\rho_A^2}\right) - \mathrm{Ln}\left(\sigma_{B1}\,\sigma_{B2}\sqrt{1-\rho_B^2}\right) + \mathrm{Ln}\,L(x)$$

This function is of the general kind known as quadratic forms, including as it does, not only linear but also quadratic and cross product terms of the variables. (The function remains of the quadratic form type regardless of the number of variables concerned). Unlike the function derived earlier under the assumption of equal covariance matrices, for which plots of the points with likelihood ratio equal to some constant yielded a straight line, plots of points of equal likelihood ratio for this quadratic form take a number of possible shapes. Just what shape these plots will take depends upon the parameter value associated with each term (c.f. Noble, 1969) which in turn depend on the parameters of the populations in question, but theoretically they may be hyperbolas, ellipses or parallel straight lines. In all tasks investigated in the present research, in which the two populations were of unequal dispersion, the resulting curves of constant likelihood ratio were hyperbolas.

One property of the equal covariance matrix situation is that the decisions made by the linear combinations are unbiased i.e. if the cut off is at $L(X) = 1$ then the same proportion of population A is displaced as is displaced from population B. When unequal covariance matrices are considered, the quadratic function with cut off at $L(X) = 1$ introduces systematic biases in favour of one or other of the populations. The degree of misclassification may be evaluated by the integrals.

$$\iint F_A(X)\ dx_1\ dx_2 \qquad \text{over the area classified as B}$$

and

$$\iint F_B(X)\ dx_1\ dx_2 \qquad \text{over the area classified as A.}$$

Obviously the degree of correct classification of each population is one minus each of these integrals.

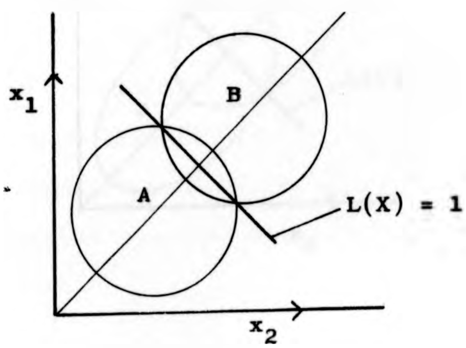
EXPERIMENT I

In this experiment the covariance matrices of the two populations between which discriminations were made were equal in each condition.

Condition I.

$$\mu_A = \begin{bmatrix} 7.5 \\ 7.5 \end{bmatrix} \qquad \Sigma_A = \begin{bmatrix} 4.0 & 0.0 \\ 0.0 & 4.0 \end{bmatrix} \quad \text{i.e. } \rho_A = 0.0$$

$$\mu_B = \begin{bmatrix} 9.5 \\ 9.5 \end{bmatrix} \qquad \Sigma_B = \begin{bmatrix} 4.0 & 0.0 \\ 0.0 & 4.0 \end{bmatrix} \quad \text{i.e. } \rho_B = 0.0$$

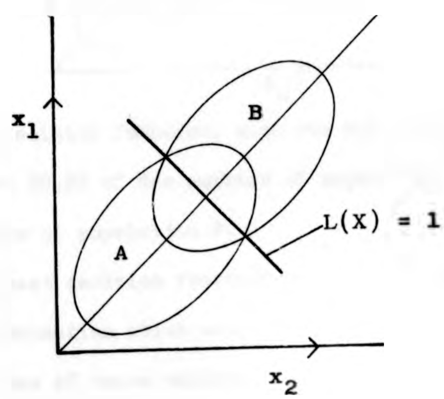


The best decision function, with cut off at L(X) = 1, correctly classified 76.0% of the members of population A and 76.0% of the members of population B.

