54

THE UNIVERSITY OF STIRLING

INDIVIDUAL DECISION MAKING IN STATIC, SEQUENTIAL AND DYNAMIC SITUATIONS

a thesis submitted for the Degree of

Doctor of Philosophy

in the University of Stirling

by

Robert Henry Ranyard, B.Sc.

December 1972.

Acknowledgements.

I would like to thank my supervisor, Mr.M.F.Moore for his guidance during the preparation of this thesis, Angus Annan who built the apparatus for the sequential experiments and many colleagues at Stirling who discussed the problems and thereby helped resolve them. I am also grateful to Mrs.S.Redman, who typed the thesis, for her careful and efficient work. The diagrams and bibliography were prepared by my wife, Margaret who naturally has my special thanks, for her considerable help.

Abstract.

Static, sequential and dynamic models of decision making situations and additive and subjectively expected utility models of decision making behaviour are defined and discussed in chapter 1. Results from conjoint measurement theory are surveyed in chapter 2 and their application to a qualitative functional analysis (QFA) of the information integration models is discussed. An important issue is how to deal with fallible data. In chapter 3 functional measurement for binary choice data by the method of minimum normit chi square is considered with a view to examining information integration models quantitatively. In the last of the 4 theoretical chapters a selective review of experimental work related to some major issues in decision theory is presented.

Six pair comparison experiments are reported in chapters 5 to 7, in five of them choices were observed and in the other (experiment 2) statements of preference were elicited. The subjects were randomly selected university students whose results were analysed individually. The alternatives available to subjects were gambles for small amounts of money, which they actually played in real play situations. Experiments 1 - 3 were set in static decision-making situations, 4 and 5 in sequential ones and experiment 6 was set in a dynamic situation. In experiment 2 the role of indifference in decision making was investigated by QFA and found to be minor. In experiments 1, 3 and 5 information integration models were investigated by QFA and functional measurement. Support for SEU models and not additive ones was found. In experiments 4 and 5 the effects on choices of current capital and previous outcome were found to be negligible.

Experiment 5 was a study of an additive information integration model in a simple, two stage, dynamic betting game. A functional measurement analysis led to its rejection.

In the final chapter the results are discussed in relation to previous work. Methodological difficulties which arose from the use of the analytic methods are considered and partly resolved. It is concluded that they are suitable techniques for the present application. The future of the information integration models is also discussed. It is concluded that as descriptive models of behaviour in static, sequential and dynamic situations they still have a very useful role to play.

CONTENTS.

CHAPTER 1.

Introduction

Much human activity involves the conscious pursuit of desirable objectives. An individual engaged in such activity may perceive that certain alternative means could lead to the desired end. Prior to choosing an action he will probably wish to obtain as much information as possible about each alternative before him. Information seeking in such situations has rightly received much study. Our interest here, though, is in what happens after the information has been gathered.

The individual will invariably be faced with a fundamental dilemma: his possible courses of action will not always lead to the outcome intended and, furthermore, any outcome will have both desirable and undesirable aspects. Man does not have the power to make perfect predictions about his future environment and even if he did have it would not be to his liking in all respects.

The behaviour of fully informed individuals, making decisions among discrete courses of action in order to obtain desired goals, is the subject of this thesis.

It is a very broad subject and it is not intended to attempt a panoramic study of its whole breadth. A specific class of models of decision making behaviour are proposed and examined empirically. This class of models will be called information integration models though this name is used by others to describe a more general class. The usage of the term here is very similar to Anderson and Shanteau's (1970) usage. Information integration models are those which assume that people evaluate information about all the attributes of an available alternative and integrate these evaluations to form some overall assessment of its subjective value. Decisions are then made by comparing the subjective values of all the available alternatives. The models will be discussed again later in the chapter.

Considerable attention has been given to information integration models over the past 20 years in a variety of evaluative situations. Recently, techniques have become available which are particularly useful for examining them in decision making contexts. Two quite different, though similarly powerful techniques will be applied in the present study. An important point to note is that the extent to which any generalization from the application of these analytic methods is valid depends critically on the degree to which the situations that behaviour is examined under resemble real life situations. Any psychological study necessarily involves the study of interrelations between the individual and his environment. The first task, then is to discuss real life decision making environments and build models of them. Models, by definition do not mirror reality but simplify it and draw out salient features of it. The models of decision making situations to be proposed should be viewed in this light. They are caricatures, abstractions of real life situations which hopefully retain some of their essence. The simplest model of decision situations discussed is one in which the temporal features are minimum. Generalizations of this which bring in some of the more important temporal features are also considered. The main aim of the study as a whole is to examine information integration models in such situations.

The structure of the study to be reported is as follows.

Most of the remainder of this chapter is devoted to the models of decision making situations which give the framework of the study. Then information integration models are described further. In chapters 2 and 3 the techniques which are used to examine the models of behaviour are set out and in chapter 4 previous empirical research relevant to the present study is reviewed. The experiments carried out in the present study are described in chapters 5 through to 7. Conclusions are drawn in the final chapter (chapter 8). Substantive conclusions are discussed in relation to previous research and the analytical methods are discussed in the light of experience with their application.

Some Simple Models of Decision Making Situations.

Three basic models have mainly been used to study decision making. They are often called dynamic, sequential and static in order of decreasing complexity, though there is some confusion in this terminology. Examples of each type will be given and the way the above terms are to be used clarified. The first two are generalizations of the static model.

A course of action in pursuit of a goal usually requires making a sequence of decisions. After the initial one, events beyond the control of the decision maker take over to some extent. As a result of the initial action and the subsequent events a new state of affairs is reached which requires the selection of a new action. The ultimate goal is not reached as the direct result of a single action. Rather it is attained through reaching towards a series of sub-goals. Fishburn (1964) gives an interesting example of this in considering a career choice problem. The following is a more simple career choice problem.

A person has just left school and has been accepted by a university where he must choose which subject to study. Once a course of study is begun the system is so inflexible that the possibility of changing courses can be ruled out. At school his favourite subject was chemistry, and his best subject english literature. He wishes to continue to study one of these. His ultimate goal can be considered to be a satisfying and financially rewarding career.

If he decides to take english literature, he knows he will find the course boring, though reasonably easy, and he will have a lot of time to enjoy the non-academic aspects of university life. If he takes chemistry he will have to work quite hard but find it interesting. His free time will be more limited. He stands the best chance of getting a good degree if he takes english literature. If he takes chemistry he will possibly get a good degree but more probably it will only be fair.

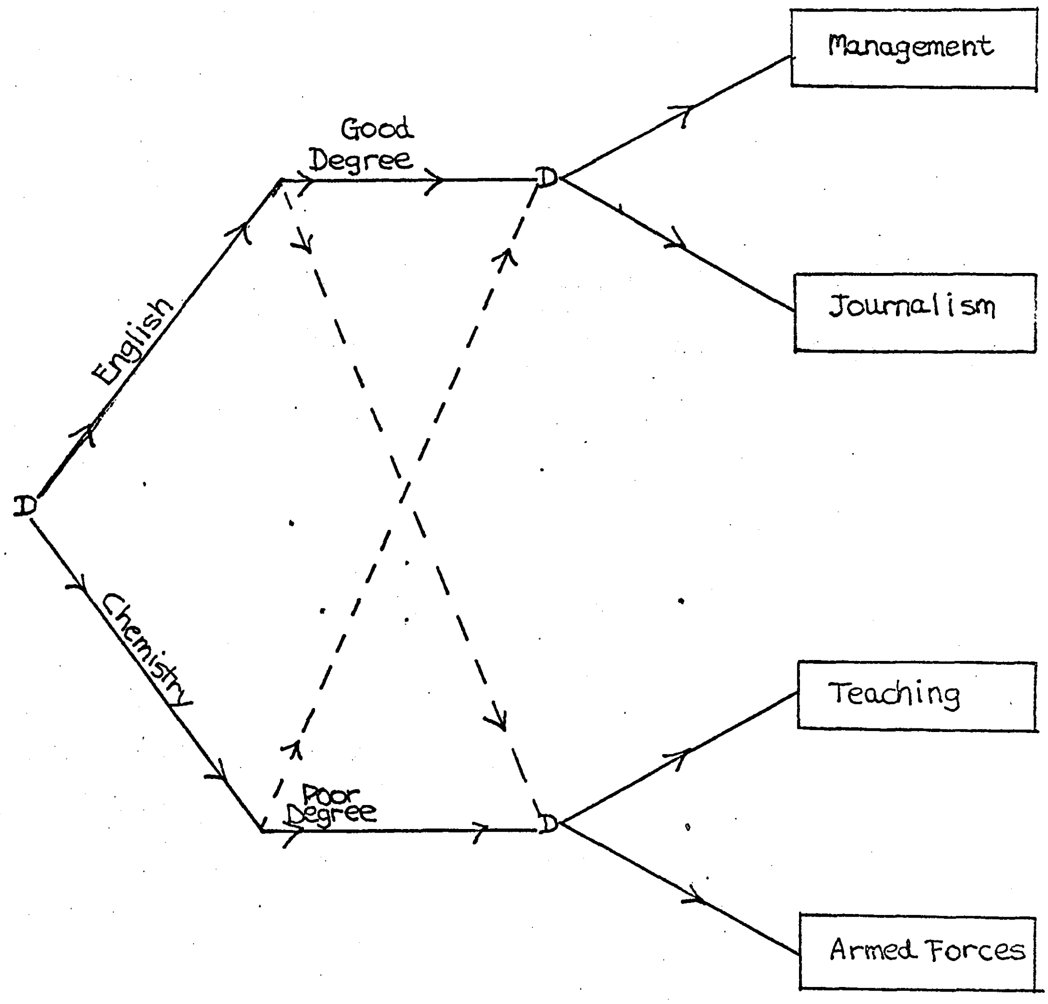
When he finishes at university he will be in a situation which requires further action. If he has a good degree in either subject he will have the opportunity (perhaps) to begin a career in some kind of journalism (technical or otherwise) or some kind of industrial management. If he has only a fair degree he may have to choose between teaching and becoming an officer in the army. Suppose he weighs up these four kinds of career in terms of job satisfaction and salary and arrives at the following conclusions:

i) management is likely to give an excellent salary but only fair job satisfaction

ii) journalism is likely to give a fair salary and excellent job satisfaction

Figure 1.1.

A Decision Tree Representation of the Two-stage, Dynamic Career Choice Problem.



Nodes marked with a D indicate the decision points.

iii) the other two are only fair in both dimensions.

It is clear that there is much for him to ponder in making his initial decision. As well as his potential ultimate career, he must consider what his chances are of a good degree for each subject choice, and what immediate pay-offs each initial action could bring. The situation can be summarized by a decision tree, as shown in figure 1.1. The decision tree representation assumes that the decision making situation can be viewed as a branching process. Some branches (marked at the node with a D) represent possible actions while others represent the possible events that are beyond the control of the decision maker. Generally, all branches are portrayed in chronological order. It is a model of the actual decision making situation and it will be a good one if it includes all the possible actions and events that are of concern to the individual making the decisions.

The above is an example of a two-stage, dynamic decision situation. The term dynamic will be used here to refer to situations in which a sequence of decisions are made such that the probability of the availability of later decisions is dependent on earlier decisions and/or events. In the literature, the term sequential is often used for this case but here the term sequential will be used to describe situations involving a sequence of decisions whose availability is independent of earlier events and actions. In sequential decision-making situations, the individuals' fortunes will fluctuate but not the availability of future actions.

As an example of a sequential decision situation, consider a taxi-driver who, after each fare, must decide which of three ranks to go to for his next fare. He knows roughly the chances at each that he will have a long wait and that he will get a profitable fare. He must weigh these factors up in deciding, while also bearing in mind that they may change throughout the day and from one day to another. For instance, one rank may be near the railway terminus. Here business may be brisk after the arrival of certain trains but only fair at other times. At other ranks his expected waiting time and his chance of a profitable fare may also be affected by events which follow some cyclic or seasonal pattern. If no action affects the future availability of his alternative ranks then this would appear to be an example of a sequential decision situation. This distinction between dynamic and sequential environments, although subtle, is rather important. By making it one can study the effect of sequential factors in the absence of certain dynamic ones, enabling a truer picture of decision making behaviour to emerge. When observing real life behaviour, and in some experiments too, the sequential and dynamic factors are confounded enabling many conflicting hypotheses to survive side by side.

A single decision, one which bears no relation to earlier events and which has bearing only on the immediate future is generally referred to as a static choice since the temporal aspects of it are minimal. (Another way of looking at a sequential situation is as a sequence of independent static ones).

It may be argued that no situation is static and they are thus too unrealistic to be worth studying. However, there are situations in which the temporal factors are of little importance For instance, suppose you had decided to take a holiday abroad for the first time and you had fixed your budget. You may feel that your past holiday experience is irrelevant to this new venture, and though you recognize the possibilities of catastrophe you are prepared to ignore them in weighing up the alternatives. Such a situation might occur when you have short-listed certain "package tour" type holidays. Your choice among these might be in essence a choice from a set of descriptions in the tour operators' brochures. Then your choice among possible holidays can be regarded as a static choice. These three examples are intended to illustrate the essential differences among the three basic kinds of situation that have featured in decision making studies.

From the general discussion so far, it is clear that what consitutes a decision situation, and what type it is, is very much in the eye of the decider. An individual faces a decision situation if he percieves it as such. It will be a static, sequential or dynamic decision depending on what he sees as important. However, in an experiment the situation can be structured so that the way it is perceived by the subject is likely to be the same as the way the experimenter perceives it. Since this has been done in the experiments which follow, decision situations will be discussed as if they are objective reality. Formal models of their structure will be given, beginning with a model of the static decision situation. This will help to make the assumptions underlying the experiments explicit.

The Static Decision Situation

The static model can be considered loosely as the sequence of events: the decision maker (DM)

i) enters decision state

ii) obtains information about the available alternatives

iii) makes choice

iv) outcome occurs

v) leaves decision state.

The "decision state" simply means that interval between the germination of the idea that action must be taken and the reaching of the goal.

Suppose, at a particular instant in time, DM enters any one of a set, S, of possible decision states. Let the state DM finds himself in be denoted Sj (where j = 1, ...n). In expanding the discussion of the structure of Sj it is necessary to define an outcome of a decision. This is done in terms of a MOTIVATING VARIABLE, which is in general multivariate and can be denoted by a vector $\underline{x} = (x_1, \ldots x_m)$. Each dimension of $\underline{x}$ represents an aspect of DM's status that he can alter by his decisions (to a greater or lesser extent) and that also have some value to him. Thus, it is assumed that these aspects can be mapped onto either a discrete or continuous variable.

In general $\underline{x}$ represents a collection of variables because obviously a choice can alter more than one aspect of DM's status. Let the status quo value of $\underline{x}$ be $\underline{x}_0$, that value which applies when he enters the state. The OUTCOME is defined as the change in the motivating variable that occurs after a decision has been made. It can be denoted by $\underline{x}_i = (x_{ii}, x_{2i}, \ldots x_{mi})$. The value of $\underline{x}$ when DM leaves the state is $\underline{x}_0 + \underline{x}_i$ if outcome $\underline{x}_i$ occurs.

A DECISION ALTERNATIVE can be defined from the outcome. Let the set of alternative decisions available in state Sj be denoted by Dj, where $d_{ij} \in D_j$ $(i = 1, \ldots k_j)$, and the number of alternatives available in S_j is k_j. A decision alternative d_{ij} is a random variable over the set of possible outcomes X_{ij} with elements $x_{ij} \in X_{ij}$. The decision alternative may be either a discrete or a continuous random variable. The marginal distributions of each dimension of $\underline{x}$ can be considered. Let $x_{cij} \in X_{cij}$, where X_{cij} is the set of possible outcomes for dimension c and decision d_{ij}. The marginal distributions in the discrete and continuous case respectively are denoted by $\left\{ P_{kcij} : x_{kcij} \right\}$ and $p\,(x_{cij})$. It will be easier to discuss the case where the marginals adequately describe all decisions, d_{ij} though it is not necessary to do this.

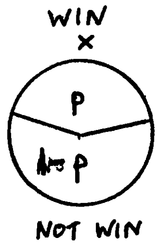
This basic model of static choice situations, where each alternative is represented by a set of marginal probability distributions has been widely used and seems quite adequate. However, it would not generally give an adequate description of how things are presented to DM even though it may describe the 'true' nature of the alternatives facing him. As suggested earlier, he is probably very ill-informed about the possible outcomes of his actions when he first perceives that an action is required, and he will first engage in some kind of information seeking activity. We have restricted our interest to how he makes his choice when he is fully informed. This is taken to mean when he has information in a form equivalent to the set of marginals. Two reflections about real-life decision making should be noted i) people rarely have accurate information when they make their choices i) there is a limit to the degree of complexity of information they can (or would wish to) handle. A choice must be made between studying behaviour either in situations characterized by information overload and inaccuracy or in situations which are relatively simple, where variables are more easily controlled. In this study the latter approach has been selected on the grounds that many interesting problems can be considered in simple situations and it is not clear as yet which models describe behaviour in them best. Another reason is so that the temporal factors of dynamic and sequential situations can be examined.

As far as static situations are concerned, this approach requires that the alternatives available to DM should be adequately described by marginal distributions which have certain properties i) simplicity, such that an "average" DM will understand them and ii) the way the information is presented should be typical of the way it is presented in the real world. It is felt that these conditions are met by discrete marginal distributions as long as there are not too many possible outcomes. Such an alternative is described by a set of values along the dimension in question and an associated set of probabilities. Alternatives are often discussed in these terms, and so decision making among discrete alternatives will mainly be dealt with.

It is worth noting that continuous alternatives can usually be adequately represented in simple value and probability terms. A step function could be fitted to most continuous distributions approaching higher criteria of adequacy as the number of steps is increased. Then, the bounds of each step and the probability of an outcome within the bounds could be given, defining the step function completely. Thus, models which apply to the discrete case are not specific to them as information about continuous alternatives can be conveyed in the same terms. There is inevitably some loss of accuracy with the step function representation but in many cases it would not be significant.

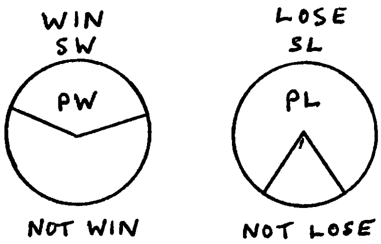
The only static situations which have been considered in the present study are those where a set of discrete, marginal probability distributions define each alternative exactly. In such cases there is complexity enough to make the study of DM's behaviour difficult. Suppose he could choose among three alternatives and his choice would affect two aspects of his status. If, for each aspect, three outcomes per alternative were possible, then his decision would involve a comparison of thirty-six items of information. The object of the study is to describe and predict how people make such comparisons. It is possible to do this in an idealized static situation so it will be used. When decision-making behaviour in non-static situations is considered the emphasis will also be on situations involving complete, exact information.

Some of the kinds of alternatives which have been used in studying decision making under uncertainty in static situations will now be described. Gambling situations have generally been used to examine static decision making under uncertainty. Subjects have been asked to choose among simple two parameter gambles of the following type (e.g. Tversky (1967b)), involving a wheel of fortune and a spinner. To play, the spinner is spun, and if it falls in the upper sector of the wheel (see diagram) the gambler wins or loses an amount x and otherwise



he wins zero. In the diagram, p, the proportion of the upper sector represents the chance of a win. This simple gamble is thus described by the couple (x,p).

Choices among duplex gambles, which have two wheels of fortune, have also been studies. For one play, the two spinners are used simultaneously. The left hand one determines the winnings and the right hand one determines the losses. Each wheel is like a simple gamble, SW being amounts won and SL being amounts lost. The events complementary to SW and SL are zero. Each gamble is summarized by the four parameters SW, SL, PW and PL. (See diagram below).



Although these alternatives involve only a one dimensional motivating variable, the information that DM has is multi-dimensional. Thus simple alternatives which involve risk can be quite complex even if the alternatives only affect DM's status in one respect.

Sequential and Dynamic Decision Situations

In order to find out if, and how people's behaviour in static situations differs from their behaviour in non-static situations it is necessary to be clear about certain important non-static situational factors. Formal definitions of sequential and dynamic decision situations can aid this. The models to be suggested are discrete time models which assume that DM passes through a sequence of decision states, the sequencing being controlled by some stochastic process. This kind of model has been used by Rapoport (1967a) as a framework for studying decision making, and is commonly used in economics and operations research.

Let there be a set of time stages, $t_1 ... t_n$ and associated with each a set of static decision states $S_1, ... S_n$. The sequential decision situation can be characterized by a sequence of random variables $\{X_i\}$ where each random variable is over the set S_i. If state $s_{ji} \in S_i$ occurs at time t_i this is denoted by $X_i = s_{ji}$. The set of conditional probabilities $p(X_i = s_{ji} | X_{i-1} = s_{j(i-1)}, .. X_1 = s_{j1})$ are sufficient to define any sequential situation, but not any dynamic one.

Consider a DM at the start of time stage t_i. He has experience of the outcomes of previous choices he has made and the motivating variable has gone though a series of fluctuations. Two important empirical questions must be: do past experiences of outcomes or current value of the motivating variable seriously affect his decisions? This study will be directed at whether effects due to these factors can be reliably reproduced under controlled conditions. In sequential situations, in the present study, DMs will only be completely informed of their alternatives at the beginning of each stage.

Also, only a particular type of sequential situation will be studied, the independent one, where $p(X_i = s_{ji}) = p(X_i = s_{ji} \mid X_{i-1} = s_{j(i-1)}, \dots X_1 = s_{j1})$, for all $i = 1, \dots n$ and all s_{ji}. This is to reduce the number of variables which must be taken into account by subject and experimenter.

Dynamic decision situations can also be characterized using a sequence of random variables $\{X_i\}$ over static decision states. This time, its characterization is completed with a different set of conditional probabilities.

$$p(X_i = s_{ji} \mid X_{(i-1)} = s_{j(i-1)},\ d_{kj(i-1)} : \dots\dots\dots : X_1 = s_{j1},\ d_{kj1})$$

where $d_{kj(i-1)}$ is the kth decisions available in state $S_{j(i-1)}$.

It can be seen that at time stage t_{i-1} the decision made will affect the probability of being in certain states at stage t_i. If DM is aware of this how does it affect his present decisions? Questions along these lines will be examined, though only for relatively simple dynamic situations. Obviously it does not take too large or complex a situation of this type before a fully informed individual will be overloaded with information. Behaviour in situations where individuals are overloaded needs to be studied, but this interesting problem, like so many others must be outside the scope of this study.

The situational factors of sequential decision situations are shared by dynamic ones. They could, and have been studied in dynamic environments. When they have been studied in this way, though they have generally been confounded with other dynamic factors. The result has been that explanations of people's decision making have been ambiguous. It has been difficult to determine which factors and combinations of factors have influenced people. In a field study observing betting behaviour at a race course, for instance, one might observe that the size of recent gains and losses has a great influence on how much is bet on the next race. In this dynamic situation one cannot tell whether this is due to the gains and losses themselves or to the effects they have on the opportunities for future bets, or both. This will be discussed with respect to other studies in chapter 4. The intention in the present study is to try and utilize the full power of the experimental method to "unconfound" such situational factors. The result will necessarily be a rather painstakingly slow development.

The formalization of sequential and dynamic decision situations has obviously not been thorough. It has been selectively directed towards the variables which will be studied experimentally. A comprehensive theoretical analysis of nonstatic decision situations does not seem necessary at this stage since it would not be possible to follow it up with a comprehensive experimental analysis. It is hoped that the discussion of the three types of decision situation has delineated the framework within which the present study of information integration models of behaviour is set. The problems which are to be investigated will be summarized after the models of behaviour have been discussed.

Information Integration Models

Information integration models were defined earlier as models which state that people integrate all the information they have about an available alternative to form an overall assessment of its subjective value. In decision making contexts it is then assumed that they compare the subjective values of all the alternatives and choose that with the highest subjective value. This brief definition will now be expanded. Further assumptions are made to obtain explicit models of behaviour.

If people assess the subjective values of whole alternatives one must assume they begin with assessments of the subjective values of the constituant attributes of the alternatives. Suppose items of information on the different dimensions can take on continuous values. The assumption is made that individuals' evaluations of these items are homomorphic to some continuous numerical scale either subordinal, ordinal or at some higher level. The counterpart to this is the assumption that the subjective values of the alternatives themselves are also homomorphic to at least a partially ordered numerical scale.

The assumptions that evaluations are scaleable are not sufficient to give explicit models in cases where alternatives are multidimensional. People also perform "integrative" operations on the sub-evaluations, which must be incorporated into the model. In a particular situation people are assumed to adopt an information processing strategy (the set of integrative operations on the constituant subjective values) on their evaluations. This is assumed to be homomorphic with a set of mathematical functional operations on the corresponding numerical scale values. That is, specific functional relationships are assumed to exist among numerical scales homomorphic to subjective value scales of the information dimensions and the whole alternatives. The explicit models of the cognitive processes underlying decision making that are proposed are all of this functional relationship type. The great advantage of making verbal explanations explicit in this way is that their consequences can be determined analytically and tested empirically.

If conclusions from studying such models are to have any general validity the dimensions of information that subjects are confronted with in the experiments must have properties advocated earlier. They must be in a form that DM can understand and in the kinds of terms that he is in the habit of using. Slovic and Lichenstein (1968 a)) suggested that people typically think of possible courses of action in terms of four risk dimensions: possible gains, possible losses, the chance of a gain, and the chance of a loss. This suggestion is in the context of static situations. For dynamic situations one could suggest additional dimensions such as favourable and unfavourable future decision states and the chances associated with them. Slovic's duplex gambles were studied by him because their four parameters correspond to the four risk dimensions and each parameter can be manipulated independently of the others. It seems a reasonable working assumption that people use these risk dimensions so following Slovic, choices among duplex gambles will be studied.

In static situations of choice among duplex gambles two simple information integration models will be studied for the most part - the additive and the subjectively expected utility (SEU) models. Let G represent a duplex gamble and SW, PW, SL, and PL its parameters. Then the ADDITIVE model states that

$$S(G) = S_1\ (SW) + S_2\ (PW) + S_3\ (SL) + S_4\ (PL)$$ and the

SEU model states that

$$S(G) = S_1\ (PW).S_2\ (SW) - S_3\ (PL).S_4\ (SL)$$

where the S_i's are subjective value scales on the information parameters and S is the subjective value scale of the duplex gambles. These functional relationship models have received much study.

It was suggested before that if the functional relationships describe the way people integrate the subjective values of items of information then the mathematical operations of the models must correspond to actual information processing operations. This is plausible for the addition and multiplication operations of the above models. Broadly speaking, addition could correspond to some averaging or aggregating operation and multiplication to a sort of weighting operation. This must be so if the models are to explain behaviour rather than merely describe it, which is the main criterion of success of the models.

The problems which the present study sets out to investigate can now be summarized. In static situations people are to be confronted with alternatives which can be described in terms of certain risk dimensions - a possible gain, a possible loss and the chances associated with them. Models which purport to explain decision making behaviour in these situations are tested. In sequential situations the effect of 2 important variables - previous outcomes and current "wealth" on decision making will be examined. Also, tests to determine whether the models considered in static situations explain behaviour in sequential situations are carried out. In dynamic decision situations alternatives cannot be described only in terms of the above risk dimensions. They must also be described in terms of such factors as "chances of being in favourable decision states in the future." Some simple alternatives are presented to decision makers in a dynamic situation and models explaining their choices are examined.

In the next 2 chapters detailed accounts will be given of two approaches to the study of simple functional relationship models like the additive and SEU ones. The discussion will focus on these two models. They are the ones which will be examined in static and sequential situations. Information integration models more appropriate to dynamic situations will be introduced prior to the experiment which considers decision making in these situations. The methods appropriate for their examination, however are those set out in the next two chapters.

CHAPTER 2.

The Qualitative Functional Analysis (QFA) of Decision Making.

Qualitative functional analysis (QFA) is a relatively new method of studying functional relationships. It does not require any measurement beyond a nominal scale. Yet it is a powerful tool which enables one to examine the lawfulness of phenomena and search for principles which the phenomena obey. In this chapter the development of QFA is traced. It is applied to decision making in static situations and principles of behaviour are suggested for examination. The approach requires that phenomena can be interpreted as relational systems. From here familiarity with the ideas of relational systems is assumed, see for instance Suppes and Zinnes (1963).

Decision making in a static environment can be interpreted in the required manner as follows. Suppose the alternatives available to DM are from some set A, and the motivating variable, $\underline{x}$ is multi-variate with k dimensions. Suppose, also that alternative decisions can be adequately represented by discrete random variables and DM has complete information about the marginal distributions of each. Let p_i and x_i be the probabilities and values associated with the i^{th} dimension of some alternative, $a \in A$. This alternative can be represented by the vector $(p_1, x_1, \ldots\ldots p_k, x_k) \in$ $P_1 \times X_1 \times \ldots\ldots P_k \times X_k$ as well as simply by an element from A. Such a vector representation could also be used in the case where decision alternatives are represented as continuous random variables. Thus, either the set, A or the product set,

$A_1 \times X_1 \times \ldots \times A_k \times X_k$ is the domain of the relevant relational system. The relations of the systems must be related to the choices DM makes. The basic assumption made is that DM's choices are dependent on his preference for or indifference between alternatives. The precise operational meaning of these terms will be discussed later, but roughly speaking if he prefers a to b (denoted $b<a$) then he tends to choose a while if he is indifferent between them (denoted $a \sim b$) then he chooses one at random. The relations $<$ and $\sim$ are known as binary strict preference relations and binary indifference relations respectively. The weak preference relation, $a \leqslant b$ denotes that b is preferred or is indifferent to a. That is, $a \leqslant b$ if and only if $a<b$ or $a \sim b$. The sequence $\mathcal{A} = \langle A, \leqslant \rangle$ and $\mathcal{A}_{2k} = \langle P_1 \times X_1 \times \ldots P_k \times X_k, \leqslant \rangle$ are relational systems called ordered structures and conjoint structures respectively. An individual's decision-making among alternatives can be interpreted as either of the above kinds of relational system.

Qualitative functional analysis postulates the existence of numerical scales (interpreted here as subjective value scales) on the system's domain and subsets of the domain which are related to the relation in a certain way. These existence, or representation hypotheses are examined, and consequences of them are stated. The representation hypotheses are proposed functional relationships among the subjective value scales. Sets of properties, consequences of the proposed relationships which must be true if the relationships among the subjective value scales hold are sought. An application of

QFA requires an empirical investigation of the consequences of the representation hypothesis.

The SEU model can be states as a representation hypothesis on $\mathcal{A}_{2k} = \langle P_1 \times X_1 \ldots\ldots P_k \times X_k \rangle$ in a form suitable for examination by QFA as follows:

Defn. The SUBJECTIVELY EXPECTED UTILITY model (SEU) for $\mathcal{A}_{2k}$ states that there exists a subjective probability function S_i on each P_i and a utility function U_i on each X_i, $i = 1, 2, \ldots k$ such that

a) for each $(p_1, x_1, \ldots p_k, x_k), (q_1, y_1, \ldots q_k, y_k) \in P_1 \times X_1 \ldots\ldots \times P_k \times X_k$

$$(p_1, x_1, \ldots p_k, x_k) \leqslant (q_1, y_1, \ldots q_k, y_k) \Leftrightarrow$$

$$\sum_{r=1}^{k} S_r(p_r)\, U_r(x_r) \leqslant \sum_{r=1}^{k} S_r(q_r)\, U_r(y_r)$$

and b) $S(0) = 0$, $S(1) = 1$, $U(0) = 0$ and $p_i \leqslant q_i \Leftrightarrow S(p_i) \leqslant S(q_i)$.

The origins of QFA are in formal measurement theory. In fact, it is an application of the results of formal measurement theory, which has been comprehensively reviewed by Pfanzagl (1968) and, more recently, by Krantz, Luce, Suppes and Tversky (1971). The term QFA aptly describes this application, though it is not in general use. Because many developments in measurement theory are very recent there is no uniformly accepted way of using it to test functional relationships empirically. The first part of this chapter reviews some of these recent developments, and arguments about how best to use them to test functional relationships are presented. Although this argument is somewhat general the specific application to examining the SEU model is to be kept in mind. Much of the discussion is motivated by the need to develop an error theory for QFA, to enable one to determine between "observed" and "true" relationships. In the later sections of the chapter the actual application of QFA to decision-making about the concrete alternatives introduced earlier are explored in more detail.

This whole theory will allow a rather thorough examination of SEU, some alternative information integration models and some specific principles of behaviour. Significance tests for them are proposed in the final sections of the chapter. It is hoped that these provide satisfactory tests of hypotheses related to qualitative aspects of information integration models.

Ordinal Measurement Models

The SEU model as defined above is an example of a representation hypothesis on $\mathcal{A}_{2k}$. It postulates a relation between the structure $\mathcal{A}_{2k}$ and some numerical scale. Some simple representation hypotheses on $\mathcal{A} = \langle A, \leq \rangle$ which postulate some relation between $\mathcal{A}$ and a numerical scale will be reviewed. Ordinal measurement models, sets of conditions on $\mathcal{A}$ will be defined and for each pair - representation hypothesis, measurement model - a theorem will be stated. The proofs of these representation theorems are omitted but can be found in the references.

H1. The INTERVAL ORDER representation hypothesis, IORH states that there exist real-valued functions, u and p on A of $\mathcal{A}$ such that for all $a, b \in A$,

i) $a < b \iff u(a) + p(a) < u(b)$.

ii) $a \sim b \iff$ not $u(a) + p(a) < u(b)$ and not $u(b) + p(b) < u(a)$

iii) $a \approx b \iff u(a) = u(b)$

where $a \approx b \iff \{a \sim c \iff b \sim c$ for all $c \in A\}$

M1. The ordered structure, $\mathcal{A}$ is an INTERVAL ORDER $\iff$ for all $a, b, c \in A$.

i) not $a < a$,

ii) $a < b$ and $b < c \iff a < c$,

iii) $a < b$ and $c < d \iff a < d$ or $c < b$.

Consider the equivalence relation, $\approx$ defined above. Let $A/\approx$ be the set of equivalence classes that partition A. Fishburn (1970) proved the following representation therem:

RT1. I.O. $\iff$ I OR H when $A/\approx$ is countable (i.e. either denumerable or finite.)

H2. The SEMI ORDER representation hypothesis, SORH states that there exists a real-valued function, u on A of $\mathcal{A}$ such that, for all $a, b, c \in A$

i) $a < b \iff u(a) + 1 < u(b)$

ii) $a \sim b \iff$ not $u(a) + 1 < u(b)$ and
not $u(b) + 1 < u(a)$

iii) $a \approx b \iff u(a) = u(b)$

M2. The ordered structure $\mathcal{A}$ is a SEMI-ORDER $\iff$ for all $a, b, c \in A$ i) it is an interval order
ii) $a < b$ and $b < c \iff a < d$ or $d < c$.

The condition ii) of M2 is called the semi-order condition and iii) of M1 the interval order condition.

Scott and Suppes (1958) proved the following representation theorem: RT2. SO $\Longleftrightarrow$ SORH when $A/\approx$ is countable.

H3. The WEAK ORDER representation hypothesis, WORH states that there exists a real valued function, u on A of $\mathcal{A}$ such that, for all a, b, c $\in$ A:

i) $a \prec b \Longleftrightarrow u(a) < u(b)$

ii) $a \sim b \Longleftrightarrow u(a) = u(b)$

M3. An ordered structure, $\mathcal{A}$ is a WEAK ORDER $\Longleftrightarrow$ for all a, b, c $\in$ A

i) not $a \prec a$

ii) either $a \prec b$, $b \prec a$ or $a \sim b$.

iii) $a \prec b$ and $b \prec c \Rightarrow a \prec c$

iv) $\sim$ on A is an equivalance relation i.e. $a \sim a$; $a \sim b \Longleftrightarrow b \sim a$; $a \sim b$ and $b \sim c \Rightarrow a \sim c$.

Suppes and Zinnes (1963) give the proof for the representation theorem: RT3, WO $\Longleftrightarrow$ WORH whether $A/\approx$ be finite or infinite. They do not cite the original sources of the proof, however.

The maps u(A) and p(A) of H1 and u(A) of H2 and H3 are numerical scales. The representation theorems prove that the set of axioms of the measurement model are necessary and sufficient for the existence of the appropriate scale. It is clear that to show that the representation is true for the set A and therefore that the scale exists one must show that the axioms of the model hold.

The function, u in all cases is the basic scale. The function, p of H1 is interpreted as ascribing a region of indifference around each point on the scale. H2 and H3 are more restricted versions of H1. In H2 the indifference regions about any point are the same length while in H3 the indifference regions are of zero length. In this sense H3 can be seen to be stronger than H2 which is stronger than H1. Furthermore, WO$\Rightarrow$ SO$\Rightarrow$ IO and WORH $\Rightarrow$ SORH $\Rightarrow$ IORH. If a scale includes an indifference region it is called an inexact scale and otherwise an exact scale. Consider the way to use these results for qualitative functional analysis when A is a set of alternatives and $\preccurlyeq$ is DM's weak preference relation. One has a hypothesis that a certain type of subjective value scale exists for DM with respect to the set of alternatives, A. One determines the $\preccurlyeq$ relation over the set A empirically, and discovers whether the qualitative conditions (axioms) of the appropriate measurement model are satisfied. Each axiom is a principle of behaviour. If they all are found to hold then the subjective value, or utility scale associated has been shown to exist, since collectively the axioms are necessary and sufficient conditions for the existence of the scale.

Let us consider the exact model a little further. The representation theorem, TR3 states that if the exact representation holds both $<$ and $\sim$ are transitive. The violation of transitivity of $\sim$ is not fatal as TR1 and TR2 show that alternative, inexact representations may hold. The violation of transitivity of $<$, however leaves us with no representation at all. It has been argued, (see Adams, 1965) that the observation of indifference between two different elements simply means that the observation procedure is insensitive to the difference which really exists. In this case $\sim$ could not be transitive and the exact model could only hold to a given approximation. Its rejection on the grounds of intransitive $\sim$

would therefore be somewhat arbitrary in general. Nevertheless, for a given set of alternatives, A and a certain observation procedure one could determine whether $\sim$ was transitive. If not, the exact model could be rejected as a suitable model of preference for the set A and the inexact models could be considered.

Unfortunately, the results of measurement theory do not provide such a clear qualitative analysis in more complex cases such as the following conjoint representation hypotheses.

Conjoint Measurement Models

Since the representations are being considered theoretically for the moment, let us denote the n-dimensional conjoint structure by $\mathcal{A}_n = \langle A_1 x \ldots A_n, \preccurlyeq \rangle$. The basic conjoint representation hypotheses are those for $\mathcal{A}_2 = \langle A_1 \text{ x } A_2, \preccurlyeq \rangle$. Only one representation hypothesis for $n > 2$ will be considered for the moment. As in the previous section, representation hypotheses and measurement models will be stated, together with theorems which link them.

<u>H4. The ADDITIVE CONJOINT representation hypothesis</u> for $\mathcal{A}_2$, ACRH states that there exist real-valued functions u_1 on A_1 and u_2 on A_2 both of $\mathcal{A}_2$ such that, for all $a_1, b_1 \in A_1$ and $a_2, b_2 \in A_2$

i) $(a_1, a_2) \prec (b_1, b_2) \Leftrightarrow u_1(a_1) + u_2(a_2) < u_1(b_1) + u_2(b_2)$

ii) $(a_1, a_2) \sim (b_1, b_2) \Leftrightarrow u_1(a_1) + u_2(a_2) = u_1(b_1) + u_2(b_2)$

A measurement model related to H4 states four axioms on $\mathcal{A}_2$, the last of which requires the following concept.

Definition. A DUAL STANDARD SEQUENCE, DSS for $\mathcal{A}_2$ is a pair of doubly infinite sequences a_{1i}, a_{2i}, $i = 0, \pm 1, \pm 2, \ldots$ from A_1 and A_2 respectively such that if $i + j = p + q$ then $(a_{1i}, a_{2j}) \sim (a_{1p}, a_{2q})$. The DSS is trivial if (a_{1i}, a_{2i}) for all i are equal.

M4. A conjoint structure $\mathcal{A}_2$ is an ADDITIVE CONJOINT MEASUREMENT model,

ACMM $\Leftrightarrow$ for all (a_1, a_2), $(b_1, b_2) \in A_1 \times A_2$:

i) $\lesssim$ is a weak order on $A_1 \times A_2$

ii) If $(a_1, a_2) \lesssim (b_1, b_2)$ and $(b_1, c_2) \lesssim (c_1, a_2)$ then $(a_1, c_2) \lesssim (c_1, b_2)$

iii) there exist $d_1 \in A_1$ and $d_2 \in A_2$ such that $(a_1, a_2) \sim (b_1, d_2)$ and $(a_1, a_2) \sim (d_1, b_2)$

iv) for any non-trivial DSS (a_{1i}, a_{2i}) $i = 0, \pm 1, \ldots.$ there exist integers n, m such that $(a_{1n}, a_{2n}) \lesssim (a_1, a_2) \lesssim (a_{1m}, a_{2m})$

Luce and Tukey (1964) proved the representation theorem: T4. ACMM $\Rightarrow$ ACRH when $A_1 \times A_2/\approx$ is not countable. Scott (1964) proposed a theorem and model for the case $A_1 \times A_2/\approx$ countable. His model consists of an infinite bundle of axioms which are necessary and sufficient for ACRH. Note that Luce and Tukey show only that ACMM is sufficient for ACRH, not that it is necessary.

The other basic hypothesis is as follows.

H5. The MULTIPLICATIVE CONJOINT representation hypothesis, MCRH states that there exist real valued functions u_1 on A_1 and u_2 on A_2 of $\mathcal{A}_2$ such that, for all (a_1, a_2), $(b_1, b_2) \in A_1 \times A_2$:

i) $(a_1, a_2) < (b_1, b_2) \Leftrightarrow u_1 (a_1) \times u_2 (a_2) < u_1 (b_1) \times u_2 (b_2)$

ii) $(a_1, a_2) \sim (b_1, b_2) \Leftrightarrow u_1 (a_1) \times u_2 (a_2) = u_1 (b_1) \times u_2 (b_2)$

A model related to H5 requires the following concept.

Definition.

The zero subset of A_1, $O_{A1} \equiv \{a_{o1} | a_{o1} \in A_1 \text{ and } (a_{o1}, a_2) \sim (a_{o1}, b_2) \text{ for all } a_2, b_2 \in A_2\}$

The ZERO subset of A_2 is

$O_{A2} \equiv \{a_{02} | a_{02} \in A_2 \text{ and } (a_1, a_{02}) \sim (b_1, a_{02}) \text{ for all } a_1, b_1 \in A_1\}$

The ZERO subset of $A_1 \times A_2$ is

$$O_{A1 \times A2} \equiv O \equiv (O_{A1} \times A_2) \cup (A_1 \times O_{A2})$$

R. Roskies (1965) proposes a model of six axioms which is sufficient for MCRH when O is not empty. A slight amendment can deal with the case O empty. A. Gioia (1967) shows that Roskies' axiom four follows from the others. Only the case when O is not empty will be considered here, and to aid understanding axiom four is included. The following relations are required for the case O not empty.

Definition. The relations s and -s hold as follows:

$(a_1, a_2) \; s \; (b_1, b_2)$ if $\{(a_1, a_2) < 0 \text{ and } (b_1, b_2) < 0\}$

or $\{0 < (a_1, a_2) \text{ and } 0 < (b_1, b_2)\}$

$(a_1, a_2) - S \; (b_1, b_2)$ if $(a_1, a_2) < 0$ and $0 < (b_1, b_2)$

or $0 < (a_1, a_2)$ and $(b_1, b_2) < 0$

M5. A conjoint structure $\mathcal{A}_2$ is a MULTIPLICATIVE CONJOINT MEASUREMENT MODEL, MCMM $\Leftrightarrow$ for all $(a_1, a_2), (b_1, b_2) (c_1, c_2) \in A_1 \times A_2$

i) $\precsim$ on $A_1 \times A_2$ is a weak order

ii) If $d_1 \notin O_{A_1}$ there exists $d_2 \in A_2$ such that $(a_1, a_2) \sim (d_1, d_2)$
If $d_2 \notin O_{A2}$ there exists $d_1 \in A_1$ such that $(a_1, a_2) \sim (d_1, d_2)$

iii) If $b_1 \notin O_{A1}$ or $b_2 \notin O_{A2}$ and if $(a_1, b_2) \sim (b_1, c_2)$ and $(b_1, a_2) \sim (c_1, b_2)$ then $(a_1, a_2) \sim (c_1, c_2)$

iv) Let $a_1, b_1 \notin O_{A1}$. If $(a_1, a_2)\ S\ (b_1, a_2)$ for some $a_2 \in A_2$ then $(a_1, b_2)\ S\ (b_1, b_2)$ for all $b_2 \in A_2$. Similarly for a_2, $b_2 \notin O_{A2}$. If $(a_1, a_2)\ S\ (b_1, a_2)$ then we say that $a_1\ S\ b_1$. Similarly for $a_2\ S\ b_2$. All of iv) also holds for the relation $-$ S.

v) Suppose $a_1, b_1, \notin O_{A1}$, and $a_2, b_2 \notin O_{A2}$. Then
if $a_2\ S\ b_2$ and $(a_1, a_2) \precsim (b_1, a_2)$ then $(a_1, b_2) \precsim (b_1, b_2)$
if $a_2 - S\ b_2$ and $(a_1, a_2) \precsim (b_1, a_2)$ then $(b_1, b_2) \precsim (a_1, b_2)$
if $a_1\ S\ b_1$ and $(a_1, a_2) \precsim (a_1, b_2)$ then $(b_1, a_2) \precsim (b_1, b_2)$
and if $a_1 - Sb_1$ and $(a_1, a_2) \precsim (a_1, b_2)$ then $(b_1, b_2) \precsim (b_1, a_2)$.

vi) the archemedian axiom, iv) of M4 holds over $A_1 \times A_2{}^+$ where $(a_1, a_2) \in A_1 \times A_2^+$ if $0 < (a_1, a_2)$.

Roskies proved the following representation theorem:

T5. MCMM $\Rightarrow$ MCRH when $A_1 \times A_2/\approx$ is not courtable. As with T4, Roskies six axioms are sufficient for H5 though not all are necessary. Roskies model is a generalization of Luce and Tukey's model, M4. A second generalization of M4 is the model proposed by Luce (1966) which is sufficient for the following representation hypothesis.

HA. The K-DIMENSIONAL ADDITIVE CONJOINT representation hypothesis KACRH states that there exist real valued functions u_1 on A_1, u_2 on A_2, ... u_k on A_k of $\mathcal{A}_k$ such that for all $(a_1, \ldots a_k)$, $(b_1, \ldots b_k)$

$\in A_1 x \ldots x A_k$:

$(a_1, a_2, \ldots a_k) \leqslant (b_1, \ldots b_k) \Leftrightarrow u_1(a_1) + \ldots + u_k(a_k) \leqslant u_1(b_1) + \ldots u_k(b_k)$

The measurement model related to this makes use of the following generalization of a double standard sequence and the concept of a standard sequence.

<u>Defn. A double infinite sequence</u> of pairs $\{a_i^r, a_i^s\}$, $i = o, \pm 1, \pm 2, \ldots\ldots$ (where for each i, $a_i^r \in A_r$ and $a_i^s \in A_s$) is a <u>DUAL STANDARD SEQUENCE, DSS</u> if for each $i = 0, \pm 1, \ldots$

i) $(a_i^r, a_{i+1}^s) \sim (a_{i+1}^r, a_i^s)$ ii) $(a_{i+1}^r, a_{i-1}^s) \sim (a_i^r, a_i^s)$,

the elements not made explicit being constant.

<u>Defn</u>. A sequence $\{a_i^r\}$ $i = 0, \pm 1, \ldots$ is a STANDARD SEQUENCE, SS of A_r if it is a member of some DSS. It is increasing if $a_i^r \leqslant a_{i+1}^r$ for all i.

<u>MA. The conjoint structure $\mathcal{A}_k$, is a K-DIMENSIONAL ADDITIVE CONJOINT measurement model</u>, KACMM if and only if, for all elements of $A_1 x \ldots x A_k$:

i) $\leqslant$ on $A_1 x \ldots x A_k$ is a weak order.

ii) For any integer j, $1 \leqslant j \leqslant k$ and any $a^i \in A_i$, $1 \leqslant i \leqslant k$ and any $b^i \in A_i$, $i \neq j$ there exists an $x^j \in Aj$ such that $(a^1, \ldots a^{j-1}, a^j, a^{j+1}, \ldots a^k) \sim (b^1, \ldots b^{j-1}, x^j, b^{j+1}, \ldots b^k)$.

iii) Consider any two inequalities $(a_i^1, \ldots a_i^k) \leqslant (b_i^1, b_i^2, \ldots b_i^k)$ $i = 1, 2$ with the property that for each $j = 1, 2, \ldots k$ there exists a corresponding permutation J of 1 and 2 such that $a_i^j = b_{J(i)}^j$ except for at most one i, which if it exists is denoted by i^* Then, $(a^1, a^2, \ldots a^k) \leqslant (b^1, \ldots b^k)$ where if i* exists $a^j = a_{i*}^j$ and $b^j = b^j_{J(i*)}$ and if i* does not exist then $a^j = b^j =$

any element of Aj.

iv) For any non-trivial increasing SS $\{a_i^r\}$, $1 \leqslant r \leqslant k$, such that any pair form a DSS, and any b in A_r there exist integers m and n such that $a_m^r \leqslant b \leqslant a_n^r$.

In the case of K= 2 the four axioms of MA reduce to those of M4. The weak order, solution of equations and Archemedian axioms (i), ii) and iv)) are readily seen to be generalizations of the corresponding axioms of M4. The generalized double cancellation condition, axiom iii) is not so obviously related to the corresponding axiom of M4. It is a concise statement of the complete set of double cancellation conditions i.e. those cancellation conditions in which a pair of inequalities imply a third. This set is finite though rather large as k increases. Fortunately many of the conditions are trivial. For instance when k = 2 the axiom includes four conditions of which only one is non-trivial. The case when k = 3 will be considered later. The set of non-trivial cancellation conditions for this case has six members. The six models presented in this and the preceding sections are the only ones from measurement theory that will be used in this study, though other representation hypotheses will be examined for which measurement models like M1 - MA are available. Now, H4-HA can be interpreted as special cases of SEU in certain simple situations. They propose that subjective value scales exist on A_i, i = 1, ... k and also on $A_1 x \ldots x A_n$ and that they relate according to a particular rule. An ideal qualitative functional analysis of these representation hypotheses would involve exhaustive examination of sets of axioms of measurement models which were necessary and sufficient for them. Unfortunately, M4 - MA are not suitable for this for two reasons.

Firstly, they are merely sufficient for H4 - HA and negation of them does not imply negation of the representation (this does not apply to Scotts model).

Secondly, it is not clear how to carry out an exhaustive empirical examination of all the axioms. (This does apply to Scott's model). In the next section some light is shed on the latter problem by the examination of the empirical status of axioms.

The Empirical Status of Axioms

Three discussions of this topic have appeared in recent years: Adams, Fagot and Robinson (1965), 1970), Pfanzagl (1968). They make the same basic points, which are clearly stated in Pfanzagl pp.106-109. What follows is derived from this. Suppose one wished to carry out a qualitative functional analysis of some representation hypothesis via an associated sufficient measurement model. It would be convenient to accept the hypothesis that the axioms are satisfied as long as there was no empirical evidence to the contrary. The empirical evidence, however, is often not complete, since only a finite set of data can be collected. Some axioms consist of infinite bundles of conditions e.g. Scott (1964). It is obviously impossible to falsify such axioms since to do so would require an infinite set of tests to be carried out. Other axioms are existential. They postulate the existence of certain elements which are in a certain relation to other given elements. Such axioms also cannot be shown to be false since failure to establish the existence of the elements does not prove they don't exist. Adams, Fagot and Robinson (1965) called both kinds of axioms objectionable

objectionable and axioms which are not objectionable they called testable. At first glance it would appear that to test the measurement model it is only necessary to test the testable axioms. However, though the objectionable axioms need not be tested they cannot be ignored as testable consequences may follow from a system of axioms as a whole which do not follow from the testable axioms alone. The situation can be clarified somewhat if the discussion is made formal.

Let $\mathcal{A} = \langle A, \quad (R_i)_{i \in I} \rangle$ be an empirical relational system with k_i- ary relations R_i and finite I. A subset $A_0 \subseteq A$ can be taken to test empirically whether certain sentences of the form

$$"R_i\ (a_1, \ldots a_{ki}) = 0" \quad \text{OR} \quad "R_i\ (a_1, \ldots a_{ki}) = 1"$$

are true for some $a_1, a_2, \ldots, a_{ki} \in A_0, \quad i \in I,$
where $R_i\ (a_1, \ldots a_{ki})$ is the characteristic function of the relation R_i. Let a finite set of such sentences be denoted by X. New sentences can be constructed from the elements, say $X_1, X_2 \in X$:

i) $X_1 \cup X_2$ $(\cup = \text{OR})$

ii) $X_1 \cap X_2$ $(\cap = \text{AND})$

iii) $X_1 \supset X_2$ $(\supset = \text{IMPLIES})$

iv) $\bar{X}_1$ (not X_1)

The set of sentences generated from X by the above operations defines a boolean algebra. This set can be denoted by the form:

$$\text{(i)}\ Y \equiv (X_{11} \cup X_{12} \ldots\ \cup X_{1\ell_1}) \cap (X_{21} \cup \ldots\ \cup X_{2\ell_2}) \cap \ldots \cap (X_{m1} \cup \ldots \cup X_{m\ell_m})$$

where $X_{ij} \in X$ or $X_{ij} \in \bar{X}$

Expressing the set of sentences generated by all the possible combinations of the operations i) to iv) in this form reduces the set to one very large sentence.

An instance of the sentence Y can be denoted by Y $(a_1, \ldots a_r)$ where a_i are the elements $\in A_0$ which appear in the sentence.

Certain important terms can now be defined.

<u>Defn.</u> A TESTABLE sentence of the R.S., $\mathcal{A}$ is a sentence of the type:

"Y $(a_1, \ldots a_r)$ holds for all $a_1, \ldots a_r$" where Y is a sentence of the type (1) above.

Such sentences have also been called universal sentences. Obviously, an axiom of the above form is called a testable axiom. An objectionable axiom is formally defined by negation.

<u>Defn</u>. The testable sentence "$Y(a_1 \ldots a_r)$ holds for all $a_1, \ldots a_r$" is a TESTABLE CONSEQUENCE of a system of axioms on an R.S., $\mathcal{A} = \langle A_i, (R_i) \quad i \in I \rangle \Leftrightarrow$ the testable sentence is satisfied for each RS $\mathcal{A}^1$ of the same type as $\mathcal{A}$ which also satisfies the axioms.

<u>Defn</u>. Let $\{S_1, S_2\}$ be systems of axioms. S_2 is PURELY TECHNICAL in $\{S_1, S_2\}$ if and only if each testable consequence of $\{S_1, S_2\}$ is a consequence of S_1 alone.

It can be shown that the axioms of a system S can be partitioned into three subsets: S_T, the set of testable axioms, S_{PT} the set of purely technical axioms and S_O the set of axioms which are objectionable but not purely technical. Adams, Fagot and Robinson (1965), (1970) concern themselves with the problem of partitioning certain measurement models in this fashion. They prove which axioms are purely technical for certain systems.

The meaning of such results for the axiom systems previously presented should be discussed. The partitioning of the axioms of M1 - M3 is a simple matter: they are all testable. Adams, Fagot and Robinson show that of the axioms of M4, the first two are testable, the third objectionable and the fourth purely technical. One might speculate on the empirical status of the axioms of M5 by analogy with the results for M4: axioms i), iii), iv) and v) are testable, ii) is objectionable and vi) is purely technical. A consequence of M4 is called the independence condition for $\mathcal{A}_2 = \langle A_1 \times A_2, \lesssim \rangle$: for all $a_1, b_1 \in A_1, a_2, b_2 \in A_2$,

i) $(a_1, a_2) \lesssim (b_1, a_2) \iff (a_1, b_2) \lesssim (b_1, b_2)$

and

ii) $(a_1, a_2) \lesssim (a_1, b_2) \iff (b_1, a_2) \lesssim (b_1, b_2)$.

This testable consequence follows from axioms one, two and three of M4, not from one and two alone. If the objectionable axiom three had been ignored then this rather basic consequence of the model would have been missed. However, from the empirical point of view, nothing is lost if the fourth axiom of M4 is ignored. This illustrates the value of being able to separate axioms of measurement models according to the testable, objectionable, purely technical trichotomy.

If one is examining a representation hypothesis via a measurement model one will be able to say which subset of the axioms of the model have empirical consequences.

It would be particularly useful if one found that the set of objectionable axioms was empty. Unfortunately, many of the models discussed in measurement theory have objectionable axioms.

The problem remains that for a finite set A it may not be possible to test the objectionable axioms fully because there are too many testable consequences of them or some may be unknown. The implications of this for qualitative functional analysis are considered further in the next section.

The Qualitative Function Analysis of the Additive and Multiplicative Representations for $\mathcal{A}_2$.

The measurement models of the previous sections arose from attempts to find sets of conditions which are necessary and sufficient for the representation hypotheses. The models for which this aim is realized, M1 - M3 are very useful for testing H1 - H3 empirically. But how useful are M4 and M5 for testing H4 and H5 in the light of the fact that these axiom systems are merely sufficient for the representations? To examine such axiom systems may be rather inefficient when the real object of study is the representation. In the case of H4, Scott's model (rather than Luce and Tukey's) affords, at first sight an efficient intermediary. However, as already pointed out, since the model consists of an infinite bundle of conditions it cannot be fully tested empirically. Furthermore, Scott showed that no finite set of conditions exist which are necessary and sufficient for H4 even when the domain is finite. It seems likely that the same is true of H5 and of certain similar representation hypotheses for $\mathcal{A}_3$ which will be presented shortly. In view of this it would seem that empirical studies of H4 and H5 should, like that of Adams and Fagot (1959) have more realistic aims.

Adams and Fagot sought a set of independent necessary conditions which contained much of the empirical content of H4. This seems to be the best that can be done empirically. A set of conditions contain much of the empirical content of a representation if data which satisfy the conditions are unlikely to violate the hypothesis. Conditions are independent if none follows from the remaining ones in the set. Adams and Fagot arrived at the following three conditions, known as weak order, cancellation and independence respectively:

H4C1 The relation $\lesssim$ over $A_1 \times A_2$ is a weak order.

H4C2 For all $a_1, b_1, c_1 \in A_1$ and $a_2, b_2, c_2, \in A_2$, if $(a_1, a_2) \lesssim (b_1, b_2)$ and $(b_1, c_2) \lesssim (c_1, a_2)$ then $(a_1, c_2) \lesssim (c_1, b_2)$.

H4C3 For all $a_1, b_1, \in A_1$ and $a_2, b_2 \in A_2$,

i) $(a_1, a_2) \lesssim (b_1, a_2) \Leftrightarrow (a_1, b_2) \lesssim (b_1, b_2)$

ii) $(a_1, a_2) \lesssim (a_1, b_2) \Leftrightarrow (b_1, a_2) \lesssim (b_1, b_2)$

An additional consequence of H4 is that the set of zeros is in general empty. This is expressed as follows:

H4C4. If $a < b$ for some $a, b \in A_1 \times A_2$ then the zero set,

$$0_{A_1 \times A_2} \subseteq A_1 \times A_2 \text{ is empty.}$$

As an interim solution, it is proposed that these four conditions contain a good deal of the empirical content of H4, and a test of them would provide a thorough test of H4.

It is clear that the model M4 has not been abandoned completely in this direct approach, since the first two of the above conditions are the necessary axioms of M4. Examining a set of conditions such as the above will be what is meant by the qualitative functional analysis of a representation in cases where the set of necessary and sufficient qualitative conditions is infinite. A suitable set of conditions for the QFA of H5 might be based on the axioms of M5. Consider, in the case O_{A1} and O_{A2} not empty, the conditions:

H5C1 (Sign) Let $a_1, b_1 \notin O_{A1}$. If (a_1, a_2) S (b_1, a_2) for some $a_2 \in A_2$ then (a_1, b_2) S (b_1, b_2) for all $b_2 \in A_2$. Similarly for $a_2, b_2 \notin O_{A2}$. All the above also holds for the relation -S.

H5C2 (Weak Order) The relation $\leqslant$ over $A_1 \times A_2$ is a weak order

H5C3 (Cancellation) For all $a_1, b_1, c_1 \in A_1$ and $a_2, b_2, c_2 \in A_2$

If $b_1 \notin O_{A1}$ or $a_2 \notin O_{A2}$ and

$(a_1, a_2) \sim (b_1, b_2)$ and $(b_1, c_2) \sim (c_1, a_2)$

then $(a_1, c_2) \sim (c_1, b_2)$

H5C4 (Sign dependence) for all $a_1, b_1, \notin O_{A1}, a_2, b_2 \notin O_{A2}$:

i) If a_2 S b_2 then $(a_1, a_2) \leqslant (b_1, a_2) \Leftrightarrow (a_1, b_2) \leqslant (b_1, b_2)$

ii) If a_2 - Sb_2 then $(a_1, a_2) \leqslant (b_1, a_2) \Leftrightarrow (b_1, b_2) \leqslant (a_1, b_2)$

iii) If a_1 Sb_1 then $(a_1, a_2) \leqslant (a_1, b_2) \Leftrightarrow (b_1, a_2) \leqslant (b_1, b_2)$

iv) If a_1 - Sb_1, then $(a_1, a_2) \leqslant (a_1, b_2) \Leftrightarrow (b_1, b_2) \leqslant (b_1, a_2)$

Note that in the multiplicative case there are no restrictions placed on the zero elements which may exist. Thus, if any are found one should begin to favour H5 rather than H4. The discrepancy of zero elements is a useful guide when one is considering which functional relationships seem plausible initially. If they exist they are often quite easily identified. The results on zeros are not testable sentences as defined in the last section, they are existential. Nonetheless they are included in the present exposition because of their usefulness in the early stages of any investigation. They will be discussed further in the application of QFA when plausible functional relationships are scrutinized.

Apart from the consequences about zeros, the consequences of H4 and H5 should ideally have certain properties if they are to be useful in empirical investigations. They should be i) testable ii) independent and iii) comprehensive in the sense of containing most of the empirical content of the representation. Their form shows that they are testable. Their apparent independence and comprehensiveness will not be taken issue with in the present study. Clearly they are quite basic consequences of the representations. The idea at present is to pursue QFA through to its application. Providing proofs of the consequences independence and comprehensiveness is regarded as an unnecessary refinement at present.

The QFA of Simple Polynomial Representations for $\mathcal{A}_3$.

The discussion is now extended to simple polynomial representations for $\mathcal{A}_3$ which are defined below.

Defn. Representation hypotheses for $\mathcal{A}_3$ that are SIMPLE POLYNOMINALS state that there exist real-valued functions u_1 on A_1, u_2 on A_2, u_3 on A_3 and u on $A_1 \times A_2 \times A_3$ such that for all elements $(a_1, a_2, a_3), (b_1, b_2, b_3) \in A_1 \times A_2 \times A_3$:
$(a_1, a_2, a_3) \leqslant (b_1, b_2, b_3) \Leftrightarrow u(a_1, a_2, a_3) \leqslant u(b_1, b_2, b_3)$.
The simple polynomial hypothesis is:

H6. ADDITIVE if $u = u_1 + u_2 + u_3$

H7. MULTIPLICATIVE if $u = u_1 \times u_2 \times u_3$

H8. DISTRIBUTIVE if $u = (u_1 + u_2)\, u_3$

H9. DUAL-DISTRIBUTIVE if $u = u_1 \times u_2 + u_3$

The basic requirement for a hypothesis to be a simple polynomial is that in $u = f(u_1, u_2, u_3)$ each u_i appears at most once, e.g. if $u = u_1^2.u_2.u_3$ or $u = (u_1 + u_2).u_1. u_3$ the hypothesis is not a simple polynomial. H4 and H5 are simple polynomials on $\mathcal{H}_2$.

It is clear that H6 is a special case of HA when k = 3. Therefore, when k = 3, MA is sufficient for H6. Measurement models which are sufficient for H7-H9 have recently been presented by Krantz, Luce, Suppes and Tversky (1971). Parallel to the discussion of H4 and H5, sets of consequences of H6-H9 should be sought which do not depend on each other and which contain much of their empirical content. That is, testable conditions suitable for QFA of the hypotheses are required. As in the last section, whether they in fact are independent and comprehensive will simply be assumed.

Necessary consequences of simple polynomial representations were first discussed by Krantz (1967) and later in Coombs and Huang (1970) and Krantz and Tversky (1971). The first suggested the kinds of conditions necessary and the other two developed from it. Krantz and Tversky called their analysis of H6-H9 ordinal analysis because they assumed that their data satisfied the weak order condition. Now, it was suggested earlier that the weak order condition in a preference context is quite a demanding principle of behaviour and it is desirable to isolate it from other qualitative principles which follow from the models. For this reason it was felt that Krantz and

Tversky's results could be improved if the necessity for the ordering assumption could be eliminated. This can be done by a careful generalization of the QFA results available for the simple polynomials on $\mathcal{A}_2$. This was done in an earlier draft of this chapter which was read to the Mathematical and Statistical Section of the British Psychological Society, June 1971. Since then, this generalization has also been carried out indirectly in Krantz, Luce, Suppes and Tversky (1971). The necessary axioms of their models sufficient for H7-H9 are very similar to the conditions derived during the course of work on this thesis. The conditions for QFA of H6-H9 to be presented were not based on Krantz, Luce, Suppes and Tversky (1971) but on earlier publications. The QFA of H6 is based on Luce's n-dimensional conjoint measurement model.

The QFA of H7 to H9 is based on Krantz and Tversky (1971) and requires the following definitions.

<u>Defn</u>. An element $a_1 \in {}^0A_1 \mid A_2$ is an EFFECTIVE ZERO in A_1 with respect to A_2 if, for all a_2, $b_2 \in A_2$ and $a_3 \in A_3$ $(a_1, a_2, a_3) \sim (a_1, b_2, a_3)$.

An element $(a_1, a_2) \in {}^0A_1 \times A_2$ is an EFFECTIVE ZERO in $A_1 \times A_2$ if, for all a_3, $b_3 \in A_3$ $(a_1, a_2, a_3) \sim (a_1, a_2, b_3)$.

<u>Defn.</u> The set A_1 is INDEPENDENT of $A_2 \times A_3$, for all a_1, $b_1 \in A_1$ and (a_2, a_3), $(b_2, b_3) \in A_2 \times A_3$ if

$(a_1, a_2, a_3) \leqslant (b_1, a_2, a_3) \Leftrightarrow (a_1, b_2, b_3) \leqslant (b_1, b_2, b_3)$

The set $A_1 \times A_2$ is INDEPENDENT OF A_3 if, for all (a_1, a_2), $(b_1, b_2) \in A_1 \times A_2$ and a_3, $b_3 \in A_3$,

$(a_1, a_2, a_3) \leqslant (b_1, b_2, a_3) \Leftrightarrow (a_1, a_2, b_3) \leqslant (b_1, b_2, b_3)$

The set A_1 is INDEPENDENT of A_2 if, for all a_1, $b_1 \in A_1$, a_2, $b_2 \in A_2$ and $a_3 \in A_3$.