Condition II.

$$\mu_A = \begin{bmatrix} 7.5 \\ 7.5 \end{bmatrix} \qquad \Sigma_A = \begin{bmatrix} 4.0 & 2.8 \\ 2.8 & 4.0 \end{bmatrix} \quad \text{i.e. } \rho_A = 0.7$$

$$\mu_B = \begin{bmatrix} 9.5 \\ 9.5 \end{bmatrix} \qquad \Sigma_B = \begin{bmatrix} 4.0 & 2.8 \\ 2.8 & 4.0 \end{bmatrix} \quad \text{i.e. } \rho_B = 0.7$$

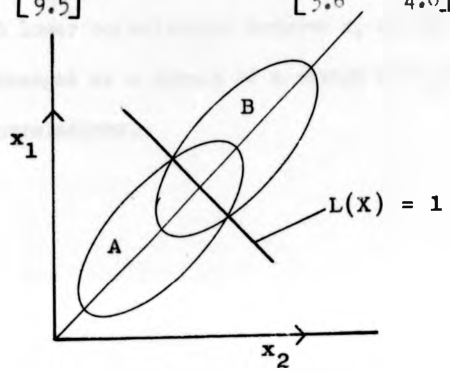


The best decision function, with cut off at L(X) = 1, correctly classifies 70.6% of the members of population A and 70.6% of the members of population B.

Condition III.

$$\mu_A = \begin{bmatrix} 7.5 \\ 7.5 \end{bmatrix} \qquad \Sigma_A = \begin{bmatrix} 4.0 & 3.6 \\ 3.6 & 4.0 \end{bmatrix} \quad \text{i.e. } \rho_A = 0.9$$

$$\mu_B = \begin{bmatrix} 9.5 \\ 9.5 \end{bmatrix} \qquad \Sigma_B = \begin{bmatrix} 4.0 & 3.6 \\ 3.6 & 4.0 \end{bmatrix} \quad \text{i.e. } \rho_B = 0.9$$



The best decision function, with cut off at L(X) = 1 correctly classified 69.6% of the members of population A, and 69.6% of the members of population B.

The best decision function in all three situations is a linear combination which weights both x_1 and x_2 equally the exact values of these weights vary as a function of the correlations.

Condition I.

$8.50 - .50x_1 - .50x_2 = \mathrm{LnL}(X)$

the line of equal probability is $x_1 = 17.0 - x_2$.

Condition II.

$5.00 - .29x_1 - .29x_2 = \mathrm{LnL}(X)$

the line of equal probability is $x_1 = 17.0 - x_2$.

Condition III.

$$4.47 - .26x_1 - .26x_2 = \text{LnL}(X)$$

the line of equal probability is $x_1 = 17.0 - x_2$.

The difference in the weightings given to x_1 and x_2 implies that with lower correlations between x_1 and x_2 LnL(X) is more greatly changed as a result of a change in x_1 or x_2 than at higher correlations.

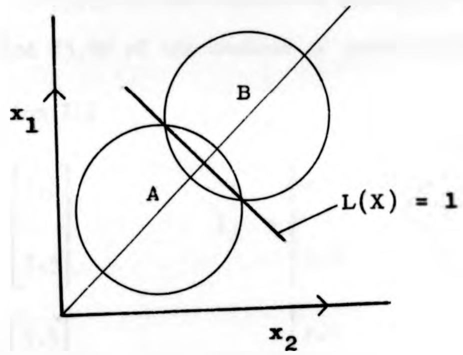
EXPERIMENT II

In this experiment the difference in correlation between x_1 and x_2 which existed between population A and population B was varied over the conditions.