$(a_1, a_2, a_3) \leqslant (b_1, a_2, a_3) \Leftrightarrow (a_1, b_2, a_3) \leqslant (b_1, b_2, a_3)$

A further set of definitions require the generalization of the S and -S relations of M5. If $a_i, b_i, \in A_i$ let a_i S b_i indicate that the function values of a_i and b_i have the same sign and let a_i - S b_i indicate that they have opposite sign ($i = 1, 2, 3$), when some simple polynomial representation is assumed. Under a representation which states that A_i and A_j are additive let (a_i, a_j) S (b_i, b_j) indicate that $u_i(a_i) + u_j(a_j)$ and $u_i(b_i) + u_j(b_j)$ have the same sign and let (a_i, a_j) -S (b_i, b_j) indicate they have opposite sign. If a representation states that A_i and A_j are multiplicative let (a_i, a_j) S (b_i, b_j) indicate that $u_i(a_i) \times u_j(a_j)$ have the same sign while -S indicates they have opposite signs. Later it will be shown that it is not necessary for our purposes to know how to determine empirically whether -S or S holds for particular elements.

<u>Definitions</u> The set A_1 is SIGN DEPENDENT on $A_2 \times A_3$ if, for each $a_1, b_1 \in A_1$ and $(a_2, a_3), (b_2, b_3) \in A_2 \times A_3$

i) If (a_2, a_3) S (b_2, b_3),

$(a_1, a_2, a_3) \leqslant (b_1, a_2, a_3) \Leftrightarrow (a_1, b_2, b_3) \leqslant (b_1, b_2, b_3)$

ii) If (a_2, a_3) - S (b_2, b_3),

$(a_1, a_2, a_3) \leqslant (b_1, a_2, a_3) \Leftrightarrow (b_1, b_2, b_3) \leqslant (a_1, b_2, b_3)$

The set $A_1 \times A_2$ is SIGN DEPENDENT on A_3 if for each (a_1, a_2), $(b_1, b_2) \in A_1 \times A_2$ and $a_3, b_3 \in A_3$

i) If a_3 S b_3 then $(a_1, a_2, a_3) \leqslant (b_1, b_2, a_3) \Leftrightarrow (a_1, a_2, b_3) \leqslant (b_1, b_2, b_3)$

ii) If a_3 - Sb_3 then $(a_1, a_2, a_3) \leqslant (b_1, b_2, a_3) \Leftrightarrow (b_1, b_2, b_3) \leqslant (a_1, a_2, b_3)$

The set A_1 is SIGN DEPENDENT on A_2 if, for each $a_1, b_1 \in A_1$, a_2,

$b_2 \in A_2$ and $a_3 \in A_3$:

i) If $a_2 \, S \, b_2$ then $(a_1, a_2, a_3) \leqslant (b_1, a_2, a_3) \Leftrightarrow (a_1, b_2, a_3) \leqslant (b_1, b_2, a_3)$

ii) If $a_2 - S \, b_2$ then $(a_1, a_2, a_3) \leqslant (b_1, a_2, a_3) \Leftrightarrow (b_1, b_2, a_3) \leqslant (a_1, b_2, a_3)$

The suggested set of QFA conditions for H6 is the following:

H6C1. (Zeros) If H6 holds then no effective zero sets exist in $A_1 \times A_2 \times A_3$.

H6C2. (Weak order) The relation $\leqslant$ over $A_1 \times A_2 \times A_3$ is a weak order.

H6C3. (Cancellation) For all $a_i, b_i, c_i, \in A_i$ where $i = 1, 2, 3$:

i) $(a_1, a_2, a_3) \leqslant (b_1, b_2, b_3)$ and $(c_1, b_2, b_3) \leqslant (a_1, a_2, c_3)$ imply $(c_1, c_2, a_3) \leqslant (b_1, c_2, c_3)$

ii) $(a_1, a_2, a_3) \leqslant (b_1, b_2, b_3)$ and $(b_1, c_2, b_3) \leqslant (c_1, a_2, a_3)$ imply $(a_1, c_2, c_3) \leqslant (c_1, b_2, c_3)$

iii) $(a_1, a_2, a_3) \leqslant (b_1, b_2, b_3)$ and $(b_1, b_2, c_3) \leqslant (a_1, c_2, a_3)$ imply $(c_1, a_2, c_3) \leqslant (c_1, c_2, b_3)$

iv) $(a_1, a_2, a_3) \leqslant (b_1, b_2, b_3)$ and $(b_1, b_2, c_3) \leqslant (c_1, c_2, a_3)$ imply $(a_1, a_2, c_3) \leqslant (c_1, c_2, b_3)$

v) $(a_1, a_2, a_3) \leqslant (b_1, b_2, b_3)$ and $(b_1, c_2, b_3) \leqslant (c_1, a_2, c_3)$ imply $(a_1, c_2, a_3) \leqslant (c_1, b_2, c_3)$

vi) $(a_1, a_2, a_3) \leqslant (b_1, b_2, b_3)$ and $(c_1, b_2, b_3) \leqslant (a_1, c_2, c_3)$ imply $(c_1, a_2, a_3) \leqslant (b_1, c_2, c_3)$

H6C4. (Independence) The sets $A_i \times A_j$ are independent of A_k for $i, j, k = 1, 2, 3, i \neq j \neq k$.

The second and third conditions are the weak order and cancellation axioms of MA when $k = 3$. The independence condition is analogous to the condition H4C3, and the set of six double cancellation conditions are the only non-trivial ones when the cancellation axiom of MA is restricted to $k = 3$. The last three of these are double cancellation conditions in the $k = 2$ sense when the two product sets are $A_1 \times A_2$ and A_3, $A_1 \times A_3$ and A_2, and $A_2 \times A_3$ and A_1. A similar condition to the trivial H6C1 could be given for H4, so the two sets of conditions parallel one another.

Conditions suggested for H7 are:

H7C1. (Zeros) All zero sets may exist and ${}^{0}A_i \times A_j \equiv ({}^{0}A_i/A_j \times {}^{A}j) \cup (A_i \times {}^{0}A_j|A_i)$ and ${}^{0}A_i|A_j \equiv {}^{0}A_i|A_k$ for all $i, j, k = 1, 2, 3$ and $i \neq j \neq k$.

H7C2. (Weak order) The relation $\lesssim$ over $A_1 \times A_2 \times A_3$ is a weak order.

H7C3. (Cancellation) For all $a_i, b_i, c_i \in A_i$ which are not effective zeros, where $i = 1, 2, 3$:

i) $(a_1, a_2, a_3) \sim (b_1, b_2, b_3)$ and $(c_1, b_2, b_3) \sim (a_1, a_2, c_3)$ imply $(c_1, c_2, a_3) \sim (b_1, c_2, c_3)$

ii) $(a_1, a_2, a_3) \sim (b_1, b_2, b_3)$ and $(b_1, c_2, b_3) \sim (c_1, a_2, a_3)$ imply $(a_1, c_2, c_3) \sim (c_1, b_2, c_3)$

iii) $(a_1, a_2, a_3) \sim (b_1, b_2, b_3)$ and $(b_1, b_2, c_3) \sim (a_1, c_2, a_3)$ imply $(c_1, a_2, c_3) \sim (c_1, c_2, b_3)$

iv) $(a_1, a_2, a_3) \sim (b_1, b_2, b_3)$ and $(b_1, b_2, c_3) \sim (c_1, c_2, a_3)$ imply $(a_1, a_2, c_3) \sim (c_1, c_2, b_3)$

v) $(a_1, a_2, a_3) \sim (b_1, b_2, b_3)$ and $(b_1, c_2, b_3) \sim (c_1, a_2, c_3)$ imply $(a_1, c_2, a_3) \sim (c_1, b_2, c_3)$

vi) $(a_1, a_2, a_3) \sim (b_1, b_2, b_3)$ and $(c_1, b_2, b_3) \sim (a_1, c_2, c_3)$ imply $(c_1, a_2, a_3) \sim (b_1, c_2, c_3)$

H7C4. (Sign-dependence) the sets $A_i \times A_j$ are sign dependent on A_k for all $i, j, k = 1, 2, 3$ and $i \neq j \neq k$.

This set is similar to (in fact based on) the set given for H5. The sets of conditions necessary for H8 and H9 are constructed to be similar to all the previous sets in that there is a zero, weak order, cancellation and independence/sign dependence condition. The weak order condition is the only one necessary for all hypotheses H6 - H9.

The suggested QFA conditions for H8 are as follows:

H8C1 (Zeros) The sets $O_{A1} x_{A2}$ and ${}^{O}A_3 \equiv {}^{O}A_3|A_1 \equiv {}^{O}A_3|A_2$ all may exist but no others.

H8C2. (Weak Order) The relation $\leqslant$ over $A_1 \times A_2 \times A_3$ is a weak order.

H8C3. (Cancellation). For all d_1, a_1, b_1, c_1, $\in A_1$, a_2, d_2, b_2, c_2, $\in A_2$ and d_3, a_3, b_3, c_3, $\in A_3$,

i) if $(a_1, a_2, a_3) \leqslant (b_1, b_2, b_3)$ and $(c_1, c_2, a_3) \leqslant (d_1, d_2, b_3)$ and $(d_1, b_2, b_3) \leqslant (c_1, a_2, a_3)$ then $(a_1, c_2, a_3) \leqslant (b_1, d_2, b_3)$.

ii) $(a_1, a_2, a_3) \leqslant (b_1, b_2, a_3)$ and $(b_1, c_2, a_3) \leqslant (c_1, a_2, a_3)$ imply $(a_1, c_2, a_3) \leqslant (c_1, b_2, a_3)$.

H8C4. (Independence/sign-dependence). The sets A_1 and A_2 are independent of each other and $A_1 \times A_2$ and A_3 are sign dependent of each other.

The following conditions are necessary for H9:

H9C1 (Zeros) The zero sets ${}^{O}A_1|A_2$, and ${}^{O}A_2|A_1$ may exist but no others.

H9C2 (Weak Order) The relation $\leqslant$ over $A_1 \times A_2 \times A_3$ is a ~~a~~weak order.

H9C3 (Cancellation) For all a_1, b_1, $c_1 \in A_1$, a_2, b_2, $c_2 \in A_2$ and a_3, b_3, c_3, $\in A_3$

if $(a_1, a_2, a_3) \leqslant (b_1, b_2, b_3)$ and $(b_1, b_2, c_3) \leqslant (c_1, c_2, a_3)$ then $(a_1, a_2, c_3) \leqslant (c_1, c_2, b_3)$

H9C4 (Independence/sign dependence) The sets A_1 and A_2 are sign dependent on each other and the sets $A_1 \times A_2$ and A_3 are independent of each other.

Krantz and Tversky suggest other cancellation conditions for H7 - H9 which require the notion of homogenous subsets. Let A_i^+ be the set of elements in A_i that have positive function values and A_i^- that set which have negative function values, $i = 1, 2, 3$. A homogenous subset of $A_1 \times A_2 \times A_3$, or $A_1 \times A_2$ in the case of $\mathcal{A}_2$ is a subset which restricts some of the sub-elements to one sign only e.g. $A_1^+ \times A_2 \times A_3^-$. With this rather loose definition, additional cancellation conditions could be given for H5 and H7-H9. However, as only those for H5 and H9 will be used in this study only those will be given here.

H5C3' Within the homogeneous sub-sets $A_1^+ \times A_2^+$, $A_1^+ \times A_2^-$ $A_1^- \times A_2^+$ and $A_1^- \times A_2^-$ if $(a_1, a_2) \leqslant (b_1, b_2)$ and $(b_1, c_2) \leqslant (c_1, a_2)$ then $(a_1, c_2) \leqslant (c_1, b_2)$

H9C3' Within the homogeneous subsets $A_1^+ \times A_2^+ \times A_3$, $A_1^+ \times A^-_2 \times A_3$, $A_1^- \times A_2^+ \times A_3$ and $A_1^- \times A_2^- \times A_3$ if $(a_1, a_2, a_3) \leqslant (b_1, b_2, a_3)$ and $(b_1, c_2, a_3) \leqslant (c_1, a_2, a_3)$ then $(a_1, c_2, a_3) \leqslant (c_1, b_2, a_3)$

Similar conditions for H7 and H8 complete the sets of useful consequences of these representation hypotheses. Although all the cancellation conditions for H7 - H9 are implied by H6C3 no two sets of conditions can both be satisfied by the same (non-trivial) data.

Now, the conditions apart from the zero ones are all testable sentences. Before the sign-dependence conditions can be tested, however an empirical method of determining when the relations S or -S holds between the sub-elements of $A_1 \times A_2 \times A_3$ is required. This is done by taking a sub-set of observations.

Suppose one required to test the sign dependence of A_1 on A_2. A particular pair, a_1, b_1 from A_1 are selected such that $(a_1, a_2, a_3) \prec (b_1, a_2, a_3)$ for some a_2, a_3. Also, a particular b_3 from A_3 is chosen. Then, for all $b_2 \in A_2$ one determines if

i) $(a_1, b_2, b_3) \prec (b_1, b_2, b_3)$

ii) $(a_1, b_2, b_3) \sim (b_1, b_2, b_3)$

iii) $(b_1, b_2, b_3) \prec (a_1, b_2, b_3)$

Then for all b_2, $c_2 \in A_2$ for which i) holds b_2 S c_2. Because of the conditions under which a_1, b_1 were chosen, all b_2 for which ii) holds must be effective zeros. If i) holds for b_2 and iii) for c_2 then b_2 - S c_2 and if iii) holds for b_2 and c_2 then b_2 S c_2. Thus, for all none zero pairs of elements in A_2 one has determined empirically whether S or -S holds. The sign dependence of A_1 or A_2 can then be tested. A similar procedure can be used to test the sign dependence of any set, or product set on any other.

Sets of consequences of the simple polynomials on $\mathcal{A}_3$, suitable for QFA have been presented. These complement those of the previous sections so that results to test a wide range of representations have been amassed. Before considering how to apply such results sets for 2 more representations can be set out.

Inexact Functional Relations.

The representation hypotheses H4 - H9 embody exact subjective scales and postulate the existence of exact functional relationships. It may be though, that data satisfy most of the principles implied by the exact models but not all. In such a case it may be possible to describe peoples' behaviour by a set of principles which follow from an inexact representation hypothesis. If such hypotheses can be stated economically they may be useful. In this section inexact versions of H4 and H5 are stated and results for their QFA proposed. They could be called representations in semi-orders since they require the semi-order condition to hold. However, to avoid cumbersome terminology they will be called the "inexact" representations, even though other inexact representations exist.

It is desirable to ensure that the zero condition, H5C1 and the sign-dependence condition, H5C5 are consequences of the inexact multiplicative representation. This is because in certain cases they are rather trivial 'principles' of behaviour which can be assumed to hold. To ensure H5C5 holds it is necessary to define signed classes, $A_1 \times A_2^+$ and $A_1 \times A_2^-$ from the zero class, ${}^0A_1 \times A_2$ as follows:

$(a,b) \in A_1 \times A_1^- \Leftrightarrow (a, b) \prec 0$

and $(a,b) \in A_1 \times A_2^+ \Leftrightarrow 0 \prec (a,b)$, where $0 \in {}^0A_1 \times A_2$.

The following inexact additive and multiplicative hypotheses are proposed.

H4'. <u>The INEXACT ADDITIVE conjoint representation hypothesis</u> for $\mathcal{A}_2$, IACRH states that there exist real-valued functions, u_1 on A_1 and u_2 on A_2 such that, for all $a_1, b_1 \in A_1$ and $a_2, b_2 \in A_2$

i) $(a_1, a_2) \prec (b_1, b_2) \Leftrightarrow u_1(a_1) + u_2(a_2) + l < u_1(b_1) + u_2(b_2)$

ii) $(a_1, a_2) \sim (b_1, b_2) \Leftrightarrow (u_1(a_1) + u_2(a_2)) - (u_1(b_1) + u_2(b_2)) \leqslant l$

H5: <u>The INEXACT MULTIPLICATIVE conjoint representation hypothesis</u>, IMCRH states that there exist real-valued functions, u_1 on A_1 and u_2 on A_2 such that for all $a_1, b_1 \in A_1$ and $a_2, b_2 \in A_2$,

a) If $(a_1, a_2) \in A_1 \times A_2^-$ and (b_1, b_2) does not, or
if $(a_1, a_2) \in 0_{A_1} \times A_2$ and $(b_1, b_2) \in A_1 \times A_2^+$
then $(a_1, a_2) < (b_1, b_2)$

b) If $(a_1, a_2), (b_1, b_2) \in A_1 \times A_2^-$ then

i) $(a_1, a_2) < (b_1, b_2) \Leftrightarrow u_1(a_1).\ u_2(a_2).\ E < u_1(b_1).\ u_2(b_2)$

ii) $(a_1, a_2) \sim (b_1, b_2) \Leftrightarrow$

$$E \leqslant \frac{u_1(a_1)u_2(a_2)}{u_1(b_1)u_2(b_2)} \leqslant 1/E$$

c) If $(a_1, a_2), (b_1, b_2) \in A_1 \times A_2^+$ then

i) $(a_1, a_2) < (b_1, b_2) \Leftrightarrow u_1(a_1).u_2(a_2) < E.\ u_1(b_1).u_2(b_2)$

ii) $(a_1, a_2) \sim (b_1, b_2) \Leftrightarrow$

$$E \leqslant \frac{u_1(a_1).u_2(a_2)}{u_1(b_1).u_2(b_2)} \leqslant 1/E$$

where $0 < E < 1$

It is suggested that the following set of consequences of H4' are suitable for its QFA.

H4'C1. The relation, $\leqslant$ is a semi-order over $A_1 \times A_2$

H4'C2. If $(a_1, a_2) < (b_1, b_2)$ and $(b_1, c_2) < (c_1, a_2)$
then $(a_1, c_2) < (c_1, b_2)$

H4'C3. The independence condition H4C3 holds.

H4'C4. The zero condition H4C4 holds.

The cancellation condition, H4'C2 is a weak version of H4C2 which applies only to the strict preference relation, $<$.

For the QFA of H5' the following conditions may be suitable for QFA:

H5'C1. The sign condition, H5C1 holds.

H5'C2. The relation, $\leqslant$ is a semi-order over $A_1 \times A_2$, within signed classes.

H5'C3. If $(a_1, a_2) < (b_1, b_2)$ and $(b_1, c_2) < (c_1, a_2)$ then $(a_1, c_2) < (c_1, b_2)$ for elements within the same signed class.

H5'C4. The sign dependence condition, H5C4 holds.

The proof that these conditions are necessary for the inexact representations is trivial. The usefulness of the inexact models is that transitivity of indifference and cancellation of indifference, two of the most stringent requirements of the exact models are not necessary. Corresponding inexact representations for $\mathcal{A}_3$, except for the additive case are not economically stated and are not expected to be particularly useful.

<u>The Qualitative Functional Analysis of Experiments in Decision Making</u>

The general QFA results of the previous sections can be applied to any empirical situation that can be represented as a relational system like $\mathcal{A} = \langle A_1, \lesssim \rangle$ or $\mathcal{A}_n = \langle A_1 \times \ldots \times A_n, \lesssim \rangle$ as long as it is meaningful to propose some functional relationship. It has already been indicated that the information integration models of the introduction are, in simple situations, special cases of representation hypotheses, and that some kinds of data on decision making can be considered as relational systems. Thus it seems that applications of QFA to information integration models of decision making can be attempted. Let us examine the problems of such an application of QFA with special regard to the SEU model.

Earlier, the general SEU model was defined as a general representation hypothesis for the case where alternatives were presented in simple probability and payoff terms. The simplest case is the riskless one where alternatives are univariate. The SEU model for this case corresponds to the weak order representation hypothesis, H3 which is renamed the ordinal utility hypothesis. From the section on ordinal hypotheses it is clear that to test the ordinal utility hypothesis for some set A it is only necessary to test the axioms of M3. Obviously, if the axioms are not met SEU is not a suitable model and some alternative must be considered. The representations H1 and H2 may suffice as inexact ordinal utility models, and could be tested by examination of the axioms of M1 and M2. In some riskless choice situations such models may be more realistic.

Suppose, in a riskless choice situation the alternatives can be characterized by elements of the product set, $A_1 \times \ldots A_k$ e.g. for a multi-variate motivating variable where an alternative has values on k dimension. An additive utility model could be considered like HA and this could be tested via some set of qualitative properties. If $k = 2$ or $k = 3$, suitable sets are known from earlier sections that are necessary for H4 and H6. In fact, the set for H4 were introduced by Adams and Fagot to test the additive utility model when $K = 2$. If the additive utility model failed, an inexact additive model could perhaps be examined.

The remaining situations to be considered involve choices among certain gambles, notably simple and duplex gambles as previously defined. The SEU model for choices among simple and duplex gambles was stated in the introduction. That for simple gambles can be regarded as a special case of H5 and for QFA one would examine the relevant set of consequences of H5. Tversky (1967a) showed that when simple gambles involving amounts to win only are considered SEU is a special case of H4. He tested it by examining the qualitative conditions associated with H4. By either method the alternative inexact representations could be examined if the predictions about indifference were violated.

Turning to duplex gambles, suppose any one of the parameters SW, SL, PW or PL is held constant. The SEU model would then be a special case of H9 and could be examined via QFA using the conditions suggested earlier. This provides a way of examining choices among the four parameter duplex gambles - each parameter is held constant in turn. For this case no economically stated inexact model is known.

Two other types of gambles to which SEU can be applied are i) simple gambles where the outcome is from a two-product set and ii) single wheel gambles where the outcome when the pointer lands in the lower sector is not necessarily zero. Krantz and Tversky pointed out that for i) SEU is a special case of H8. For ii), if $p = 0.5$ but the outcomes are allowed to vary, SEU is a special case of H4.

Thus it is possible to examine decision making by QFA in situations where the SEU or other exact information integration models are special cases of one of the hypotheses H3 - H9. In addition, QFA results for inexact representations corresponding to some of H3 - H9 are known. The only hypothesis not mentioned above is the multiplicative hypothesis, H7. If a set of gambles was constructed for which SEU was a special case of H7 the probability part would be rather artificially compounded, so such cases will be ignored. Not all the above hypotheses will be examined experimentally here, they were presented to indicate the range of situations which can be included.

The obvious way to test an information integration model by QFA is to take pairs of elements from some set of alternatives, say a, b $\in P_1 \times X_1 \times \ldots P_k \times X_k$, obtain evidence as to whether $a < b$, $a \sim b$ or $b < a$ and apply QFA to the preference relationships favoured. Two main ways of obtaining such evidence will be considered: direct observation of choices and statements of preference or indifference, both obtained when DMs are confronted with pairs of alternatives. The latter is the more 'natural' kind of data for QFA as it gives direct evidence about whether $a < b$, $a \sim b$ or $b < a$. This is not true of choice data. Suppose a is chosen from (a,b) in an actual choice situation. If it is a forced choice and even if the possibility of errors is excluded the observation is not unequivocal. Either DM chooses it because he prefers it or he chooses it at random when he is indifferent between the alternatives. Thus, choice data has the disadvantage that pairs of alternatives must be presented to people a number of times so that preference and indifference can be differentiated. It has the advantage, though, that real consequences can be made to follow a DM's choices. This is not possible in any obvious way for statements data. Both kinds of data then have advantages as well as disadvantages. The more realistic choice data is regarded as the more valid, but statements data will have good uses in preliminary studies because of its directness.

Statements of Preference

It is worthwhile to consider statements of preference in more detail. The first thing to note is that if a set of "binary preference statements" are obtained, some of them may be made in error. The "true" preference may be different from the observed preference. Therefore, if a small number of violations of qualitative consequences of an information integration model are observed this does not mean that the model must be rejected. As Krantz and Tversky (1971) note, a way of dealing with fallible data in QFA must be found. One way to do this is to superimpose an "error model" (for a certain kind of data) on an information integration model. A very simple error model for statements of preference is as follows:

Defn. The SIMPLE ERROR model for binary preference statements states that when a pair of alternatives are presented to DM the probability that he makes a statement of preference which corresponds to his true preference is $1 - \alpha$, where $0 \leqslant \alpha \leqslant 1$. The parameter, α is specific to the DM but constant for all pairs of alternatives.

Obviously more complex error models could be considered. For instance it may be that the probability of an error when the true state is strict preference is less than the probability of an error when the true state is indifference. The least restrictive error model would be that for which the probability of an error for any preference statement, i was α_i, and $0 \leqslant \alpha_i \leqslant 1$ for all observations.

Tests of Hypotheses for Statements Data.

Suppose a set of n preference statements are obtained for pairs of alternatives, one statement per pair. First one would test the observed statements to see if the consequences of some information integration model of interest were violated. If violations are found one should search for a set of binary preferences which do not violate the conditions but are similar to the observed set. Let the number of observations which are different from the observed set be k. If a good search procedure is available one should be able to find a non-violating set with minimum k value. The hypothesis that DM's true preferences are those of the non-violating set and he made errors according to the simple error model (H_1) can be tested against the hypothesis that his true preferences are those of the non-violating set but he made errors according to the general error model (Ho) which says that any probability of an error lies between 0 and 1. The hypothesis H_1, that $\alpha_1, = \alpha_2 = \ldots = \alpha_n = \alpha$ is a special case of Ho, and can be tested against it by the likelihood ratio (LR) test. The point of making the test between the models is obvious. If the test leads one to reject Ho in favour of H_1 then one can say that the information integration model holds and that errors can be described by a model with one parameter. If H_1 is rejected then one can try models that are less restricting than the simple error model but less general than the general error model, such as that suggested earlier. If all "reasonable" models are rejected one must accept Ho. This says that the information integration model holds but errors are described by a model with n parameters. Since such a model is useless it is better to conclude that the information integration hypothesis can be rejected.

The maximum likelihood (ML) estimators of α_i under Ho are $\alpha_i = 0$ when the observed preference and corresponding "true" preference are the same and $\alpha_i = 1$ when they are different. Thus the likelihood function (LF) of Ho, L(Ho) = 1. The ML estimator of α under H_1 is $\hat{\alpha} = k/n$ where k of the n observations are different from the true ones. The maximum LF of H_1 is

$$ML\ (H_1) = \frac{n!}{k!\ (n-k)!} \left(\frac{k}{n}\right)^{k} (1 - k/n)^{n-k}.$$

By the likelihood ratio test, the function $-2 \log ML\ (H_1)$ is distributed asymptotically as chi square with n - 1 degrees of freedom (since ML (Ho) = 1).

The other error model mentioned, where the probability of error is different for indifference is tested against Ho in the same way. This time the chi-square statistic would be distributed with n - 2 degrees of freedom.

Using such error models it is possible to evaluate when an information integration model has reasonable predictive power and when it should be discarded.

The above procedure only gives a test of a class of true preference statements if $ML(H_1)$ is maximised over the whole set. A search procedure to find the maximum of $ML(H_1)$ is presented in appendix 2. It is also relevant to binary choice data and it will be discussed further in this connection.

Binary Choice Data.

As indicated earlier it is necessary to obtain a number of choices for each pair of alternatives presented to DMs in order to distinguish between preference for an alternative and the random choosing of it. The design of such an experiment, using an individual DM should be described so that the assumptions made in testing by QFA can be clarified. It is the n-replicate, complete pair comparison design.

Let there be m elements in the set of alternatives, A, where elements are denoted $a, b, c \in A$. The DM makes a sequence of choices from pairs in A. On any trial he must choose one or the other, no indifference being allowed. All $m(m-1)/2$ pairs are presented n times. Both alternatives are presented simultaneously and the probability that either is on the right is $p = 0.5$. Thus it is an n-replicate design which is balanced for spacial and temporal effects. The basic assumptions made in the statistical tests to be discussed are that all the observations are independent and the n observations on any pair are identically distributed bernoulli trials. This hypothesis will be called the bernoulli model. The $n\ m(m-1)/2$ trials are presented in random order to the subjects except that no pair is presented if it appeared on the previous ℓ trials. This is so that memory factors are minimized. Assuming the bernoulli model may be reasonable when ℓ is large. The binomial parameter $p(a,b)$ for the pair $a,b, \in A$ is the binary preference probability. It denotes the probability that DM will chose a any time that (a,b) are presented.

Now, the preference and indifference relations of the representation hypotheses which are to be examined must be related in some way to the binary preference probabilities.

Preference and indifference can be defined in terms of binary preference probabilities in a variety of ways. It seems ridiculous to define indifference in any way other than the following:

$$a \sim b \iff p(a,b) = 0.5$$

If DM is indifferent between a and b he chooses at random and if he chooses at random $p(a,b) = 0.5$, If he shows a tendency to choose one over the other, i.e. $p(a,b) \neq 0.5$ then he is not indifferent between them. If this is accepted the different ways of relating true preference to preference probabilities, differ only in the way strict preference relates. In the literature, indifference has been considered as adequately represented by a probability range (e.g. Luce 1959). But here such models will not be discussed.

An errorless preference model of binary choice has commonly been called the algebraic model. The algebraic preference model states that $a < b \iff p(a,b) = 0$ (and of course, $b < a \iff p(a,b) = 1$). However, any model which predicts that no erroneous choices will occur will not survive, and as with statements data a way to incorporate errors must be found.

Two simple probabilistic models are suggested.

Defn. PROBABILISTIC MODEL I of PREFERENCE FOR BINARY CHOICE DATA states that $b < a \iff p(a,b) \geqslant \alpha$ were $0.5 \leqslant \alpha \leqslant 1$ is a parameter of the DM constant over all pairs of alternatives (obviously $a < b \iff p(a,b) \leqslant 1 - \alpha$).

Defn. PROBABILISTIC PREFERENCE MODEL 2 of PREFERENCE FOR BINARY CHOICE DATA states that $b < a \iff p(a,b) = \alpha$ where $0.5 < \alpha < 1$

In the latter case the probability of an error is constant. In both cases it would be better to estimate α from the data, but with probabilistic model I this generally leads to accepting $\alpha = 0.5$. Tversky (1969) has adopted this model with $\alpha = 0.5$ in order to test (among other things) the transitivity part of the weak order condition. Now, the interpretation of information integration models in terms of relational systems postulates that the three states $a < b$, $a \sim b$, $b < a$ are qualitatively different. If one uses $\alpha = 0.5$ though, one has the situation $p(a,b) = 0.51 \Leftrightarrow b < a$, $p(a,b) = 0.50 \Leftrightarrow a \sim b$ and $p(a,b) = 0.49 \Leftrightarrow a < b$. This is not in the spirit of the models at all since the three situations would not be discernable in DM's behaviour. Also, two situations which one could distinguish between, say $p(a,b) = 1$ and $p(a,b) = 0.6$ would, by the model, be regarded as equivalent. If one permits the binary preference probabilities to be unrestricted in the range $0 \leqslant p(a,b) \leqslant 1$ one would be far better off using models which make point predictions of them. Such models are considered in the next chapter. If the idea of qualitatively different strict preference and indifference states existing is to be of use a model with α value near to 1 so that $a < b \Leftrightarrow p(a,b)$ is high and $b < a \Leftrightarrow p(a,b)$ is low would make more sense. For this reason an α value for model I has been chosen a priori by what is hoped is a reasonable criterion. The criterion, and the α values chosen by it are given in appendix I. It is based on such considerations as 1) it would be unsatisfactory if the maximum allowable probability of error, $(1 - \alpha)$ was high 11)$(1 - \alpha)$ should be low enough so that the model is a reasonable initial hypothesis. Probabilistic model 2 is more restrictive than model I and no problems ensue from estimating α from the data in certain circumstances. Unfortunately there are other cases where it is much more difficult to test than model I. For this reason it will not be considered further. The idea of three distinct states corresponding to three point preference probabilities seems more in keeping with the spirit of the representation hypotheses, however.

Model I is a relatively weak model, It allows the probability of an error to be different for different pairs of alternatives. Preference probabilities are merely restricted to a certain range. Statistical tests based on this model will now be described.

Tests of Hypotheses for Binary Choice Data, Assuming Probabilistic Model I.

If preference is clearly defined in terms of binary preference probabilities it is possible to test, by likelihood ratio tests whether the data from an n-replicate pair comparison experiment is consistent with qualitative consequences of information integration models, assuming probabilistic model I. Tversky (1969) applied such tests in examining transitivity of preference. He assumed probabilistic model I and an α of 0.5. When $\alpha \neq 0.5$ and the set of pairs presented is large the tests are more difficult to apply. A solution will now be presented for the general case.

All models have consequences which restrict the preference probabilities in some way. They are tested against the unrestricted, maximum likelihood bernoulli model. The maximum likelihood estimate of the binomial parameter, p(a,b) is k(a,b)/n where k(a,b) is the number of times a was chosen by the subject. The unrestricted maximum likelihood of the data, UML is, therefore:

$$\mathrm{UML} = \sum_{\substack{\text{all pairs} \\ (a,b)\ A \times A}} \frac{n!}{(n-k(a,b))!\,k(a,b)!} \times (k(a,b)/n)^{k(a,b)} \times (1-k(a,b)/n)^{(n-k(a,b))}$$

Now, what is under test is whether the data violate a set of conditions predicted by some information integration model assuming probabilistic model I holds. For the likelihood ratio test the maximum likelihood of the data given these restrictions (the RML) is required.

The search procedure for this maximum is rather involved, and has been relegated to appendix 2. Suppose, anyway that the set of preferences and indifferences given the restricted maximum likelihood can be found. The best restricted estimates of the p(a,b)'s would be as follows:

If $b < a$ then best estimate of p(a,b) is $\begin{cases} k(a,b)/n & \text{if } k(a,b)/n \geq \alpha \\ \alpha & \text{otherwise} \end{cases}$

If $a \sim b$ then best estimate of p(a,b) is 0.5

If $a < b$ then best estimate of p(a,b) is $\begin{cases} k(a,b)/n & \text{if } k(a,b)/n \leq 1-\alpha \\ 1-\alpha & \text{otherwise.} \end{cases}$

Thus, the RML only restricts some of the estimates of the preference probabilities. When the likelihood ratio test is applied the asymptotic distribution of the function - 2 log (RML/UML) is distributed as a chi-square variate but unlike the statements case the appropriate number of degrees of freedom is the actual number of restrictions on the p(a,b)'s. The restricted hypothesis is rejected in favour of the more general one if the above statistic is significant. The test is conditional on what data is observed. Unfortunately, the effects of this conditionality on the distribution of the statistic under the null hypothesis are unknown. They cannot be assumed to be negligible. An even more basic problem than this must be tackled, however: the function RML must be maximized over the set of preference patterns which satisfy the consequences of the information integration model under investigation. A solution to this problem when probabilistic model I is assumed is set out in appendix 2. Unfortunately it does not apply to probabilistic model 2. The procedure involves reformulating the maximization problem as a linear programming one. Such a solution to a related problem was suggested by Decani (1969). The linear programming problems cannot be solved by standard procedures as the variables are bivalent. Such problems are called pseudoboolean linear programming problems by Hammer and Rudeanu (1968) and can be solved by their methods. A useful feature of the solutions is that they apply equally well to incomplete pair comparison data. Unfortunately the algorithms, though known to exist have yet to be converted to suitable computer programmes.

At the moment, therefore the maximization must be carried out by rather ad hoc methods, not the full solution above. Consequently the full properties of the test statistic are unknown because the effect of its conditionality on the data cannot be investigated. Despite these drawbacks the test will be applied in the experimental section as no better alternative is known.

An actual application of QFA to binary choice data is carried out as follows:

i) a set of qualitative properties, consequences of the information integration model under investigation are numerated.

ii) RML is maximized over the set of preference patterns which satisfy the conditions assuming probabilistic model I.

iii) UML is maximized.

iv) the conditions assuming probabilistic model I are tested against the unrestricted bernoulli model by the likelihood ratio test.

The actual procedures used to carry out ii) will be presented in the results sections where this is necessary. This completes the survey of QFA and the discussion of how it will be applied in the present study. Its merits and drawbacks will be considered further after it has been applied.

Chapter 3

The Functional Measurement Analysis of Decision Making.

Functional measurement is used to explore functional relationships quantitatively. Recently, Anderson (1970) has elucidated the usefulness of one approach in the study of psychophysical judgments and Anderson and Shanteau (1970) have applied it to risky decision making. Other quantitative methods have been used to explore functional relationships in decision making, including some using a relational system formulation (see Edwards (1954c), 1961), Luce and Suppes (1965)). For these, QFA would provide "goodness of fit" tests of the basic functional relationship model. Quantitative information would then be obtained by a measurement procedure within the same framework if the model appeared to fit.

However, for the present study it was decided to consider quantitative models based on a different set of assumptions, utilizing the approach of Bock and Jones (1967). Bock and Jones (1967) have developed methods for estimating interval scales of objects from preference frequencies obtained in n-replicate pair comparison experiments. The model underlying this is Thurstone's case V model for affective values. They suggest related estimation and goodness of fit tests for general multi-factor models. These can be used to explore functional relationships in an analagous way to Anderson's procedures of functional measurement. Anderson's approach cannot be used directly with choice data as it requires responses measured on a continuous scale. In the experiments to be reported, the rationale for examining functional relationships quantitatively is based on Anderson's approach but the actual estimation and goodness of fit methods used are those of Bock and Jones. The synthesis of these two approaches in the present application to static decision making will now be described.

The basic objective is to explore the functional relation between the psychological factors "attractiveness" or "worth" of alternatives facing DM and the subjective value of the various information dimensions of these alternatives. As with the QFA approach, let the alternatives be suitably described as elements $(p_1, x_1, \ldots p_k, x_k)$ of a product set $P_1 \times X_1 \times \ldots \times P_k \times X_k$. The basic assumption of any functional measurement is that the subjective and objective factors involved in the functional relation can be measured on at least an interval scale. The present application would be inappropriate if this assumption was not warranted.

Let us suppose then, that an experiment involving choices between simple gambles, that is, alternatives from $P_1 \times X_1$ is carried out, using three factor levels p_0, p_1, $p_2 \in P_1$ and x_0, x_1, $x_2 \in X_1$. Anderson would obtain a response, r_{ij} for each factorial combination (p_i, x_j), $(i, j = 0, 1, 2)$. In a results table where the P_i's are the rows and the x_j's the columns, the row means $r_{i.}$ and the column means, $r_{.j}$ are used as the basis for the scaling of the subjective value of the p_i's and x_j's. The actual scaling procedure depends on the functional relationship being examined. One may be considering any of the following information integration models:

i) $r_{ij} = s(p_i) + u(x_j)$ (additive)

ii) $r_{ij} = s(p_i)\ u\ (x_j)$ (SEU)

iii) $r_{ij} = s(p_i) + u(x_j) + su(p_i, x_j)$ (general linear)

The scales are inappropriate if the information integration model is so the fit of the model must be tested. Anderson does this by applying analysis of variance to the r_{ij}'s to test the significance of the main effects and interactions. Model i) predicts that the row x column interaction is not significant and model ii) predicts that the only significant component of the interaction is the linear x linear. Model iii) is the general case in which all the interactions are permitted. In this last case, scaling the subjective values is only meaningful in a restricted sense. The tests of these models, then consist of testing the significance of the appropriate components of the interactions by analysis of variance.

The general linear model is not really tested, but its parameters are estimated and it is accepted by default if the other models are rejected.

This simple approach could be readily extended to information integration models where the alternatives are from larger product sets. For instance, where they are duplex gambles from $P_1 \times X_1 \times P_2 \times X_2$ or $P_1 \times X_1 \times X_2$ (where $p_2 \in P_2$ is held constant). The functional relationships of interest would again be the additive, SEU or general linear models. Anderson and Shanteau (1970) in fact applied the method to rating data for simple and duplex gambles.

To apply the method to model testing from choice data, however requires further techniques since continuous measures of the subjective values of the alternatives are not directly observed. Bock and Jones (1967) have developed suitable techniques, by which the subjective values of the alternatives can be scaled under specific assumptions about the observation error. Different "specific assumptions" can be made, one such set corresponding to Thurstone's case V model for affective values (Thurstone (1959)).

Generally, though the basic assumption is that at the time of choice the subjective value, v_i of the alternative a_i is decomposable into a constant part, u_i and a variable part, e_i. For the moment, let us denote alternatives from $P_1 \times X_1$ above by $a_i, a_j \in P_1 \times X_1$. Two interpretations are possible: either the constant part represents the 'true' subjective value and the variable part measurement error, or subjective value is something which fluctuates by amounts e_i around the mean value u_i. At any rate, $v_i = u_i + e_i$ where e_i is a random variable with distribution $f(e_i)$ and $E(e_i) = 0$, Var $(e_i) = \sigma_i^2$. Alternative models can be considered where the e_i are identically distributed with common variance σ_1^2 (all $a_i \in P_1 \times X_1$). In these models, a difference process can be considered, such that the random variable $v_{ij} = v_i - v_j = (u_i - u_j) + (e_i - e_j) = u_{ij} + e_{ij}$ where e_{ij} is distributed as $f(e_{ij})$ with $E(e_{ij}) = 0$ and Var $(e_{ij}) = 2\sigma_1^2 = \sigma^2$. (This is not necessarily true for all f but it is for the ones we consider). This difference process is assumed to operate when DM is faced with a binary choice situation. That is, where he must choose from a pair of alternatives (a_i, a_j). Obviously, the u_{ij}'s are the differences in mean subjective value of pairs (a_i, a_j). Fluctuations in v_{ij}'s will be referred to from now on as random sampling error. All random sampling error in the difference processes considered have common variance.

The above model is the basis of the procedure for estimating the subjective values of the alternatives. The estimation uses the binary preference frequencies from an n-replicate pair comparison experiment.

Now, if the pair (a_i, a_j) is presented to DM then the binary preference probability i.e. the probability that he will choose a_i is

$$P_{ij} = P(v_{ij} > 0) = \int_{-\frac{u_{ij}}{\sigma}}^{\infty} f(x)\,dx = H\left(-\frac{u_{ij}}{\sigma}\right)$$

Thus the preference probabilities P_{ij} are functions of the unknown parameters of the model i.e. the u_{ij}'s and σ. Estimates of the u_{ij}'s give a scale of subjective value of the alternatives. This scale is unique up to an arbitary origin and unit. The unit can be set by letting $\sigma = 1$ and the origin by letting one of the alternatives have value 0, or by letting the subjective values sum to zero. Some models for which Bock and Jones general estimation procedures can be applied are

i) normally distributed error when $H(u_{ij}) = \frac{1}{2\pi} \int_{-u_{ij}}^{\infty} \exp\left(\frac{1}{2} x^2\right) dx$

$$= \phi(u_{ij})$$

ii) log - normally distributed error when

$$H(u_{ij}) = \frac{1}{2\pi} \int_{-\log u_{ij}}^{\infty} \exp\left(\frac{1}{2} x^2\right) dx$$

Table 3.1

Comparison of Probability of Success as Given by Four Stimulus Response Curves.

u_{ij}	Logistic	Normal	Angular	Linear
0	0.500	0.500	0.500	0.500
0.5	0.622	0.619	0.615	0.608
1	0.731	0.728	0.724	0.716
1.5	0.818	0.818	0.821	0.825
2	0.881	0.887	0.900	0.833
2.5	0.924	0.935	0.958	1.000
3	0.935	0.965	0.992	1.000
3.5	0.971	0.983	1.000	1.000
4	0.982	0.992	1.000	1.000
4.5	0.989	0.997	1.000	1.000
5	0.993	0.999	1.000	1.000

From Cox (1970)

iii) logistically distributed error when

$$H(u_{ij}) = \tfrac{1}{4} \int_{-u_{ij}}^{\infty} \operatorname{sech}^2 \left(\tfrac{x}{2}\right) dx = L(u_{ij})$$

iv) angular distributed error when

$$H(u_{ij}) = \tfrac{1}{2} \int_{-u_{ij}}^{\frac{\pi}{2}} \sin\left(-x + \tfrac{\pi}{2}\right) dx, \quad -\tfrac{\pi}{2} \leqslant x \leqslant \tfrac{\pi}{2}$$

$$= \varphi(u_{ij})$$

In these cases it is clear that u_{ij} is a monotone transform of P_{ij} and vice versa. The inverse transforms are denoted by: $u_{ij} = H^{-1}(p_{ij})$. Cox (1970) has compared the mathematical form of the inverses $\emptyset^{-1}$, L^{-1} and φ^{-1} with the inverse linear transformation of P_{ij}. One can see from table 3.1 that they are all about the same over the range $0.1 \leqslant P_{ij} \leqslant 0.9$ (remembering they are symmetrical about $u_{ij} = 0$). Outside this range they approach the limit at varying rates. The main difference between them is that the inverse linear and angular functions reach the limit at finite points while the others do so at infinity. The inverse logistic and normal curves require a relatively complex variation of the basic estimation procedures but their continuity over an infinite range far outweighs this problem. Either might form a suitable basis for functional measurement and since they agree so closely over the whole range it does not seem necessary to choose between them. Arbitarily then, solutions (due to Bock and Jones) for the normal model will be described. Parallel solutions for the logistic model exist and appear in the references.

The results, and their underlying assumptions are stated in order to avoid continual cross references to their source. In chapters 5 - 7 they are used extensively. Proofs are not given as these have been set out in detail in the primary sources, particularly Bock and Jones (1967).

Let us revert to the former notation for denoting elements $(P_i, X_j) \in P_1 \times X_1$ and consider the particular case where $P_1 \times X_1$ is a 3 x 3 product set.

For this case an n-replicate completely balanced pair comparison experiment can be carried out. The complete, n - replicate pair comparison experiment with m alternatives was described in chapter 2. The preference frequencies can be denoted S_{ijkl}/n. That is, S_{ijkl} is the number of times (P_i, X_j) was chosen over (P_k, X_l). As with chapter 2 it can be assumed that each choice for a given pair is an independently and identically distributed bernoulli variate. Thus the observed proportions, $\hat{P}_{ijkl} = S_{ijkl}/n$ are statistics distributed according to the binomical distribution, with mean P_{ijkl}, say and variance, $\sigma^2/n = P_{ijkl}(1 - P_{ijkl})/n$.

Now, it is assumed that subjective values are related to the observed proportions via the inverse normal function and the difference process:

$$Y_{ijkl} = \phi^{-1}(P_{ijkl}),$$

where $Y_{ijkl} = (\alpha_{ij} - \alpha_{kl}) + e_{ijkl}$.

Bock and Jones suggest using the minimum normit chi-squared procedure to estimate the subjective values, α_{ij}. This procedure, based on Urban (1908) uses Rao's lemma which says that if f(t) is any continuous function with continuous first derivatives of a statistic, t of a sample of n observations with mean zero and variance σ^2/n such that t tends to normality in the limit, then

$$u = \sqrt{n}(f(t) - f(0))$$

is distributed normally with zero mean and variance

$$\sigma^2 \left[\frac{\partial f}{\partial t}_0 \right]^2$$ in the limit. This result is applicable

to the present case, where the $\hat{P}_{ijkl}$'s are binomially distributed and therefore tend to normality as $n \to \infty$. The statistics $\hat{P}_{ijkl}$ correspond to t and the inverse normal transforms of them denoted $\hat{Y}_{ijkl}$ correspond to the function, f(t). It is known from applying Rao's lemma as indicated that the variable known as the normit chi-squared function,

$$Q = \sum_{\substack{\text{all pairs} \\ (i,j),(k,l)}} n\, W_{ijkl} \left\{ \hat{Y}_{ijkl} - (\alpha_{ij} - \alpha_{kl}) \right\}^2$$

is distributed as chi-squared in the limit with m(m-1)/2 degrees of freedom.

The weights, $W_{ijkl} = k^2(\alpha_{ij} - \alpha_{kl})/P_{ijkl}(1 - P_{ijkl})$ and $k^2(x)$ is the squared normal ordinate at x. The minimum normit chi-squared estimators of the α_{ij}'s are those values which minimize Q.

These estimators are normally distributed in the limit and are efficient. The W_{ijkl}'s are unknown but fortunately good approximations can be found which cause only a minor deterioration in the estimation. One such is to let $n\, W_{ijkl} = 1$

and minimize Q. This gives provisional values of the α_{ij}'s which can be used to calculate approximate W_{ijkl}'s. To obtain a final solution of course, Q is again minimized using the approximate W_{ijkl}'s. The minimization uses standard matrix algebra, and can be carried out by setting one α_{ij}, say $\alpha_{33} = 0$ This is permissible since the scale of the α's has arbitary origin. Now this gives the subjective scale values of the alternatives (p_i, x_j), let them be $\hat{\alpha}_{ij}$, and the next problem is to test the model. To do this the chi-square is partitioned as follows. The total chi-square

$$SST = \sum n\hat{Y}^2_{ijkl} \cdot W_{ijkl}$$ (the summation being over all observations). The component due to the parameters,

$$SSR = \sum_{\substack{i,j \\ \text{(all pairs)}}} \left\{ \hat{\alpha}_{ij} \times \sum_{\substack{h,l \\ \text{(all pairs)}}} n\, W_{ijkl} \; \hat{Y}_{ijkl} \right\}$$ and

The component due to the error is therefore the difference SSE = SST - SSR. Under the hypothesis that the data fit the model, SSE is distributed as chi-square with $(m - 1)\ (m - 2)/2$ degrees of freedom. Under the hypothesis that $\alpha_{ij} = 0$ for all pairs (i,j) SSR is distributed as chi-square with $n - 1$ degrees of freedom. A significant value of either indicates that the appropriate hypothesis can be rejected.

If the basic model fits then it is worthwhile proceeding to the problem of scaling the subjective values of $x_i \in X_1$, and $p_j \in P_1$ and testing functional relationships between these dimensions and overall subjective value. As with Anderson's method both of these things are closely linked. The scaling requires that one assumes a model and testing the model requires that the scaling is done.

Consider how the SEU model would be tested. The model says that the subjective value of an alternative is composed of the sum of certain parameters, which depend on the levels of the objective values, and a random error. From the estimates of the subjective values of the alternatives found by minimum normit chi-square, the parameters of the SEU model can be similarly estimated. The fit of this functional relationship is tested by partitioning the original estimation chi-square SSR into a component due to the functional relationship model, SSM and one due to departure from it, SSE. If SSE is significant the proposed functional relationship can be rejected.

To test the additive model, the original estimation chi-square is partitioned similarly. To compare the two models, SSR is split up differently again. The "main effects" parameters of the SEU model are all the parameters of the additive model. Thus, the latter is a special case of the former. In all such cases the SSR chi-square is partitioned into a component due to the parameters common to both models, a component due to the parameters specific to the more general model and a remainder. If the component due to the parameters specific to the general model is significant then this model is significantly better than the restricted one. Otherwise the latter is preferred. The procedure is described in detail in Bock and Jones. A great advantage of this approach is that a test of the error distribution being assumed can be made which is separate from the test of the functional relationship.

A further advantage is that as well as looking at functional relationships among subjective dimensions one can examine relationships between the subjective values of the alternatives and the objective values of the information dimensions. As with the purely subjective value cases, estimation procedures are only known for relationships linear in the parameters. However, some quite complex relationships can be expressed in this way. Some simple psychophysical models, special cases of the general additive and SEU models where the subjective values scales are linear functions of the objective ones will be examined by Bock and Jones' methods in chapters 5 - 7. The rationale is analgous to that for the general additive and SEU cases which will also be tested.

The main difference between the cases discussed above and the applications to be described is that incomplete pair comparison designs are used in the latter.

The necessary and sufficient condition that the estimation procedure can be applied to incomplete, n - replicate pair comparison designs is: if the number of alternatives in the set is m then m - 1 pairs are presented n times each such that each alternative is connected to every other one via a sequence of pairs. It is necessary however to have more than m - 1 pairs so that the goodness of fit can be tested. The differences in the procedure for the incomplete cases are simply that 1) n is replaced by zero in the formulae for pairs not present and 2) p, the number of pairs present replaces (n-1)n/2 as the degrees of freedom of SST.

The Validity of Applying Large Sample Procedures When n is Small.

The preceding results are asymptotic ones which hold in the limit as n tends to infinity. However, in no application does n approach infinity so in practice they can only hold to some degree of approximation. In particular, for finite n there is some bias in the estimates of subjective value, which may be large particularly if some of the observed preference frequencies are zero or one. Bock and Jones examined the simplest case where m = 2, i.e. where there is just one pair of alternatives. Suppose Thurstone's case V model holds for the pair (a_i, a_j). Assuming σ is unity, the difference in subjective values,

$Y_{ij} = \phi^{-1}(p_{ij})$. The estimation procedure for m = 2 degenerates into

$\hat{Y}_{ij} = \phi^{-1}(S_{ij}/n)$ where S_{ij}/n is the preference frequency. Now, S_{ij}/n is the minimum variance, unbiased estimate of P_{ij}, but $\hat{Y}_{ij}$ is not an unbiased estimate of Y_{ij} as the latter is a non-linear transformation of P_{ij}. Another source of bias occurs when $S_{ij} = 0$ or n. To avoid the prediction $\hat{Y}_{ij} = \pm\infty$ the rule of transforming S_{ij} to $\frac{1}{2}$ or $n - \frac{1}{2}$ is adopted, at the expense of introducing some bias. Bock and Jones investigated the extent of the bias in the estimator $\hat{Y}_{ij}$ for various values of n. They concluded that in this simplest case if $n\,p_{ij} > 1$ the bias would not be serious. However, in the case of $m > 2$ it is not clear whether this is still true. Suppose m = 4, and n is quite small, say about 10. If it is suspected that $n\,p_{ij} < 1$ for some pair of alternatives, should it be concluded that the minimum normit procedure gives estimates which are also biased? Unfortunately, for $m > 2$ this can probably not be answered simply. No simple rule of thumb corresponding to that for $m = 2$ is known. Therefore, it seems worthwhile to examine the bias in estimates in the case where $m > 2$.

Bock and Jones mention the fact that the SSE and SSR statistics are distributed as chi-square under the null hypotheses only asymptotically. They advise caution in rejecting hypotheses when the observed values border on significance because of this. However, the degree of approximation to chi square for different values of n is not known. Probably, as long as $n\ p_{ij} > 1$ for most pairs it will be adequate. In view of the extensive use of Bock and Jones methods in the experiments it was felt that evidence should be obtained about the distribution of SSE and the bias in estimation. A sampling experiment was carried out to this end, which will now be described.

A Monte Carlo Study of Bock and Jones' Procedures.

Method.

The setting for the simulation was the n-replicate, complete pair comparison experiment fully described in chapter 2. The number of objects in the presentation set, $m = 4$. Using the earlier notation let these objects be denoted a_1, a_2, a_3 and a_4. Many sets of data in this setting were simulated with $n = 7$, 15 and 60.

In a single run of the experiment n responses were simulated for each pair of objects (a_i, a_j). Let the k^{th} response to the pair (a_i, a_j) be $s_{ijk} = \begin{matrix} 1 & \text{if } a_i \text{ is chosen} \\ 0 & \text{if } a_j \text{ is chosen} \end{matrix}$

The data of main interest were the observed response frequencies denoted $p_{ij} = \sum_k s_{ijk}/n$. The responses were generated under the Thurstone case V assumptions given the following subjective values of the objects: $\alpha_1 = 1.0$, $\alpha_2 = 0.5$, $\alpha_3 = 0.0$; $\alpha_4 = -0.5$ where α_i denotes the subjective values of a_i. The only other parameter it was necessary to specify was the standard deviation of the random sampling error, σ. Values $\sigma = 0.3$ to $\sigma = 1.7$ in steps of 0.1 were used. For each combination of $n = 7$, 15 and 60 and $\sigma = 0.3, \ldots, 1.7$, 1000 runs were simulated. It was hoped that this size of experiment would enable a thorough examination of bias in the estimates given by Bock and Jones' procedure and also of the closeness of the distribution of the error chi-square statistic, SSE to the theoretical distribution.

Observations, s_{ijk} depended on the value of a randomly sampled unit normal deviate, e_{ijk} and on the difference process assumed to underly the choices. The difference in subjective value between the objects $(a_i a_j)$ at the time of choice is assumed to be $d_{ijk} = \alpha_i - \alpha_j + \sigma . e_{ijk}$. The observed choice, $s_{ijk} = \begin{cases} 1 \\ 0 \end{cases}$ if $d_{ijk} \gtrless 0$. If $p_{ij} = 0$ or 1 the convention of assuming $p_{ij} = 0.5/n$ or $(n - .5)/n$ was adopted.

The minimum normit chi-square procedure was applied to the set of p_{ij}'s thus obtained. In this application weights of 1/n were used to obtain initial estimates of the differences $(\alpha_i - \alpha_j)$. Approximate weights were calculated assuming these differences to be true. The approximate weights were then used in a second iteration of the procedure to give final estimates of $(\alpha_i - \alpha_4)$, $i = 1,2,3$. Of course, the estimates were made assuming $\sigma = 1$, so they were all rescaled by the true σ value. The error chi-square, SSE was also calculated in each simulation.

For each set of 1000 runs the mean and mean square error of each estimate was calculated. The SSE statistics were recorded in a histogram with lower bound zero, class interval one and upper bound 100.

Computations.

The computer programme to carry out this simulation was written in the Algol language for the Elliott 1900 of the University of Hull. Some details of the numerical methods used should be explained. A standard library procedure issued by Elliotts was used to generate pairs of uniform random deviates in the range 0 - 1. This procedure applied the additive congruential method and had been fully tested by its writer. No further examination of it was considered necessary. The pairs of deviates were transformed to unit normal ones by a transformation due to Box and Muller, recommended by Tocher (1963) pp 33-34. The matrix algebra used in the estimation also used standard library procedures issued by Elliotts. The only way these were tested was by applying the estimation method to one of Bock and Jones' worked examples, pp.129-132 and comparing the results, which were almost identical. The only other numerical approximations used in the estimation were those of the cumulative normal distribution function and its inverse. The approximations used were those suggested by Bock and Jones (appendices B and C) which are known to be accurate to at least 3 decimal points. The experiment was in fact carried out on the faster computer belonging to the University of Leeds. It took about 25 minutes on this machine.

Results.

First let us consider the extent of bias found. Bock and Jones' study of bias in the case of m = 2 found that bias was small compared to the MSE as long as $nP_{ij} > 1$. Also, the MSE approximated the theoretical variance as long as $nP_{ij} > 2$. Now, the present results are best reported by considering the mean estimates and MSEs as functions of n and σ. To enable a comparable discussion to that of Bock and Jones to be carried out the relationship between the values of the P_{ij}'s for different values of σ should be known. These values are shown in table 3.2. Obviously, as σ increases so do the P_{ij}s, the rate of increase being smaller for large σ.

Figures 3.1, 3.2 and 3.3, show graphs of the mean estimated $(\alpha_i - \alpha_4)$ where i = 1,2,3 for n = 7, 15 and 60. On each of the graphs the general trend is for the difference between true and estimated values (the bias) to decrease as σ increases. The bias is always smaller for larger n, when σ is held constant. These features of the curves indicate that it may be possible to describe extent of bias in terms of the values of nP_{ij}.

Before doing so let us consider the relation between the theoretical variance and observed mean square error of the estimators, since bias in estimation must be considered in relation to these. From Bock and Jones the asymptotic variance of the estimators are known. These values, for different n can be compared to the MSE's of the estimators as functions of σ. Graphs of these functions are shown for n = 7 and n = 15 in figures 3.4, 3.5 and 3.6 and for n = 60 in figure 3.7. It can be seen that in each case the theoretical variance decreases as σ (and therefore each P_{ij}) increases, though this is not so for MSE. The MSE is always very small when σ is small, which must be mainly due to the high probability of observing zero or one preference frequencies in this case. In the case of n = 7 the curve for the MSE of $(\alpha_3 - \alpha_4)$ is almost the mirror image of that of the theoretical variance. This effect is not so pronounced for the differences $(\alpha_1 - \alpha_4)$ and $(\alpha_2 - \alpha_4)$ when n = 7 but here also MSE and theoretical variance bear little relation to each other. When n = 15 the 2 curves are only similar for $(\alpha_2 - \alpha_4)$. MSE and theoretical variance are generally quite close when n = 60 but there is some divergence when σ is small or large. Thus, a general rule specifying when MSE and theoretical variance are close cannot be stated in terms of the n P_{ij} values, as it was in Bock and Jones' experiment.

Since MSE and theoretical variance are not generally close bias should be considered relative to the actual sampling variability (the MSE) rather than to the theoretical variability. Inspection of figures 3.1 - 3.7, in conjunction with table 3.2 led to the following conclusion: bias is small compared to MSE as long as $nP_{ij} > 1$ for 5 of the 6 observed preference frequencies. Unfortunately this neat summary of the extent of bias cannot be assumed to hold in other cases. However it is reasonable to attempt some generalization based on our results. A rule about the likely extent of bias given the proportion of zero or one preference frequencies observed would be particularly useful. From data obtained one could then decide how much faith to place in the scale estimated. With this end in view one could take an observed preference frequency of zero or one as reasonable evidence that $P_{ij} < 1/n$. In the present case, one would suspect that if more than one preference frequency was zero or one, large bias relative to the sampling variability would be likely. Hopefully, though bias would not be too great as long as there were no more than 2 observed zero or one preference frequencies. In general, one could assume the scale estimated was reasonably unbiased as long as 75% of observed preference frequencies were not zero or one. It is felt that this rule could be adopted in the general case until further evidence on this matter is available.

The sampling distribution of the goodness of fit statistic, SSE was also under investigation in this experiment. The theoretical sampling distribution is the chi-square with 3 degrees of freedom. The observed sampling distribution from 1000 runs for various n and σ were compared to this in 3 ways. The Kolmogerov-Smirnov goodness of fit statistic, D was tested for each distribution and their means and 95th percentiles were plotted as functions of n and σ. The values of D, which is the maximum discrepancy between observed and expected cumulative relative frequencies are shown in table 3.3. D is significant in most cases where n = 7 and n = 15 indicating departure from the theoretical distribution. There is also significant departure when n = 60 and many of the P_{ij}'s are small. The graphs of mean SSE against σ for different n are shown in figure 3.8 and the theoretical value is also plotted. The significant D values and discrepancy between observed and theoretical mean values in many cases show that with small n the theoretical distribution is a poor approximation of the actual one, particularly when the sampling standard deviation, σ is small.

With small n, the asymptotic sampling distribution of SSE should only be used as a rough guide in assessing goodness of fit. Conventionally, a 5% significance level would be adopted. The values of chi-square for this significance level observed in the experiment are plotted in figure 3.9 and compared to the theoretical value. It can be seen that using the theoretical value the model would be accepted more often when it was wrong than would be the case using the actual values. However, for n = 15 and n = 60 the observed and theoretical values are quite close except for very small σ. For these values of n the results of the experiment suggest that the theoretical 5% significance level will give a fairly realistic, though somewhat conservative goodness of fit test of the basic Thurstone scaling model. For n = 7 using the theoretical 5% level will generally provide an unrealistic and very conservative test. The graphs also show that in all cases where bias is large compared to MSE the

the theoretical goodness of fit test will be poor.

It is unknown how far the conclusions from this experiment can be generalized. This is a problem of all experiments. Since no other pertinent results are known, our opinions about the reliability of Bock and Jones' procedures when n is small will be based on the present results.

Table 3.2.

The Theoretical binomial Parameters of the 6 Pairs of Objects for Values of the Random Sampling Standard Deviation, σ.

σ	P_{14}	P_{13}, P_{24}	P_{12}, P_{23}, P_{34}.
		Theoretical Binomial Parameters	
0.3	$.28\times10^{-6}$	$.43\times10^{-3}$	.0475
0.4	$.88\times10^{-4}$	.0062	.1056
0.5	.0014	.0228	.1587
0.6	.0062	.0475	.2033
0.7	.0162	.0764	.2389
0.8	.0307	.1056	.2643
0.9	.0485	.1357	.2912
1.0	.0668	.1587	.3085
1.1	.0869	.1814	.3264
1.2	.1056	.2033	.3372
1.3	.1251	.2206	.3520
1.4	.1423	.2389	.3632
1.5	.1587	.2514	.3707
1.6	.1735	.2643	.3783
1.7	.1894	.2776	.3483

Table 3.3.

Kolmogerov-Smirnov D Statistics to test the fit of the Observed Distribution of SSE to the Theoretical Chi-square Distribution for Each Set of 1000 Runs.

σ \ n	7	15	60
.3	-.348	-.622	-.690
.4	.156	-.219	-.185
.5	.212	.113	.099
.6	.256	.178	.071
.7	.268	.123	.087
.8	.224	.124	.037[b]
.9	.184	.088	.010[b]
1.0	.180	.070	.010[b]
1.1	.133	.024	-.012[b]
1.2	.123	-.025	.013[b]
1.3	.059	-.038	.022[b]
1.4	.068	-.060	.029[b]
1.5	.050[a]	-.066	.013[b]
1.6	.049[a]	-.054	-
1.7	1.037[b]	-.068	-

All values in the table are significant at the 1% level, except the following: those with the suffix a are significant only at the 5% level and those with suffix b are not significant at the 5% level.

FIGURE 3.1

Graphs of mean estimates of $(\alpha_1 - \alpha_4)$ against the sampling standard deviation, σ for different n.

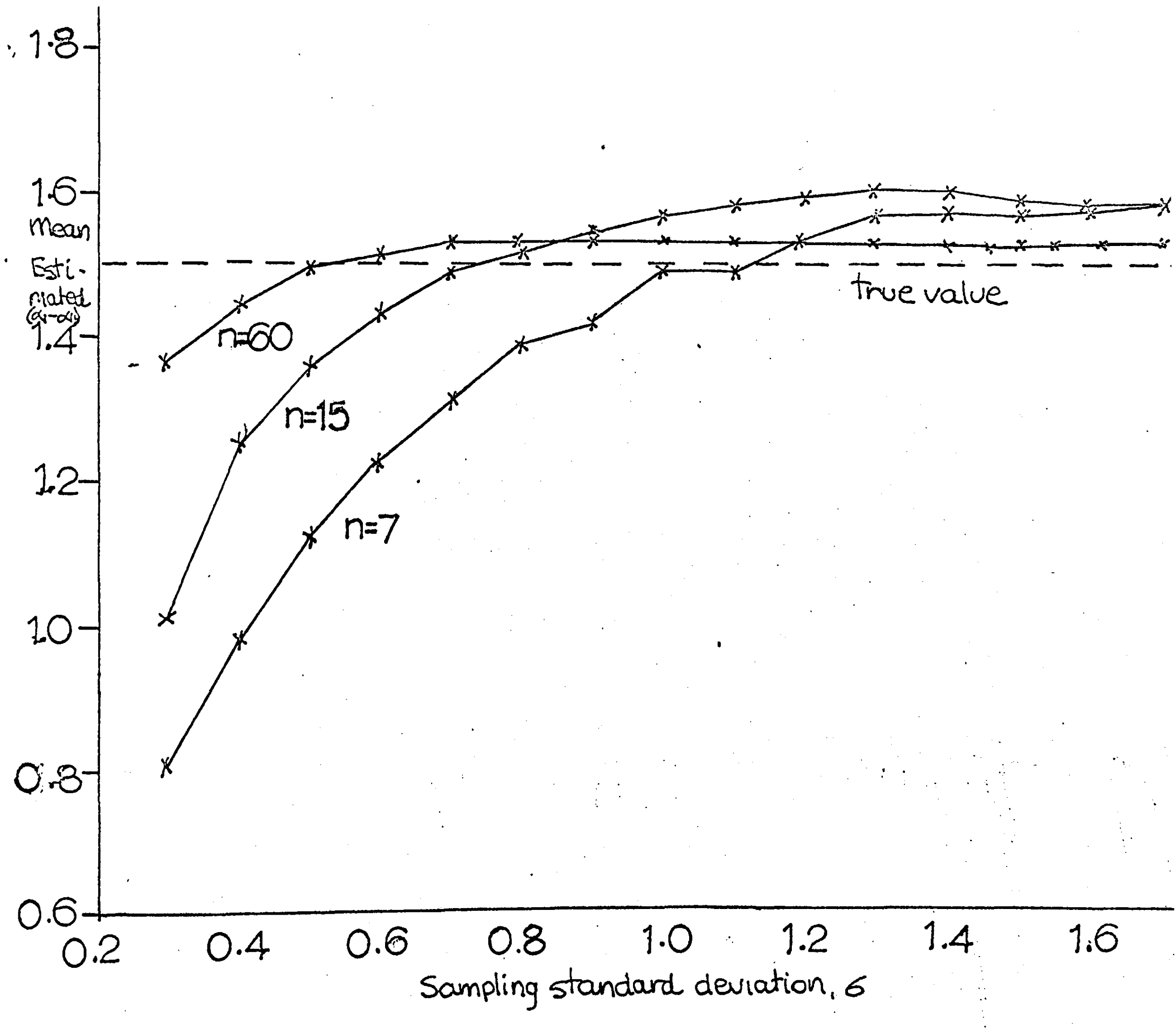


FIGURE 3.2

Graphs of mean estimates of $(\alpha_2-\alpha_4)$ against the sampling standard deviation, σ for different n.