Condition I.

$$\mu_A = \begin{bmatrix} 7.5 \\ 7.5 \end{bmatrix} \quad \Sigma_A = \begin{bmatrix} 4.0 & 0.0 \\ 0.0 & 4.0 \end{bmatrix} \quad \text{i.e. } \rho_A = 0.0$$

$$\mu_B = \begin{bmatrix} 9.5 \\ 9.5 \end{bmatrix} \quad \Sigma_B = \begin{bmatrix} 4.0 & 0.0 \\ 0.0 & 4.0 \end{bmatrix} \quad \text{i.e. } \rho_B = 0.0$$



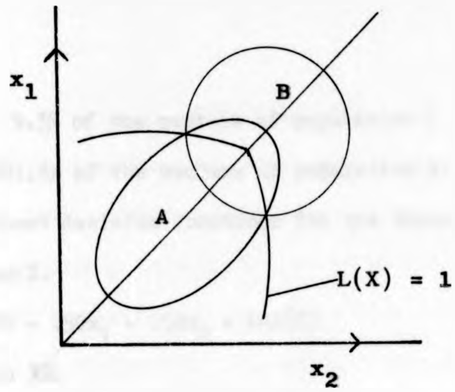
The best decision function (which is linear), with cut off at L(X) = 1, (L L(X) = 0), correctly classifies 76.0% of the members of population A, and 76.0% of the members of population B.

Condition II.

$$\mu_A = \begin{bmatrix} 7.5 \\ 7.5 \end{bmatrix} \quad \Sigma_A = \begin{bmatrix} 4.0 & 2.8 \\ 2.8 & 4.0 \end{bmatrix} \quad \text{i.e. } \rho_A = 0.7$$

$$\mu_B = \begin{bmatrix} 9.5 \\ 9.5 \end{bmatrix} \quad \Sigma_B = \begin{bmatrix} 4.0 & 0.0 \\ 0.0 & 4.0 \end{bmatrix} \quad \text{i.e. } \rho_B = 0.0$$

p.t.o.



The best decision function (which is non-linear) with cut off at L(X) = 1 correctly classifies

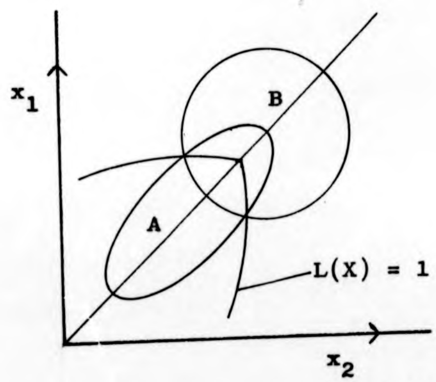
79.6% of the members of population A

and 73.0% of the members of population B.

Condition III

$$\mu_A = \begin{bmatrix} 7.5 \\ 7.5 \end{bmatrix} \qquad \Sigma_A = \begin{bmatrix} 4.0 & 3.6 \\ 3.6 & 4.0 \end{bmatrix} \quad \text{i.e. } \rho_A = 0.9$$

$$\mu_B = \begin{bmatrix} 9.5 \\ 9.5 \end{bmatrix} \qquad \Sigma_B = \begin{bmatrix} 4.0 & 0.0 \\ 0.0 & 4.0 \end{bmatrix} \quad \text{i.e. } \rho_B = 0.0$$



The best decision function (which is non-linear) with cut off at L(X) = 1, correctly classifies

p.t.o.

79.3% of the members of population A

and 81.6% of the members of population B.

The best decision functions for the three conditions are:-

Condition I.

$$8.50 - .50x_1 - .50x_2 = \mathrm{LnL}(X)$$

Condition II.

$$14.63 - .12x_1^{\,2} - .12x_2^{\,2} + .34x_1x_2 - 1.27x_1 - 1.27x_2 = \mathrm{LnL}(X)$$

Condition III.

$$15.99 - .53x_1^{\,2} - .53x_2^{\,2} + 1.18x_1x_2 - 1.39x_1 - 1.39x_2 = \mathrm{LnL}(X)$$

It can be seen that the square and cross product terms receive weights of increasing magnitude as the difference in cue correlation between the two populations increases. The functions of the curves of constant likelihood ratio may be found by rearranging the above equations, but since these functions are quadratic, the relationship between x_1 and x_2 can only be expressed using the algorithm for solution of a quadratic equation. For simplicity's sake, they are not included.