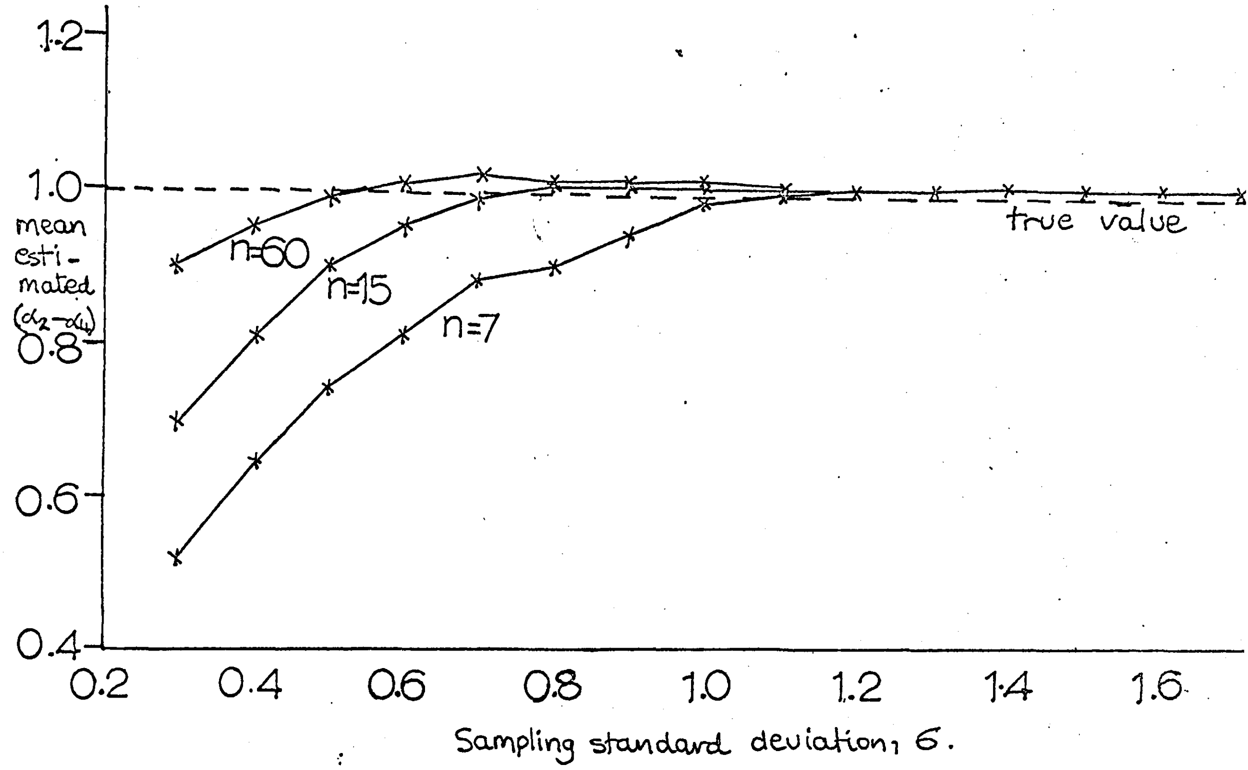


FIGURE 3·3

Graphs of mean estimates of $(\alpha_3-\alpha_4)$ against the random sampling standard deviation, σ for different n.

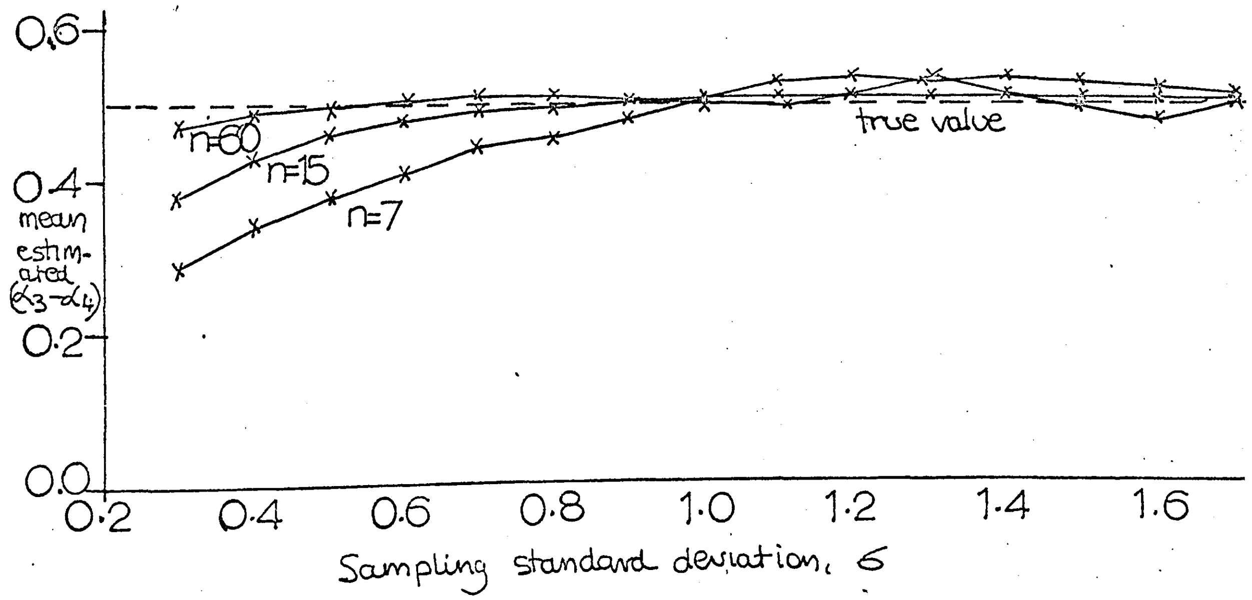


FIGURE 3.4

Graphs of Mean Square Error (MSE) and Theoretical Variance of $(\alpha_1 - \alpha_4)$ against the sampling standard deviation, σ for $n = 7$ and $n = 15$.

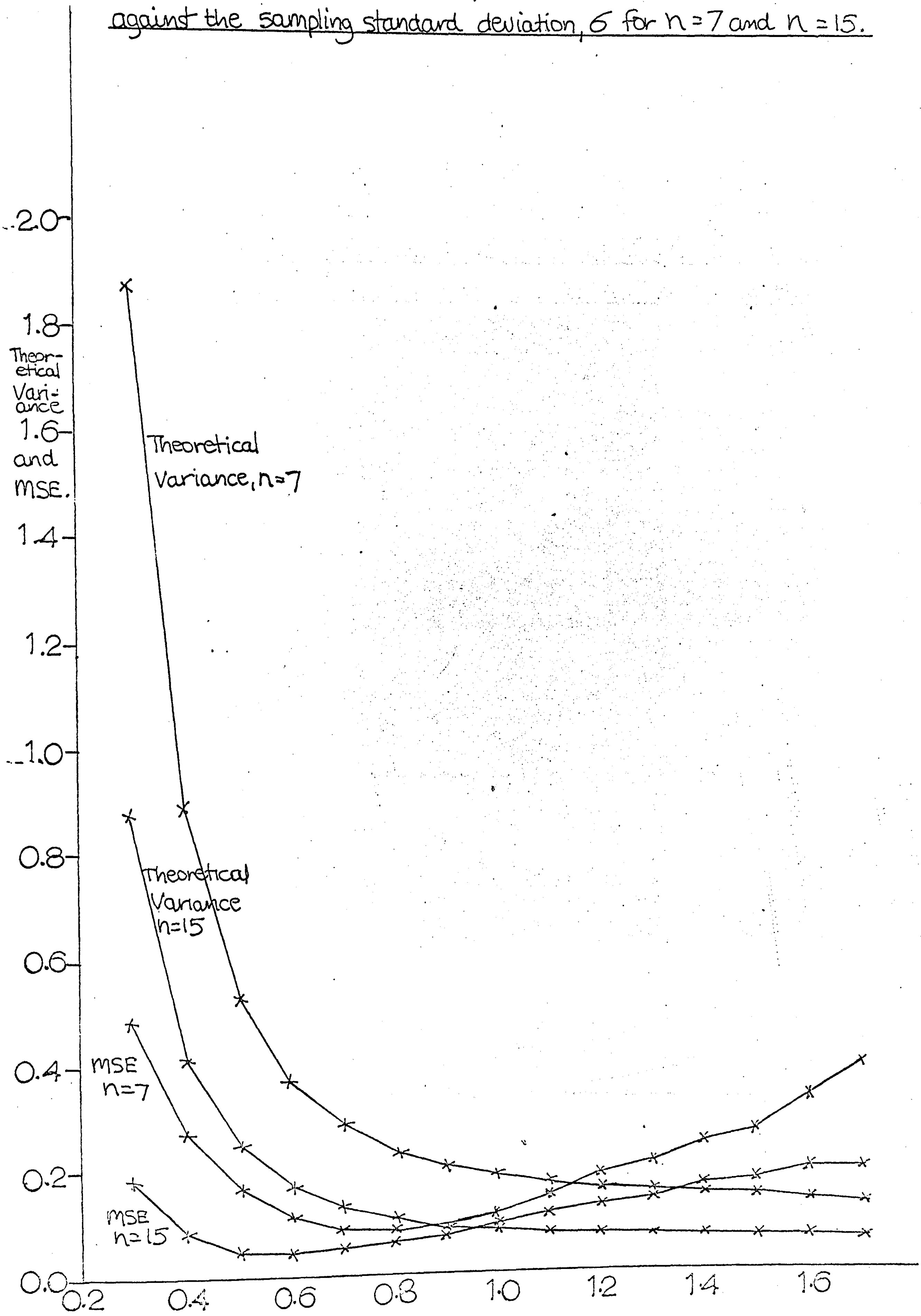


FIGURE 3.5

Graphs of MSE and Theoretical Variance of $(\alpha_2-\alpha_4)$ against the sampling standard deviation, σ for $n=7$ and $n=15$.

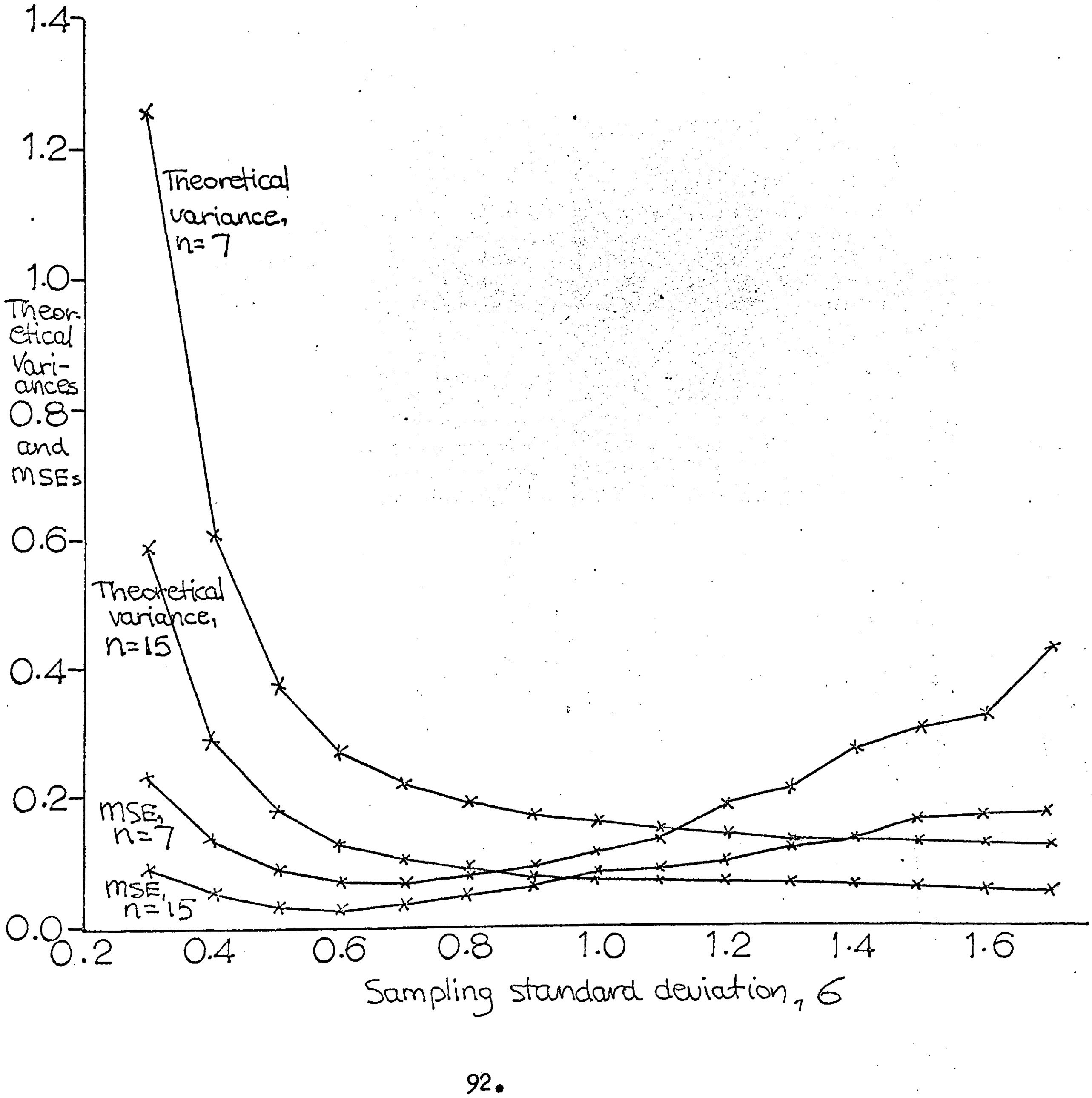


FIGURE 3.6

Graphs of MSE and Theoretical Variance of $(\alpha_3-\alpha_4)$ against the sampling standard deviation, σ for $n=7$ and $n=15$

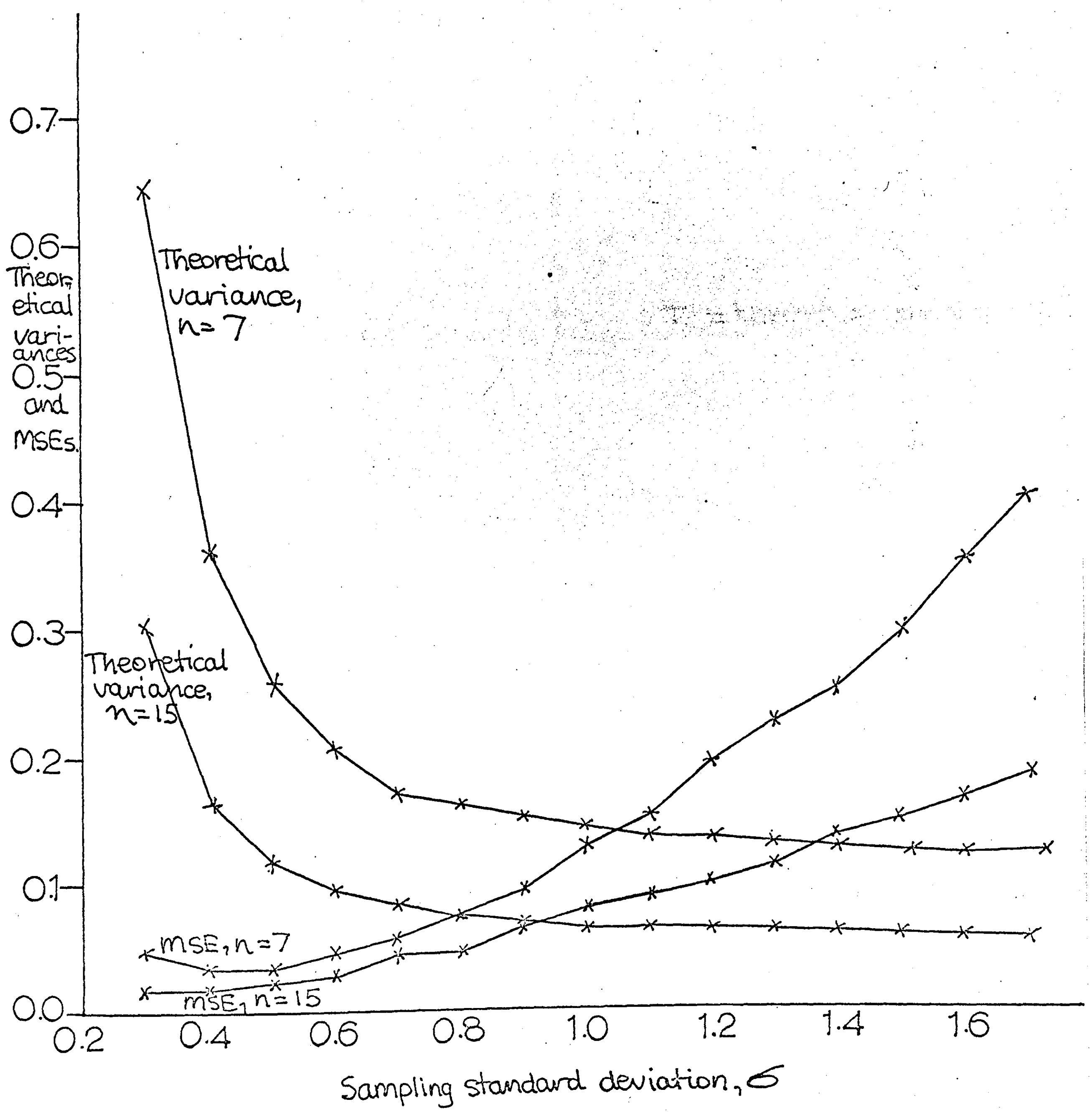


FIGURE 3.7

Graphs of MSE and Theoretical Variance of $(\alpha_i - \alpha_4)$ $(i=1,2,3)$ against the Sampling Standard Deviation, σ for $n=60$.

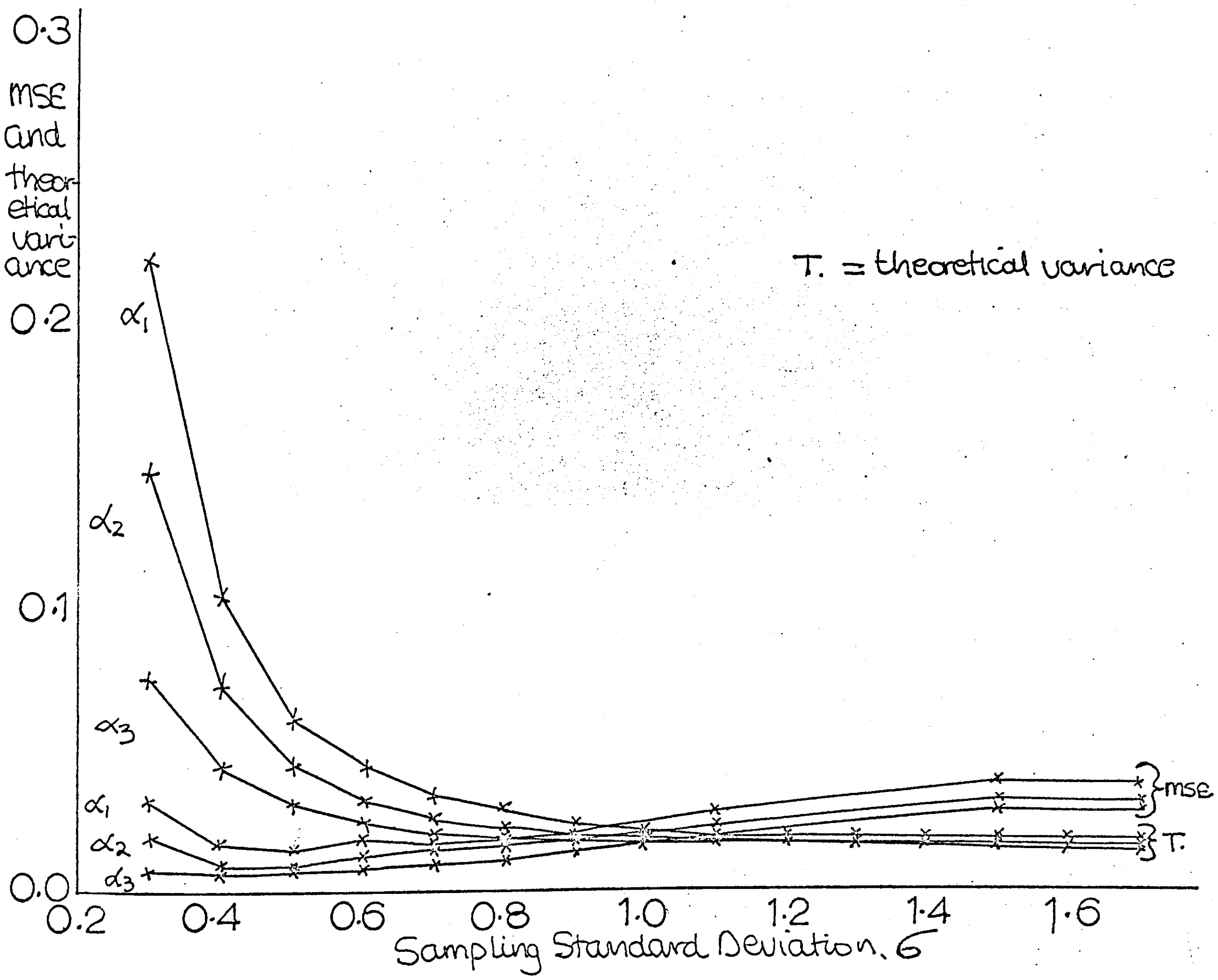


FIGURE 3.8

Graph of mean SSE against the sampling standard deviation, σ for different n.

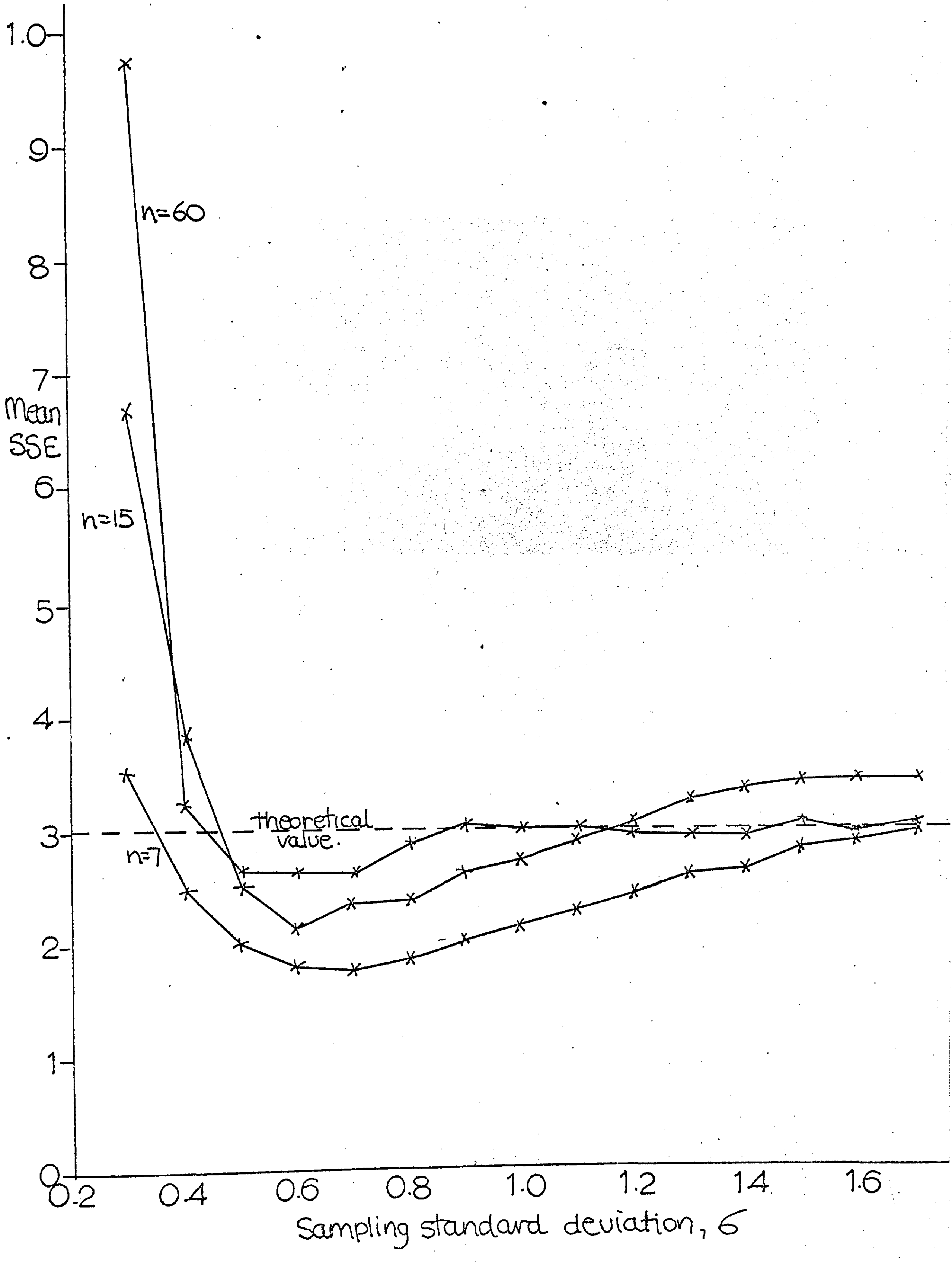
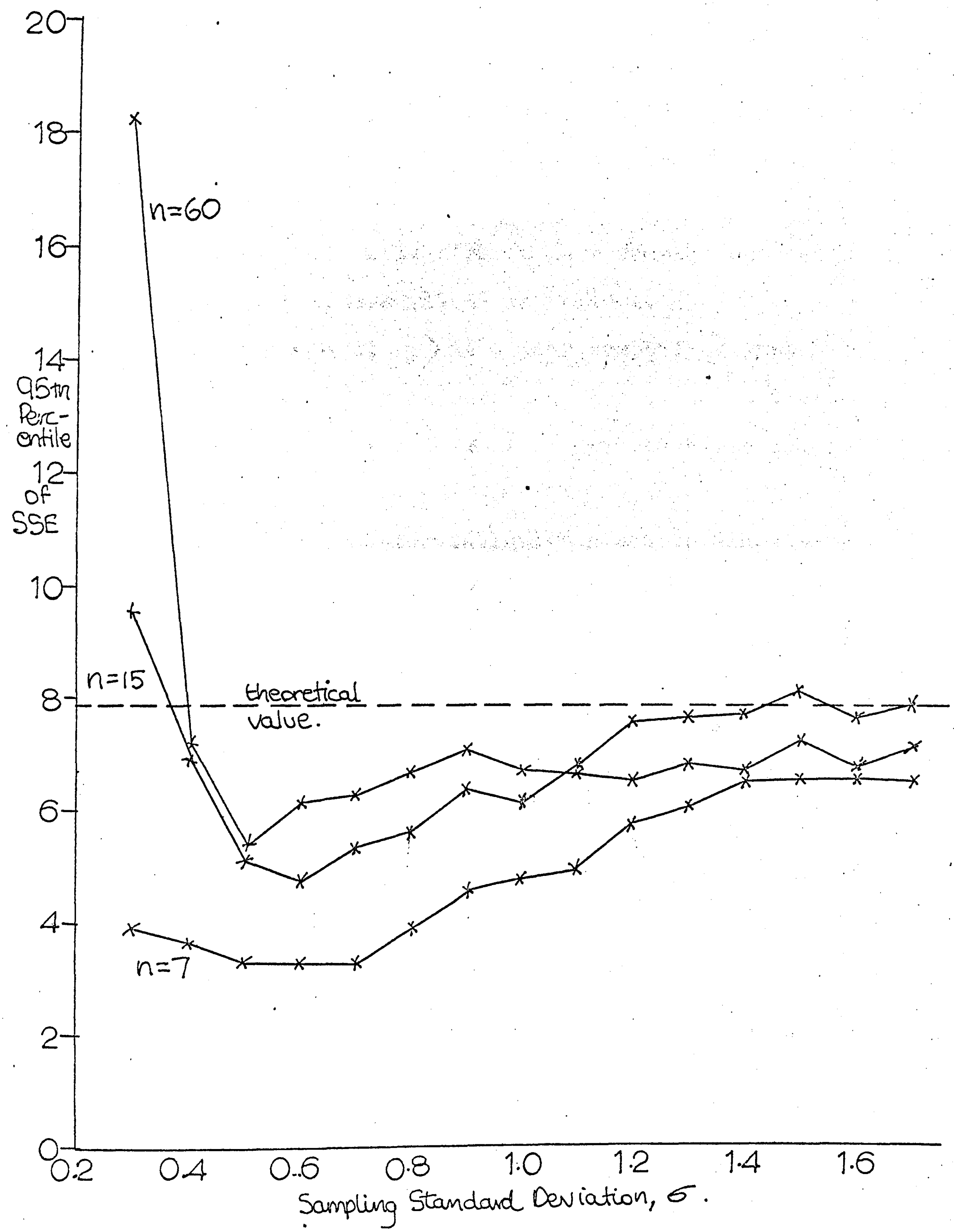


FIGURE 3.9

Graph of the 95th Percentile of the SSE Distribution Against the Sampling Standard Deviation, σ for different n.



Qualitative Vs. Quantitative Analysis.

Accounts of two approaches to the study of functional relationships and ways of applying them to choices in static situations have been given. Each analysis assumes a different specific model. The basic assumption, common to both is that a binary choice is a bernoulli variate and choices from the same pair have the same binomial parameter. Successive choices are assumed to be independent of one another. This is equivalent to saying that the models derive from a pair comparison system, as defined by Suppes & Zinnes (1963) as follows.

The system $\langle A, p \rangle$ is a pair comparison system if there exists a probability measure p on A x A (the set of pairs from A) such that for all $a, b \in A$ $p(a,b) = 1 - p(b,a)$, $0 \leqslant p(a,b) \leqslant 1$.

All the information integration models discussed, when specified according to one of the functional analyses make restrictions on p and are therefore non-trivial pair comparison systems. The exception to this is the quantitatively defined general linear model. The differences in the models are in the degree and kind of restrictions placed on p so they should be directly comparable with respect to their restrictiveness. By enumerating the permissible p functions for certain models one could derive critical tests to distinguish among them. If the enumeration is not possible analytical methods may be available. Burke and Zinnes (1965) made such a comparison of the Thurstone and the Bradley-Terry-Luce model. Such results, though eventually of great value are only of limited interest at present as the different kinds of model are not seen as competing. It is the different functional relationships which are competing and the quantitative and qualitative analysis of a single functional relation should complement one another.

If decision-making behaviour in a situation is found to be consistent with a quantitatively or qualitatively defined information integration model very strong predictions can be made. The predictions are different in kind. From functional measurement, estimates of the subjective value of any alternative varying along the appropriate dimensions can be made, either from relations among psychophysical or purely psychological scales. Then, for a pair of alternatives the preference probabilities can be predicted. From QFA, predictions about qualitative laws of behaviour can be made, which in many cases might be more useful for the theory of decision-making. From QFA, the only predictions about elements not in the original experimental set would be general ones about how they fit in with the other potential alternatives.

There are specific assumptions for each kind of model and each set may be reasonable in different circumstances. In particular, the following differences exist. The models underlying QFA predict that preference is either quite definite or there is no preference. The Thurstone type models predict that slight preferences can be found but that preference probabilities cannot be one or zero. In this respect they are conflicting models. This difference is rather fundamental and is a strong reason for considering critical tests along the lines of that suggested by Burke and Zinnes. The results of such tests, however, can be anticipated. Some situations where slight preferences can be reliably detected and others where preferences probabilities are zero would probably be found, negating both models. But models can only be rejected when better ones are available to replace them. Tests exploring the domain of a model are of course important but if a model can be shown to have a wide, though not all-embracing range then it is a good one. Both kinds of models have had considerable success in accounting for data. The experiments to be reported are attempts to explore their usefulness in situations where they have not been tried. In this way, rather than by critical tests, it is hoped that our knowledge of how people integrate the information available in making their decisions will be built up.

Chapter 4.

Review of Empirical Research on Individual Decision-making.

Some Empirical Issues.

A basic experimental problem that must be discussed is how to obtain information about a DM's preferences among the elements of some set of multi-attribute alternatives that are available to him. In previous chapters it has been suggested that actual choices should be observed as it is actual decision making that is our concern. But other dependent variables can be used. In examining many models it would appear to be more efficient to ask people to give a rating of each alternative with respect to subjective value.

An objection to this might be that people will process the information differently when rating from the way they process it when actually choosing among alternatives. Generally in rating experiments alternatives are presented singly. Processing during evaluation of single alternatives quite plausibly will be different from that when a group of alternatives are compared. However, this difficulty could be overcome if a comparative rating of DM's degree of preference, say, was elicited. Such ratings where pairs of alternatives are presented and subjects rate their degree of preference for one over the other have been used in a few studies e.g. Shanteau and Anderson (1969).

Only one study is known that compares the way people process information when rating and when choosing. Anderson and Alexander (1971) carried out a personality impression formation study making a comparison of information processing in a choice and a rating experiment. Conclusions from both were broadly in accordance with the same averaging model though there was one contradiction. Obviously this result alone is rather inconclusive as far as the general question is concerned.

Most studies using ratings are conducted in imaginary situations. An objection to the validity of these studies might be that ratings would be different when obtained under real conditions. There is little evidence about this but there is evidence that real and imaginary choices are different. Slovic (1969a)) had two groups make pair comparison choices from duplex gambles. In one group the gambles chosen were played to determine the subject's salary while in the other choices were hypothetical. In the hypothetical group DM's tended to discount the possibility of loss in making their choices but the group who knew they would have to play some of the gambles were more cautious.

In view of this it would seem preferable, whatever response is elicited that subjects should be aware that consequences which are dependent on their responses will occur. Only in a few studies have consequences been made dependent on ratings. Sjoberg (1968) and Slovic (1969b)) had subjects rate many gambles singly. They were informed prior to this that pairs of the gambles would be chosen by the experimenter (independently of how the subjects respond) and the subject would be required to play that alternative which he had rated higher. Thus, when he made any rating he knew that it would affect which gambles he would be required to play. A similar ploy could be used with ratings of degrees of preference. Average ratings for each alternative could be computed from ratings of degrees of preference and this set of measures could be used to make consequences dependent on ratings. There is no evidence about whether ratings would lead to different conclusions from those drawn from choice data in situations which are real in the above sense.

Evidence on a related question is available, however. This evidence compares choices to a different kind of rating on which consequences can be made dependent. This response is elicited by a technique, introduced by Becker, Degroot and Marschak (1964) which involves making either bids to buy or bids to sell an alternative. With the bids to buy method DM's do not have the opportunity to take the alternative and they must name the maximum amount of money they are willing to pay in order to obtain this opportunity. In the bids to sell method DM already has the opportunity and he must name the minimum amount of money he would take in return for forfeiting it. For both methods, Becker et al, (1964) have shown that the optimum strategy is to name an amount of money exactly equivalent in subjective value to the alternative. The buying or selling prices can be interpreted as ratings of attractiveness of the alternatives. They are made to have a consequence by letting the buying and selling actually take place. The question is, then, do bids and choices lead to equivalent conclusions about decision making when consequences follow them?

Lichenstein and Slovic (1971), and Lindman (1971) have examined this using a set of duplex gambles. Lindman points out that the importance of this issue is that most studies using bids have assumed them equivalent to choices and comparisons between the conclusions of bids and choices experiments have been made accordingly. Lichenstein & Slovic conducted three experiments, two comparing selling bids with choices and the other comparing buying bids with choices. The second of the selling bid experiments was the only one where bids and choices were real. They found systematic differences between gambles chosen and bids made for the same gambles. Ones with favourable probabilities were chosen while those with favourable payoffs received greater bids. Lindman conducted five experiments which supported this. Although subjects responses were imaginary the experiments ruled out contextual factors that might produce DM's inconsistency between bids and choices. The seven hypothetical response experiments are all supportive of Lichenstein and Slovic's real consequences experiment.

There is also evidence that bids and ratings lead to different conclusions. Slovic and Lichenstein (1968a)) and Andriennsen (1971) have carried out experiments where subjects rated and bid for duplex gambles. Both studies conclude that subjects gave more importance to the payoff dimensions when bidding than when rating. This is a similar result to that obtained when comparing bidding and choices.

The general conclusion of all this is that people appear to process the information available differently when they are required to bid for the alternatives from when they are required to rate them or choose among them. There is slight evidence, obtained in a hypothetical situation that they process things differently when choosing from when rating. However, it would not be surprising if results from choice and rating experiments broadly coincided especially if degree of preference rather than preference ratings were used. A decision to "play safe" and use actual choices has been taken in most of the experiments to be reported.

This is partly motivated by the lack of evidence from choice experiments relating to information integration models of decision making, particularly quantitative ones. It is unfortunate that in pursuing such evidence a penalty of inefficiency is incurred.

To obtain sufficient information about preferences among a set of alternatives from choice data it is necessary to look at average choices over subjects or trials. But, subjective values may shift from person to person or time to time. The practice of averaging over subjects is not to be recommended for phenomena like decision-making for which large differences among individuals are likely to be found. Thus, one must average over trials for individual subjects. If subjective values change systematically over trials then average choices, measured by binary preference frequencies, are meaningless. All the models of behaviour discussed require that the preference structure that underlies choice behaviour is stationary over time. Predictions about decision-making are based on this assumption. Obviously, it must be found to hold during the time that choices are being made in an experiment. Lindman, using a chi-square test, found that choices did not change during his experiment while bids did. This was for data averaged over the group, but it does suggest that choices are more stable than bids. Luce and Suppes (1965) voice the fear that in situations where the presentation set of alternatives is changed from trial to trial asymptotic choice behaviour will not be observed. This is because, they believe substantial sequential dependencies will exist between the responses. However, the opinion offered here is that in a properly randomized experiment these effects will not be systematic. They can be assumed to produce random error in the observed binary preference frequencies.

Even so, it is still necessary to try and test whether any systematic response dependencies or trends are present. Most choice studies do not do this, often because insufficient data is available. In the experiments to be reported here, however, certain tests against obvious memory effects and general trends have been used. The experimental paradigm generally used was described in chapter 2. A set of m pairs of alternatives are presented to the subject n times each. The mn trials occur in a predetermined sequence, random except that if a pair is presented at trial k it has not appeared on the previous l trials. The hope is that due to the similarity of the choice tasks that are interspaced between successive presentations of a pair, and also because of the time duration between such presentations the subject forgets what he previously chose. Also it is hoped that if l is about m/2 or m/3 the constraint does not bias the randomization too much. Before the procedure can be successful, then m must be fairly large. Before general tests of dependency or trend can be used n also must be large. On the other hand m x n must not be too large or the subject will get too bored. A successful experiment therefore, will require a careful choice of m, n and l. It is hoped that Luce and Suppes' objections do not apply to this paradigm, but if they do it can be revealed by the precautionary tests.

Finally, in this discussion of empirical issues Edwards (1969) and Edwards, Slovic & Lichensteins' (1965) comments on general experimental procedures should be noted.

The earlier paper reports experiments investigating choices among gambles under different conditions. Some conclusions were: "shorter experimental sessions, individual administration and real gambling aid in motivating subjects and thus help to prevent boredom induced distortions in preferences." Unfortunately, no attempt was made to determine which of these factors was most important. Nevertheless the recommendations seems intuitively reasonable so they will be adopted as far as possible. In the later paper, a review Edwards (1969) makes the point that "gambling experiments in university settings are open to a number of objections: the stakes are trivial, the subject population is rather special, and the experimental conditions must necessarily be somewhat unrealistic." Unfortunately, his solution: "setting up a research laboratory in the Four Queens Casino in downtown Las Vegas" is not always available. Of Edwards three points the first two are empirical. It is to be hoped that wider domains will be explored fully in time. The third point is a problem that faces all experiments. Realism is what is sacrificed in order to attain greater control of variables. However, in the choice experiments to be reported, the situations are real in that real consequences follow decisions made. Before reporting these experiments some relevant substantive issues will be discussed.

Other Reviews.

There are many reviews of empirical work on human individual decision making behaviour from all points of view. (Edwards (1954c), (1961), Luce & Suppes (1965), Becker & McLintock (1967) etc.) The purpose of the remainder of this chapter is not to add another such comprehensive review. Recent trends will be discussed and work which gives direct evidence for and against some functional relationships among subjective values will be considered. In particular, work using methods similar to those described earlier will be considered in detail.

Studies Related to the QFA Approach.

A few studies have employed QFA (as understood here) to explore functional relationships among subjective values. The first was that of Fagot (1956) reported in Adams & Fagot (1959) which has already been discussed. In the analysis a rather useful procedure was adopted to give a full test of the additive hypothesis, H4. The hypothesis states that $(a,b) \precsim (c,d) \Leftrightarrow u(a) + u(b) \leq u(c) + u(d)$. A complete set of binary choices from a two product set and the above relation define a set of numerical linear inequalities. As they point out, for such a system of inequalities theorems are known which enable one to establish if a solution exists, and if so to find it (or them). Unfortunately this is only true for the additive representation. Where the additive model can be shown to hold by this method QFA is redundant. However, where it does not hold QFA can be used to examine the nature of the model's failure. Fagot did this for subject's data which violated H4 paying particular attention to the ordinal model H3 (that which says the weak order condition is satisfied). If QFA reveals that violations are not systematic then a search for the largest subset of observations which do not violate H4, the additive model can be made. If the number left out of this set is small then these observations can be put down to "errors" of judgement. Fagot's experiment elicited subjects choices for job applicants and showed considerable support for the additive utility model.

Before discussing the next experiment which used QFA, consider the following gambles $(x,\frac{1}{2},y)$ of the wheel of fortune type, where x is won or lost with probability $\frac{1}{2}$ otherwise y is won or lost, and let $x,y \in X$. QFA of the SEU model can be carried out by considering the special case of H4, postulating the existence of a single utility scale, u such that $(x,\frac{1}{2},y) \leqslant (z,\frac{1}{2},w) \iff u(x) + u(y) \leqslant u(z) + u(w)$. Now, since the elements all belong to the same set it seems a reasonable assumption that the above relationship between the utility sums is equivalent to the following one between utility differences: $u(x) - u(z) \leqslant u(w) - u(y)$. A choice between gambles $(x,\frac{1}{2},y)$, $(z,\frac{1}{2},w)$ is assumed to give evidence about utility difference according to this equivalence. That is, the relation $(x,\frac{1}{2},y) \leqslant (z,\frac{1}{2},w)$ can be assumed equivalent to a utility difference relation $(x,z)\,Q(w,y)$. An alternative QFA can then be carried out to test the SEU model using the following hypothesis. Let us call the Utility Difference Representation Hypothesis on $\mathcal{A}_0 = \langle A \times A, Q \rangle$ that hypothesis which states that there exists a real valued function, u on A x A such that for all $a,b,c,d \in A$, $(a,b)\,Q(c,d) \iff u(a) - u(b) \leqslant u(c) - u(d)$. The advantage of using this to examine SEU rather than H4 is that it makes full use of the facts that all elements which occur belong to the same set and that any pair can be compared to any other pair. Fagot (1959) has carried out a QFA for the strict utility difference representation hypothesis, the same as the above except that it does not allow utility differences to be equal. Suppes & Winet (1955) have proposed an axiom system sufficient for the full representation hypothesis. Adams, Fagot & Robinson (1971) have suggested a subset of this axiom system which contains much of the empirical content of the hypothesis and would therefore be suitable for QFA. Thus, the results to carry out QFA of the utility difference representation of the SEU for the above gambles are known. Only one experiment using these results is known, that of Fagot (1959).

His omission of the possibility of equality of utility differences greatly simplified his analysis. A group of 10 undergraduates had to give statements as to the differences in utility of pairs from a set of class grades, A,B,C,D and E. It was assumed that the alphabetic order was their preference order for the grades. If at most one of their statements was reversed, 7 of the 10 S's would have satisfied all the qualitative consequences of the utility difference model that Fagot considered. The remaining 3 would have satisfied a strong sub-model. This is strong evidence for Fagots strict utility difference hypothesis. No other studies have tried to examine the utility difference hypothesis by QFA. Most studies examining the hypothesis have been concerned with obtaining an interval scale of the elements. Tests of the models have been made by using the scale obtained to make predictions about further choices. This approach is useful because it shows that models using the relational system formulation can yield quantitative data. What it does not reveal, though is the source of erroneous predictions, as a thorough QFA would do. These scaling studies (e.g. Hurst & Seigel (1956), Seigel (1956) Coombs & Komorita (1958) Davidson, Suppes & Seigel (1959) have been extensively discussed by Luce & Suppes (1965).

Two studies examining choices among risky alternatives by QFA are those of Tversky (1967a), and Coombs, Bezimbinder & Goode (1967). Tversky obtained data by the Marschak bidding technique and in the latter study pair comparison choices were observed in addition to bids. Tversky's experiment will be discussed in more detail when quantitative studies are considered since most of the analysis is based on functional measurement.

Eleven prisoner S's were asked to give selling prices for certain sets of alternatives: i) risky ones, where they could win a number of cigarettes, bags of candy or both with a given probability (the complementary event being zero) ii) riskless ones, comodity bundles of x cigarettes and y bags of candy. There were three sets of risky and one set of riskless alternatives. The SEU hypothesis was tested for 3 of the sets, the fourth being reserved to test predictions using a variant of the linear programming method used by Fagot (1956). The observations are on at least an ordinal scale and the application of this procedure is a full ordinal functional analysis of the additive or multiplicative hypothesis. Solutions of the appropriate linear programmes were found for the largest subset of each set of data. The Kendall rank correlation coefficient, t between each solution and its data matrix was computed as a measure of the degree of additivity for that set. Twenty six of the 33 data sets examined were perfectly additive and the worst had t = 0.950. This data, then gives strong support for additivity and it was not necessary to apply any further analysis, such as QFA.

The Coombs, Bezimbinder and Goode study actually precedes that of Tversky. Two experiments were carried out to test qualitative predictions of expected value (EV), subjectively expected value (SEV), expected utility (EU) and SEU models. Subjects choices between pairs of simple gambles and bids for them were observed. The SEU model tested was actually a strict SEU model such that no indifference was permitted. As observed in connection with Fagot (1959) this greatly simplifies the analysis. The probabilistic preference model: $p(a,b) = \pi \Leftrightarrow a \prec b$ was assumed. Essentially, the test of the four models was as follows. An unrestricted estimate of π and also an estimate under the restriction of some implication of an expectation model was made. If the latter estimate was significantly lower than the former then the model was rejected. This is because the expectation model giving the result can only explain the data by assuming greater inconsistency than the unrestricted model which assumes only constant π.

The strict (i.e. no indifference) SEU model was tested by estimating π subject to the following consequence of it. For simple gambles $(p,s) \in P \times S$:

If $(o,t) \prec (p,s)$ and $(r,s) \prec (q,t)$ and $(p,u) \prec (o,v)$ then $(r,u) \prec (q,v)$, for all $o,p,q,v \in P$ and $s,t,u,v \in S$.
It was Tversky's paper which pointed out that the above condition, called the triple cancellation condition is equivalent to Coombs et al's theorem 5. Both the experiments reported by Coombs et al were designed to make this test of SEU as well as similar tests of EV, SEV and EU. The experiments, using widely differing samples of subjects gave very strong evidence that the above condition was not violated. It should be noted, though that even if it had been violated the finding could be explained by proposing that subjects had been indifferent between some of

the gambles.

There have been a number of studies which have tested the transitivity condition. Transitivity, that is of the weak preference relation. Its strict preference and indifference components have not been considered separately. Various kinds of stochastic transitivity, consequences of different choice models based on preference probabilities have been defined. That most directly related to "algebraic" transitivity is generally called weak stochastic transitivity, which holds for all a,b,c, $\in$ A:

$$p(a,b) \leqslant \tfrac{1}{2} \text{ and } p(b,c) \leqslant \tfrac{1}{2} \Rightarrow p(a,c) \leqslant \tfrac{1}{2}.$$

This is equivalent to transitivity as defined earlier, with preference defined according to probabilistic model one and $\alpha = 0.5$.

The experimental evidence relating to this condition is reviewed by Luce and Suppes (1965). It generally supports transitivity. Most of the evidence against it is discussed by Davis (1958). He suggests that most indications of intransitive choices that have been found can be ascribed to S's being indifferent among certain of the alternatives presented to them. He concentrates his argument on data about the proportion of circular triads found in experiments by Edwards (1953, 1954a)b)) and May (1954). Davis says of Edwards' interpretation of his finding: "Circular triads were found in Edwards' first experiment (1953) in about 20% of the total number of times they could occur...... (He) says that although circular triads might be expected to occur if the subjects were indifferent the results of the vote count showed that they did not choose at random. This argument of Edwards is logically fallacious..... it does not follow from the premises that if there is significant agreement between subject choices and that there are circular triads in 20% of the possible triads of these choices, that there is agreement on the circular triads among subjects."

Davis performed two experiments, one using similar alternatives to those used by May and the other using ones similar to Edwards. He tested to see whether stable intransitivities could be found by carrying out complete pair comparison experiments and then replicating them using the same S's. If the proportion of triads that were circular on both replications was greater than that expected by chance he would conclude that stable intransitivities had been found. In neither experiment was he able to draw this conclusion.

A study by Tversky (1969) is the only one to reliably produce intransitivity of choices. Again indifference is not treated separately from strict preference, but inspection of the data indicates that the intransitivity found cannot be explained by S's indifference among some alternatives. Two n-replicate pair comparison experiments were carried out, the first using simple gambles and the second using job applicant profiles. This study differed (as far as one can ascertain) from every previous investigation of transitivity in one important respect. The difference between the values of the alternatives on one of the dimensions was either near to being indiscriminable or such that it did not reliably reflect a real difference. Tversky suggests that under these conditions people may choose according to a decision rule based on a lexicographic semi-order: "Consider, for example a situation in which three alternatives, x,y and z, vary along two dimensions, I and II and where the values of these dimensions are given by the following payoff matrix.

		Dimensions	
		I	II
	x	2ε	6ε
alternatives	y	3ε	4ε
	z	4ε	2ε

.... Suppose the subject (S) uses the following decision rule in choosing between each pair of alternatives: if the difference between the alternatives on dimension one is (strictly) greater than ε , choose the alternative that has the higher value on dimension I. If the difference between the alternatives is less than or equal to ε choose the alternative that has the higher value on dimension II. It is easy to see that this seemingly reasonable decision rule yields intransitive preferences when applied to the above matrix."

This decision rule, first considered by Davidson, McKinsey & Suppes (1955) is called by Tversky a Lexicographic Semi-order, (LS). It is a special case of an additive difference model, where DMs consider the utility difference of two alternatives along each dimension and form the overall utility difference of the alternatives additively across the dimensions. As well as his discussion of the special case one of the major contributions of Tversky's study is his discussion of additive difference models of preference.

The first experiment employed five simple gambles each represented by a wheel of fortune. The payoff information was given but the probability of a win was not explicitly given. S's had to determine differences in probability by comparing the win areas of the wheels. An n-replicate pair comparison of the five alternatives was carried out and the transitivity condition was tested by a likelihood ratio test similar to that suggested earlier for QFA with probabilistic preference model I. In fact, the tests suggested earlier are extensions of Tversky's test in this experiment.

Of 8 subjects tested, 7 were found to violate weak stochastic transitivity. Again by the LR test, only one subject appeared to violate the LS model. It should be remarked that these subjects were "screened" in a pre-session and selected as being likely to show intransitive choices. It was said that this was in order to find S's who would apply the LS decision rule but equally, S's who were poor at discriminating differences among the areas representing the probabilities could have been inadvertently selected. Luce & Suppes (1965) and Irwin (1958) have discussed the relation between discrimination and preference. They point out that people can only prefer one thing over another if they can discriminate between them. Thus, one could put the intransitivities in this experiment down to poor discriminability of the probability dimension. In which case the experiment does not reveal anything basic about the structure of preference.

The second of Tverseky's experiments is not subject to the above criticism. The alternatives used were three-dimensional job applicant profiles similar to those considered by Adams & Fagot (1959). An applicant's profile was represented as a bar diagram, the heights of the three bars representing his level on each of the three factors intellectual ability, emotional stability and social facility. Again, S's were screened in a pre-test to find those likely to use the LS rule. This time, in addition, stimuli were selected for each S which were expected to yield intransitive choices. A 3-replicate pair comparison of 10 alternatives was carried out. Comparisons of the observed with the expected number of circular triads given that S's were stochastically transitive and given that they used the LS rule were made. The observed value of this statistic was close to that expected under the LS rule for 11 of the 15 S's tested. The observed proportion of circular triads for the group was significantly different from that expected under stochastic transitivity but not from that expected under LS. Inspection of the alternatives used does not suggest that S's would be unable to discriminate between different values on any dimension. A better interpretation is that S's supposed small (but discriminable) differences in intellectual capacity scores did not reliably reflect a true difference. A conclusion from these results can be suggested. Where the information is perceived as approximate, as not being an accurate representation of the real world then the LS decision rule is often followed. Violation of transitivity has not been demonstrated in situations where DM's information is accurate and complete.

Some studies have been carried out which can be interpreted as being favorable to the inexact SEU model but unfavorable to the exact SEU model. The first of these was in the extensive study based on the utility difference representation hypothesis by Davidson, Suppes & Seigel (1957). In this their main concern was measurement of the utility of money and the subjective probability of certain chance events. They did, however, conduct an experiment to obtain choices between gambles of the $(x,\frac{1}{2},y)$ type discussed earlier where the alternatives were gramaphone records. They used the linear programming method to test the additive difference hypothesis by converting the choices appropriately to linear inequalities. The exact model failed, but instead of proceeding as Adams & Fagot did and putting violations down to error they introduced a positive constant into the inequalities on the "greater" side and solved the new linear programme minimizing the positive constant. A solution to this always exists and they tested its suitability by the number of correct predictions it made. Their conclusions were favorable to their linear programming model, which, as indicated, is the inexact SEU model in another guise.

Wallesten (1971) tested the inexact SEU model using simple gambles by presenting pairs of gambles (p,x), $(q,y) \in P \times X$. He let one of the four parameters be adjustable and let the subject adjust it so he was indifferent between the gambles. Letting the utility of zero be zero and assuming SEU gives the equality $s(p)u(x) = s(q)u(y)$. Not surprisingly, when a set of such equations were obtained with overlapping p's and x's no solution could be found. Consider though that the inexact SEU model predicts that $s(p)n(x)/s(q)u(y) \leq \varepsilon$ when two gambles (p,x) and (q,y) are judged indifferent. If logs are taken this gives linear inequalities which can be solved minimizing ε. Essentially, this was Wallesten's procedure. He determined $s(p)$

and U(x) scales and considered whether scale values were consistent with certain independent adjustments of indifference. This was not as good a test of the inexact SEU model as QFA would give, but the evidence was favorable to the model. Both of the above studies attempted to fit exact SEU scales, failed and put this failure down to the nature of indifference.

Let us now see what line of enquiry with QFA is suggested by previous research related to it. Most of the studies which tested rperesentation hypotheses did so with some linear programming method, not by QFA. Such methods cannot be used to pinpoint failures as QFA can, and they are limited in that results for additive representations only are known. Other studies used techniques to estimate subjective scales and then tested their predictions. It is felt that QFA stands in the relation of a "goodness of fit" test to these scaling methods. QFA can give adequate grounds for rejecting the representation which the scaling procedures assume. Rejecting representations on the grounds of poor predictions is generally rather arbitrary.

Some studies have employed QFA and observed the frequency of violations of some consequence of a representation. These can be differentiated by their attitude to the predictions made about indifference by exact representations. Many studies "lumped them in" with strict preference and otherwise ignored them. One study explicitly excluded indifference (Coombs et al (1967) and Davis (1958) actively discussed it in his discussion of transitivity of choices.

Studies applying linear programming techniques have also had different assumptions about indifference which affected their treatment of violations of models. Some lumped indifference and preference together and called violations "errors" while others explicitly separated indifference from preference and assumed violations of conditions occurred when pairs of alternatives were in the same indifference region. These contradictory views of indifference suggest that empirical evidence about its true nature should be sought. This is attempted in chapter 5.

The results of chapter 2 are particularly applicable to simple and duplex gamble alternatives. Two previous studies, Coombs et al (1967) and Tversky (1969) applied QFA to simple gambles and found conflicting results with respect to the SEU model. No previous studies have applied QFA to duplex gambles. These seem two good reasons for attempting to gather further evidence from an application of QFA to these cases.

Functional Measurement Studies.

There has been much discussion of the random utility models, of which the models of the last chapter are special cases, and their relation to binary (and more than binary) preference probabilities (Becker, Degroot & Marschak (1963), Morrison (1963), Burke & Zinnes (1965), Luce & Suppes (1965), Becker & McLintock (1967). It has generally been held that the available statistical techniques to estimate the parameters of even special cases of these models are unsuitable for choice data obtained from individuals. Obviously results from their use on large sample group data (Bock & Jones (1967), Sanders (1961)) are inadmissible as evidence about the way individuals make decisions. The results that those of the last chapter are based on, by Berkson (1955), Cox (1958), Bradley & Terry (1952) etc. are discussed, but the point has been made that "......the known properties of their estimates are large sample properties...." (Becker, Degroot & Marschak (1963a)). This statistical fact has been seen as an insuperable barrier to the use of the estimation and goodness of fit procedures previously discussed. Empirical work has therefore been deflected into testing certain relationships among the preference probabilities predicted by the general random utility and other models, e.g. Becker, Degroot & Marschak (1963b,c). Many of these have been in situations where choices from more than a pair of alternatives have been observed. However, this deflection has not led to a great deal of empirical work for, as Luce & Suppes (1965) point out ".....nasty statistical problems that have hardly begun to be formulated, let alone solved" appear. They conclude that "Because the results of tests of observable properties of preference probabilities are almost never clear cut, one is left with a distinct feeling of inconclusiveness." Probably because of such problems with this approach and because parameter estimation appeared hazardous using individual binary choice data, other response modes have gained in popularity. All the evidence about peoples' subjective values to be reported uses some kind of rating response, either actual numerical or graphical ratings or bids to buy or sell alternatives. The evidence against the assumption that choices, bids and ratings can be used interchangeably has already been cited. In evaluating substantive evidence to be discussed it should be remembered that the following hypothesis: when a person rates one alternative higher than the other he may consistently make the contrary choice, has not been disproved.

It would be rather odd if the present trend of using rating type response modes continued and weighty evidence built up about how people make decisions, none of which included observations of actual decisions. The need to carry out experiments using actual choice data is one of the main reasons for attempting a quantitative study using choices.

Recent results from experiments which collected rating and bidding data to see which quantitative information integration models are favoured are conflicting. Two studies by Tversky (1967 a),b)), the first of which has already been discussed, broadly supported the classical SEU model. In the first, bids for simple gambles, cigarettes or packs of sweets were won with some probability and bids for "commodity bundles" of cigarettes and sweets were obtained from prisoner subjects. In the second, bids for simple gambles where money was won or lost were observed. In both studies analysis of variance of the bids (logs of the bids in the case of gambles) showed good support for the general SEU information integration model. In the second study, the classical SEU model, with utility of money assumed as a power function of money, u(o) = 0 and u(1) = 1 gave excellent prediction of the bids. The study shows that the utility scale of the bids was about the same as the utility scale of the value part of the gambles, (so that subjects showed no utility for gambling), and also that the subjective probability of complementary events summed to unity i.e. that $(s(p) + s(\bar{p}) = 1$. In the first study, however, acceptance of the hypothesis of no utility for gambling led to rejection of the hypothesis that $s(p) + s(\bar{p}) = 1$ and vice versa. Thus the general SEU model, but not the classical one was accepted.

Slovic and Lichenstein (1968a)) and Andriennsen (1971) conducted studies examining bids and ratings for duplex gambles. They considered an additive model,

$R = c_0 + c_1 PW + c_2 SW + c_3 PL + c_4 SL$ where the c_i's are individual parameters, PW, SW, PL and SL are the gambles' parameters and R represents the bidding or rating response. Results of both studies indicated a good fit of this model, but unfortunately, it was compared mainly to the EV model, (it was considerably superior to this), not to more plausible alternatives. Evidence of whether models including interaction terms were significantly better than the above model was discussed by Slovic & Lichenstein. They carried out a pilot study, which, they said, showed that more complex models did not fit better than the above. No details are given, however and it is not clear whether this test was of group or individual data. They conclude that most variance not due to the model can be ascribed to error, though they do not say how they arrive at this. Andriennsen's study investigates whether the above model can be extended to other situations, in particular to a "skill" situation. He concludes, among other things that in many cases the model is no better than EV and that responses are largely determined by situational factors not related to the gamble's parameters.

In a study of ratings for "offers to gamble", mainly of lotteries which cost a fixed amount, Sjoberg (1968) studied the general SEU model using a technique based on factor analysis that he developed in an earlier paper, Sjoberg (1966). He found that for most people the model accounted well for the data. For some subjects a simpler additive model similar to Slovic & Lichenstein's fitted the responses but for most of them the general model,SEU was necessary. The study by Anderson & Shanteau (1970) discussed earlier also found evidence that probabilities and payoffs were integrated multiplicatively. They observed ratings for duplex gambles and found that no kind of straight additive model was satisfactory though the SEU information integration model was broadly successful. There were significant interactions, however, which followed no particular pattern. These could not be accounted for by the model.

These studies indicate that in many cases a simple additive model can account quite well for the way people make ratings and bids for gambles. Where it has been compared to the general SEU model, however, the latter has generally been superior. In simple riskless situations, where the number of dimensions of the alternatives is not too large the additive model has been found adequate, and other models have not led to significantly better fits (Tversky 1967a), Shanteau and Anderson (1969).

It can be noted at this stage that the finding that people reverse their preferences in bidding and choosing among gambles, Lichenstein & Slovic (1971),Lindman (1971) cannot be accounted for by any utility model of the kind under consideration. They can account for behaviour within the bidding and within the choosing situations but the models which apply in the two situations are different. This obviously delineates an outer bound of the model's domain. However, the only explanation offered by either authors is an "information processing" one which is just as piece-meal as the above suggestion that different models apply in different cases.

Some closely related work on the study of clinical judgements should be noted. A clinical judgement is not a decision as has been understood in the foregoing. That is, it is not a selection from a set of multi-attribute alternatives under the direction of prevailing motivation. However, assessments about multi-attribute alternatives are made in clinical judgements and where these can be interpreted as judgements of preference the two fields overlap. It is not surprising, therefore that information integration models based on the general linear model have been considered extensively in clinical judgement studies. Certain features of the kinds of situations generally considered in clinical judgement studies that are not shared by those which are our current concern are:

i) no information about the costs involved in giving different judgements is explicitly given.

ii) the uncertainties involved are not discussed.

iii) the unreliability of the information source is not discussed.

iv) the number of dimensions that each alternative has is large.

Because of these dissimilarities between the two kinds of situations clinical judgement studies will not be considered in detail. The review by Goldberg (1968) covers the field and shows how information integration models based on the general linear model can account for judgements about complex stimuli made by experts and naive subjects in a variety of situations.

Before leaving functional measurement studies work on a class of models similar to the information integration ones should be discussed. In early years the main competitors to the

SEU and additive models were those based on the suggestion that people are primarily influenced by the alternatives' expected value and secondarily by the higher moments, variance and skewness e.g. Coombs and Pruit (1960). Related to this is the idea that people have specific probability preferences. Much experimental evidence has been adjudged to support these ideas (Edwards (1953, 1954a)b)d)), Coombs and Pruit (1960), Van der Meer (1963), but Slovic and Lichenstein (1968a) point out that probabilities, payoffs and moments have all been confounded with one another so that considerable ambiguity has resulted. Slovic, Lichenstein and their associates have carried out a series of studies observing choices and bids for gambles (mainly duplex gambles) where this confounding of variables was avoided. They have shown, rather convincingly that people base their choices and preferences on the specific information given about chances and amounts of wins and losses (Lichenstein (1965), Slovic and Lichenstein (1968a),b)), Slovic(1969b)), Lichenstein, Slovic and Zinc (1969), Payne and Braunstein (1971)) and not on subtle combinations of them such as the moments. Also they were unable to find any DM whose subjective assessment of any probability dimension appeared to be non-monotonic with objective value. They concluded then, that probability preferences, if they exist are not very powerful determiners of choice.

The demolition of the "moments" and "probability preference" theories is the most conclusive result to date of the study of information integration models. Results about which specific model accounts for the data best is conflicting. A criticism of some of the studies cited is that the model tested for goodness of fit was not compared to other models. The basic criticism of recent quantitative studies of information integration models, however is that they provide evidence about decision making yet decisions are not actually observed. The need to back up this evidence with experiments where actual decision making takes place is one of the main reasons for the experiments to be described.

Sequential Decision Making Studies.

There have been four studies of decision making in sequential and dynamic environments where DMs have had full information. Edwards (1962a)) reviewed the main areas of research in sequential and dynamic situations and only one of these generally included full information. That which he called "static sequential" which seems to include such situations as when a sequence of choices between bets are made which are then played after every choice. This corresponds to what has been defined earlier as the "independent sequential situation". Now, earlier it was suggested that this situation was suitable for the study of the effects: previous outcomes of choices and current status (in the case of gamble choices, current capital). In the independent sequential situation the above factors are not confounded with dynamic ones. Dynamic studies will be considered later but first let us discuss the sequential ones.

Edwards (1962b)) has reviewed his own work in such situations. He found in his probability preference experiments (particularly 1954d)) that "S's can be made to win exceedingly large amounts of money, or be made to lose substantial, but smaller amounts of money, without significantly changing their choices," (1962b)). Lichenstein (1965) examined the effects on bids for gambles of the amount of money previously won or lost in gambling. She found no effect. Greenberg and Weiner (1966), surprised at these findings which, they suggest, are counterintuitive investigated the factors reinforcement history and amount of money possessed in a 3 x 3 factorial experiment. There were about 14 S's assigned to groups representing each factorial combination. They were given an initial stake (at one of 3 levels) and a twenty number bingo card containing nine winning, nine losing and two neutral numbers. An accomplice picked numbered discs, supposedly at random from a bag 9 times. The numbers chosen were rigged to give three levels of reinforcement: 8 wins and 1 loss or, 4 wins, 4 losses and a neutral, or 1 win and 8 losses. It was found that reinforcement history, but not amount of initial capital affected a subsequent choice from 100 offers to gamble.

An earlier study, Dale (1962) found that sequential information affected peoples choices. People chose repeatedly one of three alternative bets in an imaginary setting under the instructions to do as well as possible in the long run. They were standard (x,p,y) bets with expected values -1, 0 and +1 points. Half of the subjects were shown, in addition to the basic information about the bets, the points gained from playing the bets forty times in an earlier session. It was clear that when people were given the additional sequential information they chose that gamble with the highest expectation significantly more often. It is not clear whether the actual sequence of outcomes experienced effected their choices.

Siguel (unpublished doctoral dissertation, 1969) had people choose between pairs of standard gambles under two conditions differing in information given, each subject choosing under each condition. In one condition subjects were reminded prior to each choice, of their previous choice and its outcome and in the other they were also told their cumulative winnings or losses. For all 12 subjects, who made over 2000 choices each in all, very marked choice and choice-outcome sequential effects were observed under both information conditions. Two features of the design may have contributed to this i) response bias was not controlled for and ii) the information given to S's prior to each choice (their last choice and its outcome) may have induced them to repeat their last choice (or alternatively to change it!)