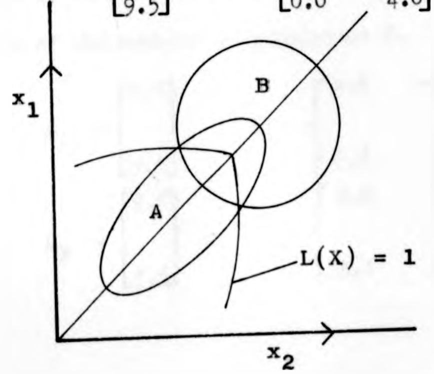
EXPERIMENT III

In this experiment all conditions have what are, except at a trivial level, the same statistical properties. Condition II and III are simply transformations of Condition I either by addition of a constant to one of the cues or by the multiplication of a cue by a constant and subsequent addition of a constant. These transformations do not affect the difficulty of the task in terms of the proportions of the members of populations A and B who are correctly placed by the best decision function (which is in all cases non-linear).

Condition I.

$$\mu_A = \begin{bmatrix} 7.5 \\ 7.5 \end{bmatrix} \quad \Sigma_A = \begin{bmatrix} 4.0 & 3.6 \\ 3.6 & 4.0 \end{bmatrix} \quad \text{i.e. } \rho_A = 0.9$$

$$\mu_B = \begin{bmatrix} 9.5 \\ 9.5 \end{bmatrix} \quad \Sigma_B = \begin{bmatrix} 4.0 & 0.0 \\ 0.0 & 4.0 \end{bmatrix} \quad \text{i.e. } \rho_B = 0.0$$



The best decision function, (which is non linear) with cut off at L(X) = 1 correctly classifies

79.3% of population A,

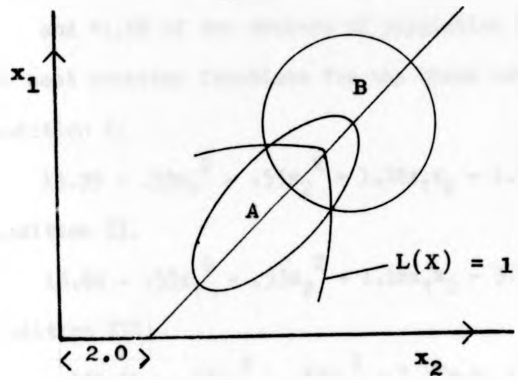
and 81.6% of population B.

p.t.o.

Condition II

$$\mu_A = \begin{bmatrix} 7.5 \\ 9.5 \end{bmatrix} \qquad \Sigma_A = \begin{bmatrix} 4.0 & 3.6 \\ 3.6 & 4.0 \end{bmatrix} \quad \text{i.e. } \rho_A = 0.9$$

$$\mu_B = \begin{bmatrix} 9.5 \\ 11.5 \end{bmatrix} \qquad \Sigma_B = \begin{bmatrix} 4.0 & 0.0 \\ 0.0 & 4.0 \end{bmatrix} \quad \text{i.e. } \rho_B = 0.0$$



The best decision function , (which is non-linear) with cut off at L(X) = 1 correctly classifies

79.3% of the members of population A,

and 81.6% of the members of population B.

Condition III

$$\mu_A = \begin{bmatrix} 7.5 \\ 9.5 \end{bmatrix} \qquad \Sigma_A = \begin{bmatrix} 4.0 & -3.6 \\ -3.6 & 4.0 \end{bmatrix} \quad \text{i.e. } \rho_A = -0.9$$

$$\mu_B = \begin{bmatrix} 9.5 \\ 7.5 \end{bmatrix} \qquad \Sigma_B = \begin{bmatrix} 4.0 & 0.0 \\ 0.0 & 4.0 \end{bmatrix} \quad \text{i.e. } \rho_B = 0.0$$