Now, the evidence from all these studies regarding the effects of reinforcement history and current capital is not unanimous. Furthermore, most of them can be criticised on methodological or statistical grounds. Dale's result is not in dispute. It shows clearly that people were affected by information about previous outcomes. But, this is not the same as being affected by the actual experience of these outcomes.

Edwards (1954d)) and Greenberg and Weiner (1966) used games which were rigged. That is, the probabilities of events were not what they were told. Their results only apply, then to situations where DM's information is inaccurate. Since it was inaccurate, it is reasonable to suppose DM's might become aware of this even though they were not told. In the other studies: Edwards (1954a) b) d)), Lichenstein (1965) and Siguel (1969) the information given was representative of the true probabilities and payoffs. The major criticism of these studies, though is that the sequential factors previous outcomes and current capital were confounded. This is not so important to Lichenstein's result as her finding was negative. Edwards and Siguel's results are difficult to interpret. Was it the capital, the reinforcement history or both which affected choices? Other criticisms are: i) as indicated earlier, Siguel's result may be an artefact ii) all the studies used rather insensitive tests of effects due to sequential factors. Clearly then, further experimentation is necessary. Experiments are required in which i) the principle factors are not confounded ii) sensitive tests of changes in choices are employed iii) accurate information is given to DM's iv) the effect of response bias is controlled.

Dynamic Decision Making Studies.

Sequential factors have been examined in dynamic situations in four studies. Cohen et al (1969) gave subjects an initial stake and observed the amount they bet on either a red or a black number of a roulette wheel (they could choose the colour). They found that the previous choice and the previous outcome affected both their choice of colour and their choice of amount to bet. It is to be expected that the effect on choice of colour would diminish in the face of more basic differences between the alternatives. The same kinds of effects on amount bet were found by Hachauer (1970) in a field study of roulette playing behaviour. In both studies it is not possible to say whether previous outcome or current capital were the major determinant of amount bet because these things were confounded. Rapoport and Jones (1970) and Rapoport et al (1970) examined an optimal model for a similar dynamic situation to the roulette playing one. In the first study subjects bet play points on one of two alternatives a fixed but unknown number of times. He could bet as much of his stake on any trial as he wished and he won or lost this amount. In the second study a sequence of bets were made where four possible outcomes might occur. At any trial his capital had to be divided among all four alternatives and he won four times what he had placed on the outcome which occurred. In both of these studies, unlike the previous two, the effects of current capital and previous outcome were examined separately. Both were found to have a positive effect on proportion of capital wagered on the most likely outcome. It is felt that previous outcome had a marked effect because the probabilities involved were not explicitly given (though they were well learned). If the probabilities had been known exactly it is suspected that subjects would not have allowed themselves to be influenced much by previous outcomes. The more striking effect, though was that due to amount of capital possessed at the time of choice. Proportion wagered on the most likely outcome was a u-shaped function of this.

In these four studies in dynamic situations positive effects on amount, or proportion wagered were found due to either, or both of the sequential factors, previous outcome and current capital. This is rather surprising in view of the lack of effect they have been found to have in purely sequential situations. For the moment, let us choose the interpretation that most of the effects of sequential factors were due to current capital. This certainly had the more marked effect in Rapoport and Jones (1971) and Rapoport et al (1971) and since the sequential factors were confounded in the other studies they do not provide evidence against the chosen interpretation. The problem is, then that the effect of current capital on decisions made in dynamic situations is not the same as its effect in sequential ones. This solution to this problem may lie in the following fact.

In all of the above studies when current capital changes so does the chance of arriving at certain future decision states. People very likely modify the amount they bet as a function of the probable future decisions available rather than as a function of current capital itself. The ambiguity of interpretations epitomises previous studies of dynamic decision making. The independent variables are not well controlled and any number of models can "explain" what is going on. Before the foregoing studies could be useful considerable clarification of the meaning of their findings is required.

The only extensive programme of research in dynamic decision making under relatively complete information has been carried out by Rapoport. His approach is to consider quite complex dynamic decision situations that have been studied as business or economic problems and develop or borrow the optimum decision theory for them. These are proposed as initial hypotheses of actual behaviour which are then tested. The two studies mentioned above perhaps provide the most successful of the optimal models that have been considered by Rapoport. Despite the effects of the sequential factors (which optimal models would not predict) the strategy people adopted was surprisingly close to optimal. Other studies by Rapoport have unfortunately suffered rather badly from the confounding of independent variables. In a study of a Markovian decision task involving transitions among three decision states (Rapoport, 1968) the optimal model appeared to be quite successful. However, behaviour could also be explained by the rather simple, far from optimal strategy of making decisions to increase the likelihood of a transition to the most favourable decision state. Three earlier studies considered a situation known as the readers' control problem (Rapoport 1966a), 1966b), 1967b). The readers' control problem is a multistage dynamic decision problem where DM is required to make a series of decisions to minimize a cost function which depends on the state of affairs prior to a decision, the decision made and a chance event. The expected cost is to be minimized over all stages and the optimum strategy is known from dynamic programming theory.

In all three experiments the group mean decisions followed the optimum ones closely but an examination of individual decisions revealed errors which could not be explained in terms of the optimal model or variations of it.

It is felt that rather little progress has been made in the study of dynamic decision making because the situations studied have been complex. The power of the experimental method, to control situational factors and thereby strip away as much ambiguity as possible, has not been fully utilized. It was indicated in the introduction that progress in the study of decision making under uncertainty would be painstakingly slow. Unambiguous experiments in simple situations are here preferred.

Chapter 5.

Static Experiments.

Qualitative and quantitative techniques for examining information integration models in static situations have been described. The models can be examined qualitatively using statements of preference and binary choices and quantitatively using the latter. The techniques will be used to compare different models in static situations where DM's are presented with pairs of simple or duplex gambles. The sub-goals within this broad objective will now be introduced, first with regard to QFA and then functional measurement.

The alternatives presented in the experiments will be simple gambles and duplex gambles where the parameter PL remains fixed at PL = 0.5. The information integration model of main interest as far as QFA is concerned is the SEU model. This is a special case of the multiplicative representation hypothesis, H5 when the alternatives are simple gambles and it is a special case of the dual-distributive representation hypothesis, H9 when the alternatives are duplex gambles. To test SEU, therefore one could examine the complete set of appropriate conditions i.e. H5C1 - H5C5 for the simple gambles and H9C1 - H9C4 for the duplex.

First though, these should be examined logically. The zero conditions are rather trivial. It can be assumed for every subject that the scale values associated with probabilities and payoffs of zero are themselves zero. For these elements the conditions on zeros can easily be shown to be satisfied. It is not expected that other zeros would exist so until the contrary is demonstrated the zero conditions can be disregarded. Two other rather trivial assumptions about peoples choices can be made. The first is reasonable when the motivating variable is money while the second is more general and is hypothesized in the SEU model (see chapter 2). They are as follows:

i) In the general definition of SEU, if each x_i, $y_i \in X_i$ is a monetary quantity $u_i(x_i) \leqslant u_i(y_i) \Leftrightarrow x_i \leqslant y_i$ for i = 1, ..., k.

ii) For each p_i, $q_i \in P_i$, $s_i(p_i) \leqslant s_i(q_i) \Leftrightarrow p_i \leqslant q_i$ for i = 1, 2,....k

It can easily be shown for simple and duplex gambles that the sign dependence/independence conditions predicted by SEU follow from these two assumptions, which amount to what has been called the "sure thing principle". Thus, if one is prepared to accept the sure thing principle (see Edwards, 1954c)) it is only necessary to test experimentally the weak order and cancellation conditions. Similarly if one was interested in the inexact SEU model for simple gambles the empirical task would be to test the semi-order and cancellation of strict preferences conditions. Accepting the sure thing principle has a further consequence: for simple and duplex gambles all simple polynomial functional relationships other than the SEU model can be rejected. This is because the sure thing principle violates the zero, independence and sign dependence consequences of all non-SEU models.

Although these conclusions seem reasonable, a word of caution should be expressed about rejecting hypotheses because trivial predictions are violated. This can be illustrated with an empirical study made by Coombs and Huang (1970). They carried out an ordinal analysis of peoples' perceived risk of certain gambles. The gambles were of the wheel of fortune type of the form $(x, p = \frac{1}{2}, y)$ where an amount x was won with probability $\frac{1}{2}$ and otherwise an amount y was lost with probability $\frac{1}{2}$. Such gambles are completely characterized by their expected value, $E = (x + y)/2$ their range, $R = x - y$ and the probability of a win, $P = \frac{1}{2}$. A further factor, the number of plays of a gamble, C can be included. Gambles from the set E x R x C can be examined with respect to their riskiness. Coombs and Huang (1970) considered the distributive simple polynomial model, i.e. the model which says that the percieved risk of the whole gamble, PR is a distributive function of the percieved risk of the components, E, C and R: $PR = (b_1(E) + b_2(R))\, b_3(C)$. The ordinal analysis they performed supports this model. Krantz and Tversky, (1971) in discussing the model state that support also can be found from the effective zero classes which exist. However, it is easily demonstrated that this is not so. Any gamble with a range of zero must have a percieved risk of zero, regardless of its expected value or the number of times it is played. This, as can be seen from the results on zeros reported earlier, must be incompatible with the above distributive model. The wisdom of rejecting the model on the basis of this should be questioned, though, as 'gambles' with a zero-range are clearly trivial. If the model is shown to be satisfactory in the domain of non-trivial gambles much stronger reasons for rejecting it are necessary. This may also be true in considering peoples' preferences for gambles. Violations of results about zeros predicted by some model would not be sufficient grounds for rejecting it. While the diagnostic value of zero results is not in dispute it should be recognized that real tests of a model must be based on its performance in non-trivial cases.

The sure thing principle unlike the zero conditions applies to quite a wide range of choice situations. It is not concerned with only a few trivial alternatives. Any models which demonstrably violate it, therefore, need not be considered further. In the experiments of this chapter then, QFA will be used to examine the ordering and cancellation consequences of the exact SEU model only. Two previous experiments involving simple gambles, Coombs, Bezembinder & Goode (1967) and Tversky (1969) have addressed themselves to the triple cancellation and transitivity of strict preference conditions respectively. No previous study, however, has focussed on cancellation and transitivity of indifference. This is considered an omission which should be rectified since the exact SEU model makes very strong predictions about indifference. If violations are found which cannot be put down to error then the inexact SEU model must be accepted as the more plausible alternative. In this chapter, this comparison between exact and inexact SEU models is made when simple gambles are considered but no particular consequences of SEU are emphasized in the case of duplex gambles. Here a general QFA is attempted since no other study using such alternatives is known. If the general QFA shows that the SEU model is satisfied more rigorous tests of particular conditions should follow.

As well as these substantive problems some methodological questions will be raised. In chapter 2 it was noted that the statistical tests associated with QFA have not been fully worked out. Therefore it was necessary to attempt certain ad hoc solutions, whose success is also under investigation.

Previous studies investigating functional relationships quantitatively have considered additive and SEU models. For both simple and duplex gambles this will be done in the experiments of this chapter. As well as this it will be necessary to examine the extent of bias in estimates of subjective value for the particular designs used.

In experiment 1 all the procedures previously discussed will be used to examine the statements of preference and binary choices of a single DM who is confronted with simple gambles. Because data from only one individual has been collected this is to be regarded primarily as a feasibility study of the techniques.

In experiments 2 and 3 more substantial amounts of data are collected: statements of preference for simple gambles in experiment 2 and binary choices for duplex gambles in experiment 3. QFA is applied in experiment 2 and both QFA and functional measurement are carried out in experiment 3.

Experiment 1.

The examination by QFA and functional measurement of a single subjects binary choices and statements of preference for simple gambles.

The aims of this experiment were, for the single subject:

i) to test the SEU model and probabilistic model 1 for binary choice data by QFA,

ii) to test the SEU model and the simple error model for statements of preference by QFA,

iii) to compare additive and SEU models by Bock and Jones' method of functional measurement using the binary choice data.

In any empirical test of the consequences of a representation hypothesis the pairs of alternatives used is crucial. In this experiment it was decided to consider a 4 x 3 factorial set of amount to win x probability gambles and a similar set of lose ones. In order to emphasize the consequences of the SEU model for indifference, small differences in probability and payoffs were selected. If i) and ii) show that the SEU model is successful this is despite the fact that the gamble sets were such that transitivity and double cancellation of indifference could be tested. This selection of gambles close in subjective value also given an advantage with respect to iii) above, in that it is likely to reduce the chance of observing zero or one preference frequencies. A limitation of the experiment, though is that subjective scales of only a small range of simple gambles can be obtained.

Experimental Design.

A win set and a lose set of simple gambles were constructed. Let the win set be P x SW and the lose set P x SL, where $P = \{0.40, 0.45, 0.55, 0.60\}$, $SW = \{+75p, +80p, +85p\}$ and $SL = \{-85p, -80p, -75p\}$. A gamble from the win set is one where $sw_i \in SW$ is won with probability $p_j \in P$ and nothing is won with probability $(1-p_j)$. Similarly, in a gamble from the lose set, $sl_i \in SL$ is lost with probability $p_j \in P$. All pairs from the win set and all pairs from the lose set such that neither dominates the other are presented to the subject (S). This gives 18 win pairs and 18 lose pairs.

In the first part of the experiment binary choice data was collected. Each of the 36 pairs were presented to the subject 7 times. The juxtaposition of gambles in a pair, with respect to left-right was random. The order of the 252 trials was also random except that if the pair had appeared during the previous 10 trials it would not appear at the present trial. The responses available to the subject were either 'chose gamble one' or 'chose gamble two'. No indifference judgement was allowed.

In the second part of the experiment statements of preference were obtained. The 36 gamble pairs were presented in random order and on each trial S stated which he preferred or whether he had no preference. The order of presentation of the pairs and the juxtaposition of gambles in a pair were both random.

Subject

The Subject was a male research assistant in the Department of Psychology, University of Stirling aged 24.

Procedure

For the binary choice part, a stimulus tape was prepared for a standard ICL teleprinter. After the instructions were given S ran the tape through the reader until the first pair were printed in the following form:

	Choice One	Choice two.
Amount to lose *	75p	80p
Chance	.60	.55

Choice ....

* Win appears here on win pairs.

After the word 'choice' there was some runout on the input tape and S stopped the reader. He then punched either '1' or '2) depending on his choice and restarted the reader, until the next pair had been typed on the teleprinter printout. The operation was repeated 252 times until the experiment was completed. During the experiment, the teleprinter punch was switched on giving a paper tape record of the stimuli and responses. This took about 2 hrs. 15 mins.

The subject, before he began, was told that at the end of the experiment, one trial would be picked at random and he would be required to play the gamble chosen on that trial. He was to be given a stake of 75p before this, so he stood to lose only a few pence and to gain up to 160 p over all. This was intended to give some realism to his choices and provide motivation for the tedious task of watching the teleprinter print. He actually won 130p.

Approximately two weeks later, the 36 gamble pairs, randomized for left-right position were presented to the subject in booklet form. The booklet was made up from teleprinter printout with the gamble pairs written in the same form as for the first part of the experiment. This time, S was asked simply to indicate, with an imaginary choice situation in view, which he preferred or if he had no preference. He wrote one, two or a dash accordingly and went through the booklet page by page in about fifteen minutes.

Results and Discussion

The raw data for the binary choice situation is shown in the final columns of tables 5.1 and 5.2. These columns give the number of times the subject chose gamble two of each pair. The raw data for the statements of preference can be derived from the "stated preference" columns of the tables.

First, a result was obtained pointing to the validity of the proposed analysis of the binary choice data in terms of the binary preferences probabilities. The following two factors may make the analysis invalid:

i) subjects remember their previous choice on a particular pair and tend to repeat it.

ii) they change their minds about some preferences half way through the experiment.

Suppose the raw data is put in the form of 36 binary sequences of 1s and 2s, one for each pair presented. A 1 or 2 at the jth position of the i^{th} sequence indicates that the subject made this choice at the jth presentation of the i^{th} pair. If either of the invalidating factors does have an effect then the number of runs in this data matrix will differ markedly from the number expected under the null hypothesis of no effect. For the data obtained, the expected number of runs, $E(R) = 76$, the standard deviation, $SD(R) = 3.5$ and the observed number of runs, $R = 75$. These figures were obtained by calculating R, $E(R)$ and Var. $(R) = \{SD(R)\}^2$ for each sequence and summing over all sequences. It can be concluded from the above figures that the two factors are not likely to have affected choices. The basic Bernoulli model, which says that choices between any pair are determined by a constant preference probability is not disconfirmed. Further analyses based on the binary preference frequencies shown in figures 5.1 and 5.2 can therefore be carried out.

Qualitative Functional Analysis

For the choice data the unrestricted maximum likelihood preference pattern under probabilistic model I can be found. This pattern, and also the pattern of observed statements of preference can be augmented by the assumed preferences among pairs in which one gamble dominates the other. This defines, for each kind of data, a strict preference relation over the set of simple gambles, with an indifference relation as its symmetric complement. One can test both relations to see if the qualitative properties predicted by the SEU model are violated. From chapter 2 it is known that to do this it is sufficient to test conditions H5C2 and H5C3', the weak order and cancellation conditions respectively. This test showed that a small number of violations of the conditions occurred. The question was posed as to what is the minimum number of preferences that must be changed in order to eliminate these violations. If the pattern of preferences most consistent with the data does not satisfy SEU, perhaps one very similar to it does. Let the k value of a preference pattern be the number of preferences different from the pattern that best fits the data. The following algorithm was used to try to find some pattern consistant with SEU having the minimum k value. It was applied to both kinds of data.

i) Take the pattern of preference most consistent with the data

ii) Find the relation that takes part in most violations of SEU

iii) change it to the value which reduces the total number of violations most

iv) repeat the last two operations until no violations of SEU occur.

This algorithm worked to give a value of k = 3 for the statements and k = 1 for choices. The preference patterns obtained are shown in tables 5.1 and 5.2. They can only be regarded as estimates of the min k patterns as nothing is known about the algorithm which selected them. It probably worked because the violations of the initial pattern were few. The results of the algorithm can be regarded as estimates of the maximum likelihood patterns consistent with SEU assuming the relevant probabilistic model. Assuming these are reasonable estimates, the SEU model can be tested by the likelihood ratio test against the general "point probability" models. The chi-square statistics used in these tests, with their degrees of freedom are shown in table 5.3.

Also in table 5.3 the SEU model is compared to two other information integration models: the expected value (EV) hypothesis and a "level of advantages" hypothesis. The level hypothesis is as follows:

i) label the most advantageous probability 3, the next 2 and the least advantageous 1.

ii) similarly label the values componants in rank order

iii) define the level of advantageousness of a gamble as the sum of its probability and value ranks.

iv) define the level hypothesis as that which predicts S is indifferent between gambles of the same level and otherwise predicts that he prefers the gamble with the higher level.

It can be seen from table 5.3 that by the LR test of choice data, level and EV can be rejected at the .001 level but SEU cannot be rejected even at the 0.05 level. The comparison of the three models with respect to the statements data can be made by comparison of min k values as well as the three LR tests. Both comparisons show that SEU is superior to the simple information integration models, though EV and level cannot be rejected at the .05 level. These findings are not surprising as SEU defines a class of models, and the one tested was chosen for its goodness of fit while the other two are particular models, each defining a single preference pattern. However it does show that the analysis has led to an improvement of goodness of fit over two quite plausible simple models. The evidence so far points to the conclusion that the SEU model can account for the subjects' choices and it can also account quite well for the subjects' statements of preference when errors are defined as above. Neither choices nor statements contradicted the consequences of SEU concerning indifference, because indifference did not play a prominent role in the subjects responding. Unfortunately, few tests of the conditions could be made so few violations could be observed.

Thus, QFA was not applied to the inexact SEU model.

Functional Measurement Analysis.

Strictly speaking, Bock and Jones' analysis requires non-zero estimates of binary preference probabilities and a large n in an n-replicate pair comparison experiment. However, n was quite small (7) and it can be seen from tables 5.1 and 5.2 that many (21 out of 36) preference frequencies were zero or one. The analysis will be carried out under the assumption that despite these facts, the asymptotic results hold to a reasonable approximation. The implications of this will be considered later.

The analysis was carried out, using only the observations actually made, within each gamble set. Thus, goodness of fit statistics were calculated and parameters estimated from two incomplete, 7 - replicate pair comparison experiments. The observations not made could have been assumed and the analysis could have been based on complete pair comparison results. However this would have led to a complete domination of zero and one preference frequencies, which was not necessary because the actual observations were sufficient for the analysis. A consequence of using the incomplete design was that the extreme favorable and unfavorable gambles of each set were not included. Thus an interval scale of only ten of the gambles in each set was estimated. Further to this, minimum normit chi square estimation of the parameters of certain sub-models and tests of their goodness of fit were calculated.

Table 5.4 gives chi-square statistics (some obtained from differences of chi-squares) of goodness of fit for the basic Thurstone model and two additive multi-factor sub-models. The significant subjective values chi-squares lead us to reject the hypothesis that the subjective values of the gambles are all equal (within each set). If any "departure" statistic is significant one can conclude that the data deviates from the model under test more than would be expected by chance. From the table it can be seen that the basic Thurstonian model can be accepted and one can consider the various information integration sub-models. The table shows the chi-square for departure from the general additive model and the additive model discussed by Slovic. This latter is where a gamble's subjective worth is a weighted sum of its probability and value components. Thus, the two parameters of the model are the two weights. The data from both the win and the lose set does not depart significantly from either of the additive models. The difference between the "departure" chi-square statistics is not too great, considering the difference in their degrees of freedom (particularly for the win set). Therefore it is reasonable to accept Slovic's more simple model. Particularly as this only has two parameters and predictions can be made based on the objective values of the gambles. In figure 5.1 and 5.2 the subjective values predicted by this model are compared to those estimated directly from the data. These confirm that the model fits quite well. The parameter estimation for the model is shown in table 5.5.

Slovic's model can be considered a special case of the linear SEU model. In the linear SEU model the subjective values are linear functions of the objective values, and are assumed to integrate multiplicatively. Thus,

$$u(g) = (a + b.p)(c + d.x) = ac + bcp + ad\ x + bdpx$$

where $u(g)$ is the utility of the gamble, a,b,c,d are parameters of the model and p and x are the objective probability and value components of the gamble. In a pair comparison situation the constant, ac cancels so there are effectively only three parameters, bc, ad and bd. Slovic's model is the special case where bd = 0. Bock and Jones' methods can be used to test the fit of the linear SEU model and also the specific hypothesis that bd = 0. For both sets of gambles the hypothesis bd = 0 could not be rejected at the 5% level of significance.

Discussion.

The conclusions from the qualitative and quantitative analyses of the choice data are consistent with one another. Within the win and the lose set of gambles the QFA of the SEU and additive hypotheses is the same. Thus, acceptance of one implies acceptance of the other. Because sign dependence with respect to the sign of the outcomes manifestly held, though it is reasonable to call the QFA a test of SEU. It should be noted that the QFA analyses of the choices and statements are not really comparable, even though they were obtained from the same DM. Six of the eighteen "min-k" preferences were different in each set. Differences are to be expected because i) the statements were hyperthetical and the choices were not ii) the different data were collected two weeks apart. The QFA does not really discriminate between the general SEU model and one which proposes the two components are additive within each set.

To do this functional measurement is necessary, since if either the additive or the multiplicative integration rule holds within sets the QFA will not be contradicted.

The expected test by QFA of the consequences referring to indifference did not materialize. One cannot conclude that a thorough test of indifference has been carried out, despite the gambles being chosen to do so. For this reason statements of preference for the gambles used in this experiment will be collected from a group of subjects in experiment 2. Choices will not be collected because the algorithm available for applying the QFA is unreliable.

The functional measurement supported Slovic's additive model. It may be invalid, however, because of the factors discussed in chapter 2. Additional bias in parameter estimates may be caused by the arbitary nature of the incomplete design. The scale depends on the specific pairs used and random fluctuations in a few observations may affect the scale considerably. Sampling experiments should be carried out to test this source of bias. As a considerable amount of effort would be involved, this has been left till the results of later experiments are discussed.

The overall conclusion of the experiment is in favour of the hypothesis that Slovic's additive model holds within win and lose sets. It remains to be seen whether the same applies in other experiments, where more data is collected.

TABLE 5.1.

Win Gamble Pairs and Minimum-K Preferences

Pair No.	Gamble One Amount	Gamble One Chance	Gamble Two Amount	Gamble Two Chance	Minimum-K Preferences: Stated Preference Data	Minimum-K Preferences: Binary Choice Data	Raw† Choice Data
1	85	.55	80	.60	2**	0	3
2	85	.55	75	.60	1	1	2
3	85	.45	80	.60	2	2	7
4	85	.45	80	.55	2	2	7
5	85	.45	75	.60	2	2	6
6	85	.45	75	.55	2	2	5
7	85	.40	80	.60	2	2	7
8	85	.40	80	.55	2	2	7
9	85	.40	80	.45	0	2	5
10	85	.40	75	.60	2	2	6
11	85	.40	75	.55	2	2	5
12	85	.40	75	.45	1 (2)*	1	0
13	80	.55	75	.60	2	0	4
14	80	.45	75	.60	2	2	7
15	80	.45	75	.55	2	2	7
16	80	.40	75	.60	2	2	7
17	80	.40	75	.55	1	2	7
18	80	.40	75	.45	2	0	4

* The numbers in brackets show the preferences used as initial values for the min k algorithm. Where there is no numbers in brackets the final and initial values were the same.

† This gives the number of times (out of 7) gamble two was chosen.

** 1 and 2 shows whether gamble 1 or 2 was preferred, 0 indicates indifference.

TABLE 5.2.

Lose Gamble Pairs and Minimum K Preferences.*

Pair No.	Gamble One		Gamble Two		Minimum-K Preferences		Raw Choice Data
	Amount	Chance	Amount	Chance	Stated Preference Data	Binary Choice Data	
19	75	.45	80	.40	0	1	0
20	75	.45	85	.40	2 (1)	1	0
21	75	.55	80	.40	2	2	7
22	75	.55	80	.45	2	2	7
23	75	.55	85	.40	0	2	6
24	75	.55	85	.45	0	1 (0)	4
25	75	.60	80	.40	2	2	7
26	75	.60	80	.45	2	2	7
27	75	.60	80	.55	0	1	2
28	75	.60	85	.40	2	2	7
29	75	.60	85	.45	2	2	6
30	75	.60	85	.55	1	1	0
31	80	.45	85	.40	0	1	1
32	80	.55	85	.40	2	2	7
33	80	.55	85	.45	2 (0)	2	7
34	80	.60	85	.40	2	2	7
35	80	.60	85	.45	2	2	7
36	80	.60	85	.45	0	0	3

*All the notation in this table is the same as in table 5.1.

TABLE 5.3.

Likelihood Ratio Tests of Hypotheses 'level', EV and SEU against the unrestricted hypothesis.

Hypothesis under test	Stated Preference Data			Binary Choice Data	
	K value	Chi^2	d.f.	Chi^2	d.f.
EV	11	44.0	35	290.4***	26
LEVEL	12	45.8	35	236.1***	29
SEU*	3	19.9	35	13.1	16

* the values given are those for the min-k, SEU consistent preferences of Tables 5.1 and 5.2.

*** these values are significant at the .1% level.

TABLE 5.4.

Goodness of Fit Tests for the Functional Measurement.

Source of Variation	Win Set Chi-square	d.f.	Lose Set Chi-square	d.f.
Subjective Values	24.68**	9	36.67**	9
Departure from Thurstone model	3.32	9	1.34	9
Total	28.00	18	38.01	18
Departure from General Additive	2.57	4	3.55	4
Departure from Slovic's additive	3.37	7	10.45	7

** Significant at the 1% level.

TABLE 5.5.

The Parameter Estimation for Slovic's Model.

Parameter	Win Set Estimate	Win Set Standard Error	Lose Set Estimate	Lose Set Standard Error
Probability Weight	0.151	0.0338	-0.191	0.0372
Value Weight	0.142	0.0433	0.230	0.0503

Figure 5.1.

~~G~~raph of subjective values estimated directly from the data against those predicted by Slovic's model.

WIN SET

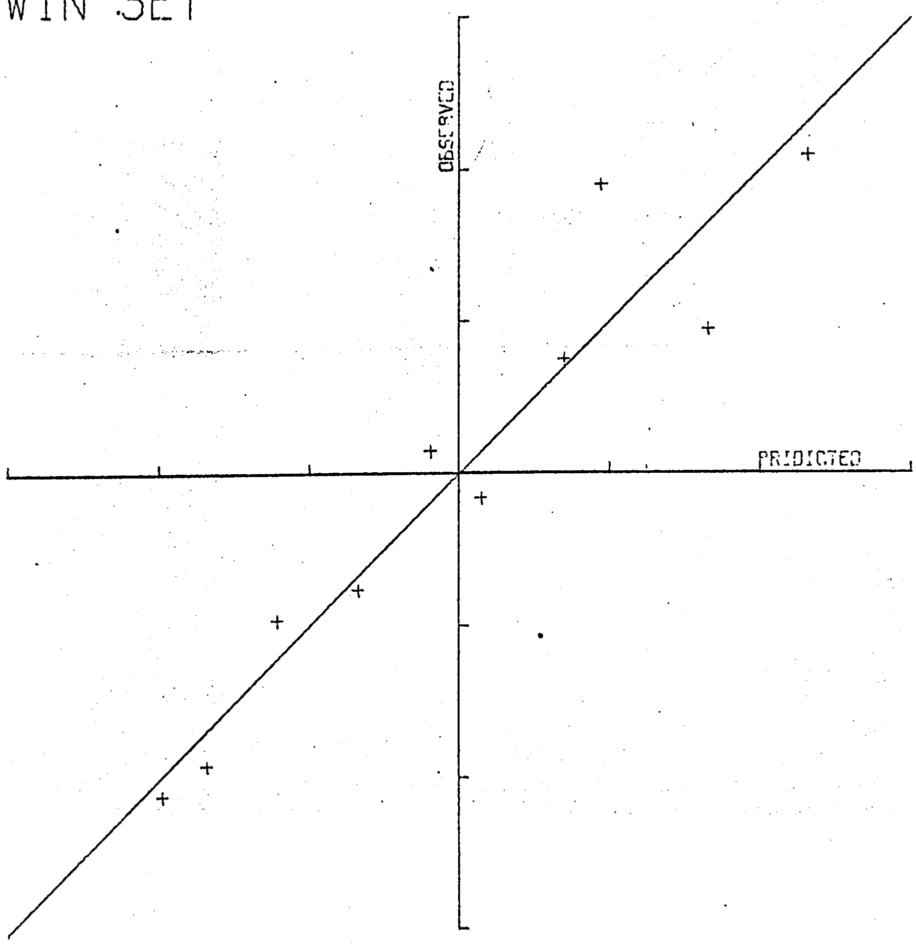
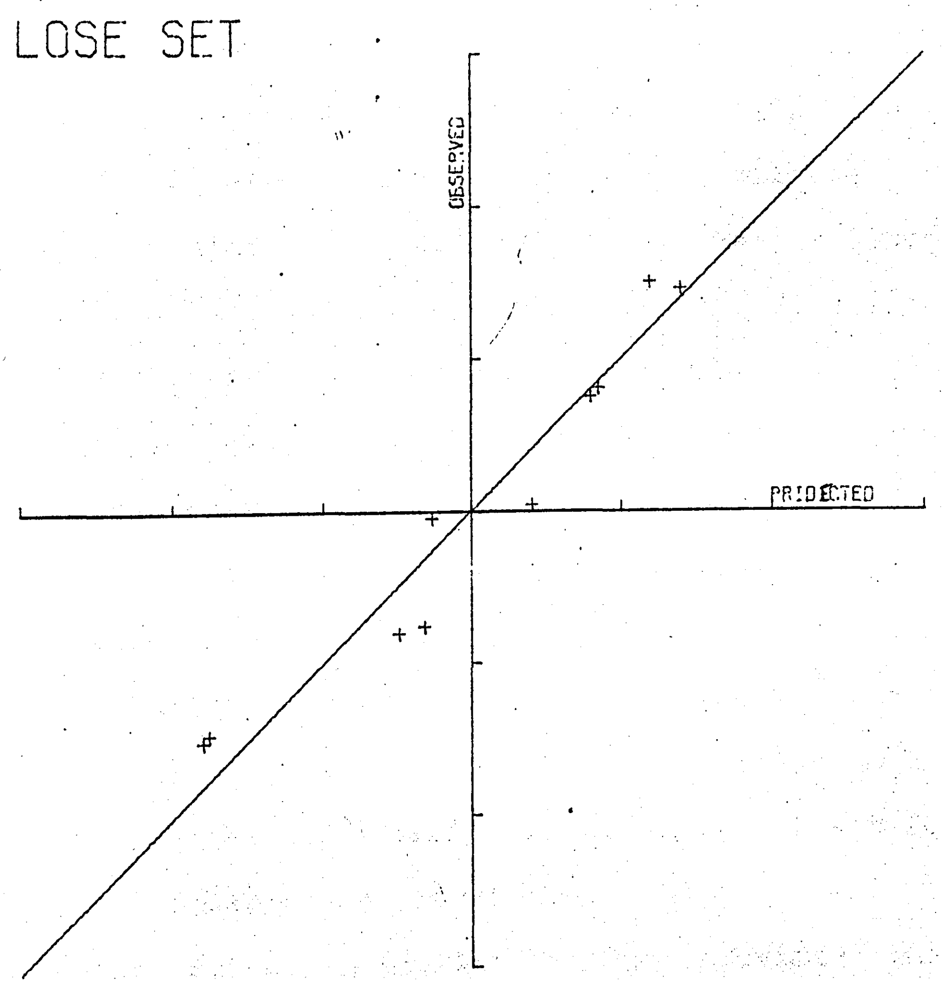


Figure 5.2.

Graph of subjective values estimated directly from the data against those predicted by Slovic's model.



Experiment 2.

A Comparison of the Exact and Inexact SEU models by QFA Applied to Statements of Preference for Simple Gambles.

The exact SEU model received support from both the binary choice and statements of preference data in experiment 1. Not the algebraic, exact SEU model, that is, but one which postulates that errors in statements of preference occur. For the subject observed, "indifference" played only a minor role in both statements and choices. There was insufficient data to examine whether indifferences had the structure predicted by the substantive part of the exact SEU model. The model says that if a person is truly indifferent between certain alternatives then these indifferences should be in accordance with the double cancellation condition and be transitive. The likelihood of this was questioned in chapter 2 and the inexact additive and multiplicative representation hypotheses were proposed as alternatives to the exact models. These algebraic inexact models predict that strict preferences have a well defined structure but that indifferences do not conform to any laws such as the above. Under the exact models, people have a very clear idea of when alternatives have the same value and when they are different. Under the inexact model, alternatives must be further apart in subjective value than some critical value before people are prepared to express a preference.

The motivation for the present experiment was to compare the exact and inexact SEU models using the QFA results presented in chapter 2. The simple gambles of experiment I are ideal for this because there are only small steps along each dimension so that differences in subjective values are likely to be small. It was decided to take a direct approach and obtain a complete set of statements of preference from a group of subjects and simply count the number of tests and violations of the conditions for QFA previously presented.

It was noted in chapter 4 that the implications for indifference of exact representation hypotheses have rarely been recognized. Little is known about the role of indifference in decision-making. The analogy between preference scales and, say, loudness scales may lead one to expect to find a "just noticeable difference" for preference. However, when the alternatives are clearly discriminable processes underlying discrimination and preference will be entirely different. Therefore, in view of the lack of direct experimental evidence there is no a priori reason to expect one result rather than the other.

The basic question this experiment sets out to answer, then, is whether the exact algebraic SEU model has similar substantive relevance for statements of indifference as it does for statements of strict preference. If it does, well and good, but if not, two explanations could account for the lack of structure of indifference: i) statements of indifference mainly occur at random and do not reflect "true" indifference ii) the subjective difference in worth between two gambles must be greater than some threshold before one is considered preferable. It would be difficult to discern which explanation is correct if the data is fallible. This will be discussed further in the results section.

Design and Procedure.

The experiment was a replication of the statements of preference part of experiment I except that a group of ten subjects took part. Each was given a booklet containing the pairs of gambles, the presentation order and juxtaposition of gambles in a pair being randomised differently for each subject.

The following instructions were given to the group, all first year undergraduates at Hull University who had volunteered to take part.

"This is an experiment, about peoples' preferences for simple gambles involving small amounts of money. The gambles can be represented by a wheel of fortune (which was drawn) with a spinner. To play, the spinner is spun and if it lands in the upper sector an amount x is won, if it lands in the lower sector nothing is won. The size of the upper sector indicates the chance of winning.

You will be shown pairs of these gambles and simply asked which one you would prefer to play if you had to play one of them. If you have no preference then you should say so. Here is a typical pair (show). The wheels of fortune are not actually drawn, but the chance of winning is indicated by a decimal. Next to the word 'choice' you should write '1' or '2' according to which you prefer, and if you have no preference write a dash.

As well as these 'win' gambles there are pairs of 'lose' gambles. Here is such a pair. The question with these is: which would you dislike playing least, if you had to play one of them? Again, write '1' or '2' or '-' for no preference. I will show each of you 36 pairs, half of them win pairs, half of them lose pairs. They're all mixed in together so when you come to the next pair, first check to see which type it is. Just work through them in your own time, do not go back to any you did earlier at any stage."

The subjects followed the instructions through and indicated their preferences in, on average, fifteen minutes.

Results and Discussion.

The raw data for the group showed that there were quite wide individual differences in the tendency to use the indifference category. Three subjects used it not at all and one subject used it fifteen times out of thirty six. The median number of times it was used was four.

The basic analysis was simply to count the number of tests and the number of violations of the ordering and cancellation conditions predicted by the exact and inexact SEU models within the win and the lose sets. The group results, only are reported (see table 5.6) since the proportions of violations of any condition were similar for all subjects. It can be seen that the data did not allow many tests of the conditions involving indifference and such tests as there were came from only four of the subjects. This is unfortunate since these are the only conditions predicted by the exact SEU model which are not predicted by the inexact model. All those conditions common to both models could be tested many times from each subjects' results, however.

Taking the group as a whole, it can be seen that the conditions involving strict preference were violated a small percentage of times while those relating to indifference were violated on nearly all the occasions on which they could be tested. Of those conditions involving the strict preferences the double cancellation had the greatest percentage of violations, then transitivity, then the semi-order condition and finally the interval order condition. This order, perhaps, conforms to expectations. Double cancellation implies that people can partially order differences, or ratios of elements along the dimensions while transitivity implies that they can partially order the alternatives themselves. One would not expect that they could do the former better than the latter. The interval order and semi-order conditions can only be violated if the indifference category is used, and so the large number of tests together with the small number of times the indifference category was used accounts for the overall low percentage violations of these conditions.

One can conclude from these four conditions that the inexact SEU model was broadly satisfied. To draw the same conclusion for the exact SEU model it would be necessary to observe low percentage violations of all six conditions tested. Since this is not the case it can be said that the exact SEU model has no substantive relevance for statements of indifference.

It was indicated earlier that one way to follow this up with a more positive conclusion would be to consider two explanations for the structureless nature of indifference: i) indifference plays a minor role in the information integration and statements of indifference occur at random and ii) indifferences occur when the alternatives are closer together in subjective value than some critical interval. To discriminate between the two it would be necessary to construct a (preferably interval) scale of the alternatives and observe if the indifferences were random, regardless of scale proximity. The experiment was not designed to do this. A partial scale could be constructed in some cases, but because of the fallible data the "best" scale would generally not be unique.

Whether the exact SEU model, together with the simple error model could account for the percentage violations found in the group was examined in a small sampling experiment. Monte Carlo data was generated from "statistical DM's" who behaved exactly according to the model, with different values of the single parameter, α. The parameter, α was the probability of an erroneous statement. The "true" preferences and indifferences were set up from random subjective value scales of probability and value, both monotone increasing. A data matrix of pairwise statements of preference and indifference was set up as follows: i) if the sure thing principle made a prediction the statement was taken to be in the predicted direction ii) otherwise, a random number, R, was selected between zero and one a) if $R \geqslant \alpha$ the statement predicted from the subjective value scales and a multiplicative information integration rule was assumed b) if $R < \alpha$ one of the two statements not predicted as in a) was picked at random. In any run, the percentage violations and number of tests of the transitivity and double cancellation of indifference and strict preference were calculated. Since, with this method of generation, the number of "true" indifferences is small (zero in theory) it is manifestly clear that the indifferences play a minor role and occur mainly at random, as hypothesized. Forty simulations for each value of α starting at 0.15 in steps of .05 to 0.40 were carried out. The results are summarized in table 5.7. This gives the mean and variance of the percentage violations of each of the four conditions. It also shows in how many simulations it was possible to test the conditions, and of these what the mean number of tests was. A value of α between 0.25 and 0.30 would appear to approximate the group results satisfactorily. Therefore, it is confirmed that the exact SEU model with simple error could account for the results summarized in table 5.6. One would expect the same to be true for the inexact SEU model with simple error, probably with a lower fitted α.

It has not been possible to reject either of the possible explanations of the results, even though the predictions about indifference were not remotely born out. The notion of fallible data had to be introduced to explain the violations of the strict preference conditions. If one accepts that the data is fallible it makes very little difference whether the exact or inexact model is accepted because of the minor role of indifference. Since indifference does not seem to be too important it will not be investigated further. The experiment has drawn attention to the rather strong predictions about indifference made by the exact SEU model and shown that they do not hold. The sampling experiment puts the finding in perspective. Indifference is a relatively rare phenomenon in the context of simple gambles and it does not seem that the exact SEU models' lack of predictive usefulness with respect to indifference damages its general usefulness to any great extent.

TABLE 5.6.

Violations of the Testable Conditions Considered in Experiment 2.

Property	Percentage of Violations	Number of Violations	Number of Tests	Number of Subjects
Transitivity of Indifference	100.0%	28	28	4
Double Cancellation of Indifference	91.5%	32	35	4
Transitivity of Strict Preference	4.5%	153	3350	10
Double Cancellation of Strict Preference	10.2%	157	1544	10
Semi-order Condition	1.4%	429	30459	10
Interval-order Condition	0.9%	220	23662	10

Table 5.7.

Sampling Experiment. Summary of Results from 40 Simulations at Each Value of Alpha.

Alpha	Transitivity Condition							
	Indifference Relation				Strict Preference Relation			
	Percent Violations		No. of Tests		Percent Violations		No. of Tests	
	Mean	Standard Dev.	Mean	N*	Mean	Standard Dev.	Mean	N*
0.15	100	-	2.11	9	4.07	3.77	424.6	40
0.20	100	-	1.73	15	5.62	3.77	423.1	40
0.25	100	-	2.19	26	7.19	4.22	408.1	40
0.30	93.62	23.67	2.12	26	6.99	3.87	409.2	40
0.35	95.00	21.89	2.50	20	7.45	4.28	410.6	40
0.40	100	-	2.78	32	10.61	4.50	404.9	40

Alpha	Double Cancellation Condition							
	Indifference Relation				Strict Preference Relation			
	Percent Violations		No. of Tests		Percent Violations		No. of Tests	
	Mean	Standard Dev.	Mean	N*	Mean	Standard Dev.	Mean	N*
0.15	100	-	1.88	9	5.52	3.65	359.6	40
0.20	100	-	1.92	13	7.44	3.91	359.8	40
0.25	100	-	2.06	16	7.77	3.26	354.5	40
0.30	96.82	11.22	2.96	27	9.39	4.14	342.3	40
0.35	100	-	2.09	23	9.21	3.81	351.7	40
0.40	100	-	3.90	30	10.29	4.08	336.5	40

*N is the number of simulations in which at least one test of the condition was possible. The mean number of tests given is for this N only.

Experiment 3.

Decisions for Duplex Gambles with One Parameter Fixed: QFA and functional measurement of information integration models.

The aim of this experiment was to extend the findings of experiment I to DM's choices for more complex gambles, thus applying the techniques for investigating functional relationships to situations where they had not been previously applied. Duplex gambles have more parameters than simple ones. It is therefore necessary to consider a larger set of gambles if the investigation is to be representative. But the number of pairs that can be used in a pair comparison experiment is finite, so the proportion of pairs that can be taken for a representative set of gambles must be small. This limitation only allows a general, "goodness of fit" type of investigation by QFA and functional measurement at this stage. If such a study of a wide range of gambles is successful rigorous tests of a more specific nature may be worth attempting. For the moment a more general, hopefully more representative study of decisions for duplex gambles with one parameter fixed is proposed.

Experimental Design.

The stimuli were pairs of duplex gambles from a set of such gambles. Parameter PL was 0.5 for each of the gambles in the set which consisted of every combination of parameter values: SW, SL = $\{$ 4/-, 2/-, 1/- $\}$ and PW = $\{0.2, 0.4, 0.8\}$. The basic design is an n-replicate, pair comparison one. A complete design would require the comparison of $27 \times 26/2 = 351$ pairs. As usual, it can be assumed that people choose according to the sure-thing principle when a pair is presented in which one gamble dominates the other. This reduces the number of pairs which should be included by nearly a half. This is still far too many for an experiment in which n is any reasonable size. It was decided that $n = 7$ would be large enough for the purposes of the experiment. With $n = 7$ there is a reasonable chance that for many pairs (a_i, b_j) from the set, $np_{ij} > 1$ so that the minimum normit chi-square procedure will be valid.

The choice of pairs to be observed is critical to the experiment. Certain criteria for selecting a sub-set of the possible pairs derive from the aims of the experiment. First of all, they should be representative of the whole set. The smallest representative set would be one which formed a cyclic design (see David, 1963). In a pilot experiment when a set which satisfied the cyclic condition was chosen at random it was found that the data was not suitable for minimum normit chi-squared analysis. Zero and one preference frequencies occurred too often. To reduce the chance of obtaining such preference frequencies the following device was used, familiar from experiment I. Let the elements of the parameters of the gambles be numbered 3, 2, 1 in order from most to least favorable. Define the level of each gamble as the sum of its parameter's numbers. This partitions the set into six subsets, the number of gambles in each being shown below.

Number of Gambles	1	3	6	7	6	3	1
Level	9	8	7	6	5	4	3

Certain properties of this grouping with respect to dominance can be listed:

i) No gamble dominates another of the same level

ii) Gambles only dominate others of a lower level

iii) The single gamble at level 9 dominates the rest and all gambles dominate that at level 3.

iv) The gambles at level 8 dominate those at level 4.

With respect to dominance, then the level structure is hierarchical. It is to be expected also that the subjective difference between gambles will tend to increase as the difference between their levels increases. Thus, to try to select gambles close in subjective value it might be reasonable to choose pairs of gambles of the same level. Levels 9 and 3 were omitted. A "cycle" of pairs within each level was picked at random for each remaining level, giving 25 pairs. These are not sufficient to estimate the scale values of the 25 "middle" gambles by minimum normit chi-square. To remedy this, 4 pairs of gambles, each of which consisted of gambles from adjacent levels were selected. The 29 pairs were sufficient to enable goodness of fit of the basic Thurstone model to be tested.

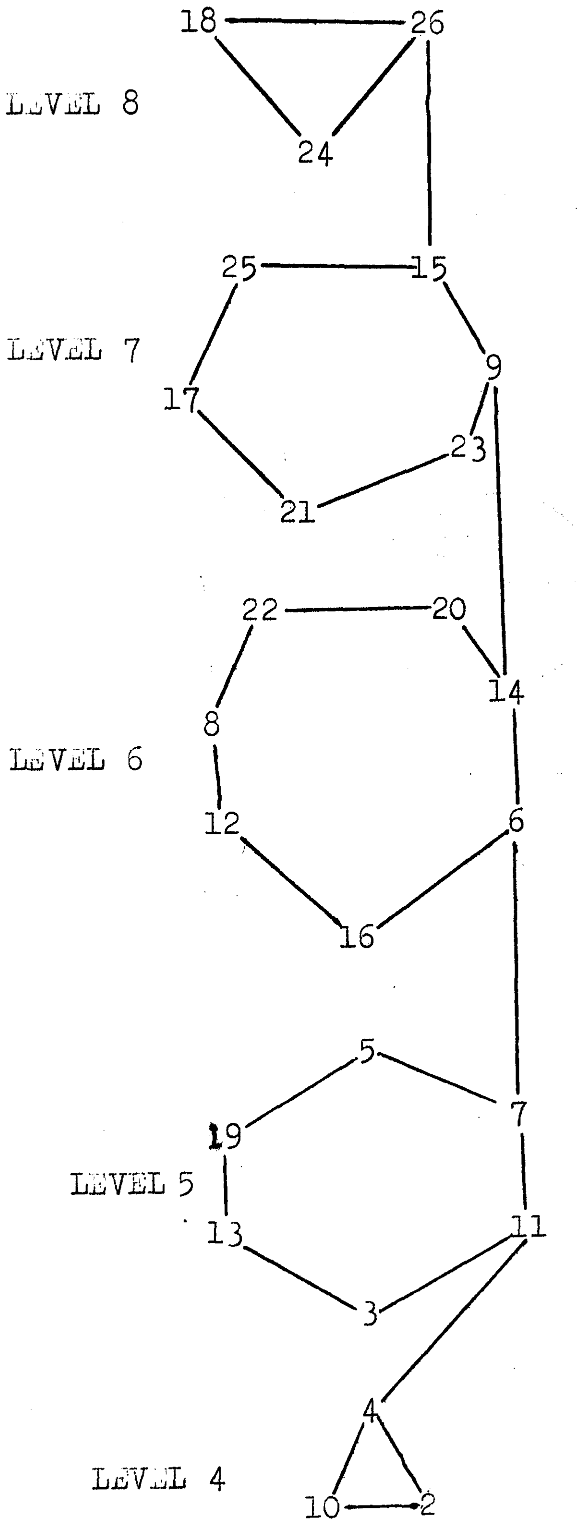
This "cyclic-chain" set of pairs is illustrated in figure 5.3. From the foregoing it is clear that the pairs were selected with the quantitative rather than the qualitative analysis in mind. Coombs, Bezembinder & Goode (1967) note that it is difficult to select a subset of pairs to test cancellation as, depending on DM's actual preferences, any small subset can lead to few, if any, tests.

By selecting triads of alternatives and presenting all pairs in the triad, transitivity can always be tested, however. With the pairs selected transitivity will definitely be tested for the gambles at levels 8 and 4, regardless of DM's choices, but any other tests of cancellation and transitivity will depend on what DM chooses. This is seen as the major problem for QFA. For moderate n and moderate number of pairs a complete pair-comparison is out of the question. An incomplete design leading to a partial QFA is all that can be carried out.

Once the pairs to be used have been decided upon the design of the experiment is straight forward. The 29 pairs were each presented to DM seven times on cards as shown in figure 5.4. The 7 x 29 = 203 trials were carried out in random order subject to the constraint that no pair appeared that had appeared during the previous 10 trials. The gambles on the cards were assigned the left and right position randomly, and the response obtained on each trial was DM's decision as to which gamble of the pair he would rather play should the occasion arise.

Figure 5.3.

Cyclic-chain Design for Experiment 3.



Key.

Gamble Code	Parameter Values SW	PW	SL
26	4	.8	2
25	4	.8	4
24	4	.4	1
23	4	.4	2
22	4	.4	4
21	4	.2	1
20	4	.2	2
19	4	.2	4
18	2	.8	1
17	2	.8	2
16	2	.8	4
15	2	.4	1
14	2	.4	2
13	2	.4	4
12	2	.2	1
11	2	.2	2
10	2	.2	4
9	1	.8	1
8	1	.8	2
7	1	.8	4
6	1	.4	1
5	1	.4	2
4	1	.4	4
3	1	.2	1
2	1	.2	2

A line between two gambles indicates that this pair were presented.

Figure 5.4.

A Typical Gamble Pair as shown to the Subjects.

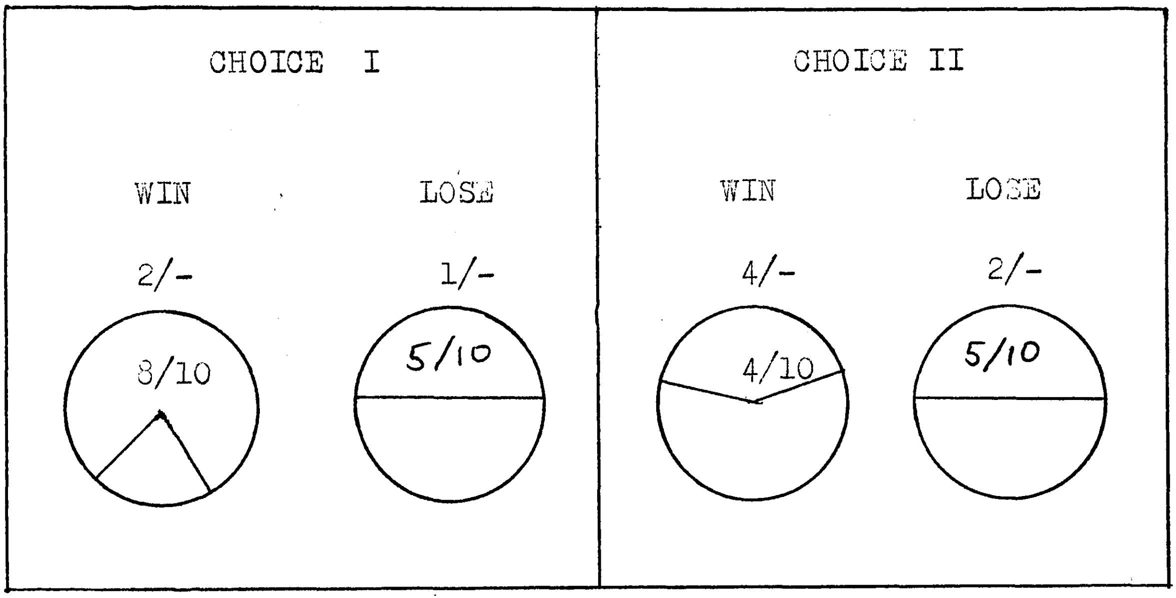
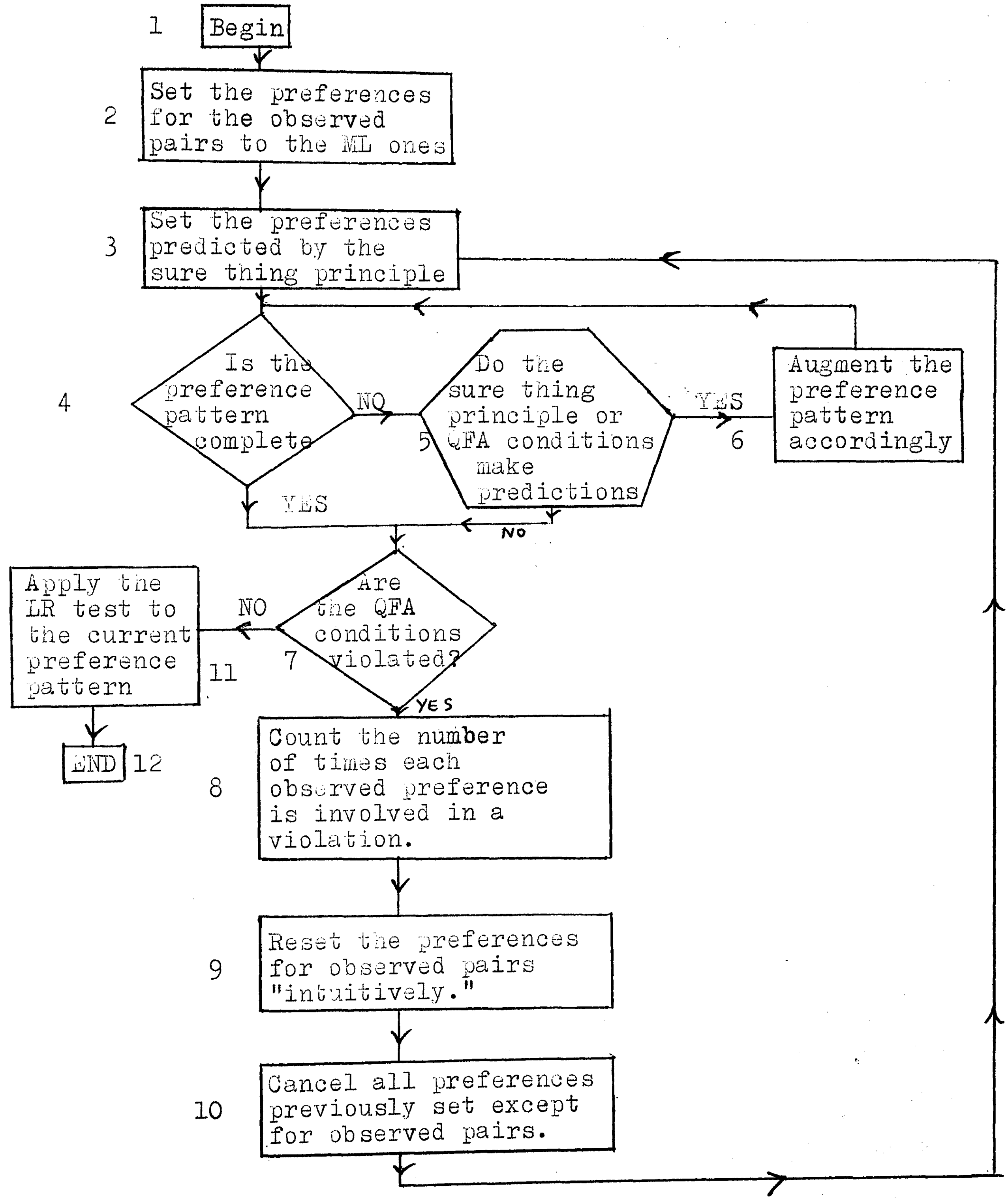


Figure 5.5.

Flow Chart of the "Algorithm" to Test the SEU Model by QFA.



Subjects.

The subjects were 10 undergraduate students at the University of Stirling, 6 male and 4 female, between the ages of 17 and 21. They were all volunteers who had been told there was the possibility of financial gain.

Procedure.

Each subject, (DM) carried out the experiment individually in a quiet room. The experimenter told DM that the experiment involved making choices between gambles where small sums of money could be won or lost. The duplex gamble was then introduced, with a piece of apparatus consisting of two duplex gambles painted on a card, some poker chips and two small roulette wheels, specially converted to wheels of fortune. The DM was given some poker chips, each representing one shilling, He chose one of the gambles, which he then played, and won or lost chips according to the outcomes. After making a few choices he was given a booklet of cards containing pairs of gambles as in figure 5.4. He was asked to go through the booklet, ticking the gamble from the pair that he would rather play. These gambles were such that one dominated the other. If any DM chose the non-dominant gamble it was assumed that he had not understood the nature of duplex gamble which was explained again. At the end, the experimenter picked one of the pages of the booklet at random by throwing a dice. DM was given some chips representing shillings with which to play the gamble he chose on the page picked at random. This he did using the wheels of fortune.

It was then explained that what had preceded (from the time the booklet was produced) was a "miniature" version of the experiment. The actual experiment would differ from it in certain ways: i) it would not be so easy to decide between the gambles (this information was supplied to prepare subjects for gamble pairs where neither was dominant).

ii) a different, and larger set of pairs would be used (the stack of gamble cards was produced at this stage)

iii) when they had made all their choices they would be given a 10/- stake, and they would play three of the gambles they had chosen for real money. They would take away with them the total amount of money they ended up with (greater or less than 10/- depending on their luck). Therefore, three pages from the experimental gambles would be selected by the experimenter at random. The gambles DM had ticked on these pages would be the ones he played.

A different randomization of order and "left-right" was used for each subject. While DM was making his choices the experimenter left the room. DM's were asked to go through the stack of gamble pairs page by page without going back to previous ones.

Results and Discussion.

The problem that subjects may recall their earlier choice on some pair and repeat it was discussed in experiment I. In this experiment, subjects were asked whether they noticed any repetition in the pairs presented. All but one said they did not, and for the subject who did, his grounds were his belief that only a small number of combinations of gambles were possible. They all said that each page looked similar, but most of them could not say whether they had previously seen the pair before them.

The "runs test" used in experiment I was also carried out here. The results of the test are shown in table 5.8. It can be seen that the observed number of runs was less than two standard deviations away from the expected number for all but two of the subjects. Under the assumption that the sampling distribution of the statistic is not too peculiar it can be assumed that the "bernoulli variate" hypothesis is reasonable for these DM's. The number of runs for S5 was more than four standard deviations away from the expected number, and for S9 it was more than two away, both observed values being less than those expected. This can be interpreted as evidence that the DM's probably remembered what they responded last time and tended to respond the same. Overall, though the runs test favours the "Bernoulli" hypothesis.

A more careful check was carried out on whether there was any trend in the response sequences.

As with the runs test, consider (each) DM's responses as a 7 x 29 matrix of 0's and 1's, one row per pair of gambles and the columns representing replications. A zero indicates one of the gambles was chosen and a one the other. It is not meaningful to look for a general increase in the likelihood of a 1 from the first to the last replication, as the labelling of a response as '1' is arbitary. It is meaningful, however, to test for significant heterogeneity of trend in the set of binary sequences for each subject. That is, to test whether there is more trend in the responses sequences, regardless of direction, than one would expect by chance. Jonkheere and Bower (1967) have developed a non-parametric test to investigate this property of a set of binary sequences. It tests the hypothesis that the extent of heterogeneous trend is greater than would be expected by chance under a "Bernoulli variate" model. This test seems to be the most sensitive test of trend available and it is also suitable for small amounts of data as from this experiment. Its other attributes are discussed in the paper cited. The conditions under which Jonkheere and Bower's "W" statistic for heterogeneity of trend approximates the chi-square distribution were not met in the present case. Therefore, the transformation of W they suggest was calculated. This is shown for each subject in table 5.9. It is denoted Z as under the null hypothesis (no trend) it has approximately a unit normal distribution. The null hypothesis cannot be rejected at the 5% level (two tailed) for any subject. Thus, it can be concluded that no subject changed their strategy to any noticeable degree. These two tests both support the assumptions regarding stable response probabilities that underly the substantive analyses which follow. It is worth noting that such tests are really necessary and should be carried out as a matter of course. It may have been that trends or memory factors were prominant which would have made the analysis invalid. The conclusion that the experimental design was not suitable would have resulted from finding widespread trend or memory factors.

The binary preference frequencies are recorded in table 5.10. QFA and functional measurement will be based on these.

Q.F.A.

The QFA is based on the probabilistic definition of preference discussed in chapter 2 and used in experiment I. First, three specific models will be tested against the maximum likelihood, bernoulli model. They are the EV and level hypotheses tested in experiment I, and the indifference model. The last named is that which predicts DM will be indifferent between any two gambles. The EV model is a special case of the SEU model while the level model is a special case of the additive model. The indifference model is a special case of both. They can all be tested against the general unrestricted model by the LR test, the results of which are shown in table 5.11. For subject 1 to 9 all the chi-square statistics are significant at least at the 1% level, so that these simple models can all be rejected. For subject 10, however, the indifference model cannot be rejected ($0.1 < p < 0.9$). Since it is a special case of the additive and SEU models no further QFA of these models is necessary for subject 10. Acceptance of the indifference model implies acceptance of them both. The SEU model will be tested by QFA for the other 9 subjects.

The method of testing the SEU model by QFA, assuming the probabilistic model, was discussed in chapter 2 and applied in experiment I. The only consequences of SEU relevant are the transitivity and cancellation conditions, H9C2, H9C3 and H9C3'. The maximum likelihood set of underlying preferences consistent with these conditions can be tested against the unrestricted point probability hypothesis by the LR test. The main problem, as before is the maximization of the likelihood of the data under the restrictions. The algorithm which is known to do this efficiently has not been programmed. Unfortunately, an adaptation of the less efficient algorithm which worked in experiment I did not converge to a solution in the present case.

The ad hoc method which had to be used to give a statistical test of the qualitative consequences of SEU listed above will now be described. It is far from satisfactory but it worked. A flow chart of the method is set out in figure 5.5. The first step is to set the preferences for the observed pairs to the maximum likelihood ones and also to set the preferences predicted by the sure thing principle. Now, in experiment I the preference pattern so obtained was tested to see if it violated the conditions. An amendment to this operation, shown in boxes 4 5 and 6 of the flow chart was necessary as the preference pattern in the present case was incomplete. The incomplete preference pattern must be examined to see if predictions follow from the sure thing principle and/or the QFA conditions. The predictions are then set to augment the preference pattern. Tests and augmentations are repeated until no new predictions result. Only then is the test of violations of the QFA conditions carried out. If no violations are found the pattern maximizes the likelihood of the data. Otherwise, the number of times each ML preference was involved in a violation was counted and this information used to reset some of the preferences on the observed pairs "intuitively". By simple inspection it was possible to deduce which changes would be likely to reduce the violations most. Of course, a certain amount of trial and error was necessary. This whole operation was repeated until a set of preferences was found which resulted in no violations of the consequences of SEU. This set was used as an estimate of the maximum likelihood pattern consistent with SEU. It was then tested, assuming the probabilistic error model by the likelihood ratio test. Hopefully, figure 5.5 makes it possible to understand this "algorithm", It probably worked because the observed preferences, the sure thing principle and the consequences of SEU left the strict preference and indifference relations considerably under determined.

The results of the QFA of the SEU model are set out in table 5.12. For each subject the pattern closest to the ML preference pattern which satisfies SEU is shown. The numbers in brackets indicate the ML preferences for those pairs where these are different from those of the pattern satisfying SEU. It can be seen that for every subject it was necessary to change only a few preferences before a pattern which did not violate SEU was found. For subject 5 the initial, ML pattern was satisfactory. For subject 4 the most number of changes, 8 were necessary. The median number of changes made before a suitable pattern was found was 2. Changes first attempted in the "intuitive" box of the algorithm were indifferences involved in a moderate number of violations, which were changed to the most likely strict preference.

In the penultimate row of table 5.12 chi-square statistics for the LR test of the ML, SEU consistent preference pattern against the unrestricted bernoulli model are shown. These chi-square statistics should, under the null hypothesis be distributed with degrees of freedom shown in the bottom row of the table. No chi-square is significant at the 5% level for any subject. Therefore, the broad, "goodness of fit" QFA test of the SEU model, assuming probabilistic model 1 has lead to its acceptance for each subject.

Functional Measurement.

An important factor in the functional measurement is the proportion of zero or one preference frequencies that occurred. If they dominate, the estimates and goodness of fit statistics will be severely distorted. From table 5.10 it can be seen that the selection of pairs to make non-zero or one frequencies more likely was quite successful. The most non-zero or one preference frequencies observed was 14, for subject 3 and the least was 0, for subject 10. The median number was 9.

The goodness of fit of the basic Thurstone model is shown in table 5.13. The significant chi-squares of estimation for subjects 1 to 9 (1% level) mean that in each case the hypothesis that all the affective values are zero can be rejected. The only error chi-square value to reach the 5% level of significance is that for subject 3. For this subject the Thurstone model can be rejected, but for the remaining 9 it is accepted. Subject 10's non-significant estimation and error chi-squares lead to the conclusion that he was indifferent between the gambles of all pairs and chose at random.

The above conclusions all rest on the assumption that Bock and Jone's asymptotic results hold in the present case. This assumption will be examined in a small sampling experiment when the rest of the functional measurement results have been presented. This approach - discussing the findings in terms of the asymptotic results and questioning the assumptions separately - is taken because in general results would be discussed in terms of asymptotic theory. It is not generally possible to discuss empirical results in terms of simulated statistical distributions.

Since the basic model seems satisfactory certain multiplicative and additive information integration models can be considered. In all cases the estimation chi-square for subject 10 was non-significant so the random behaviour hypothesis about this subject was not changed. Subject 10 will not be discussed further. Since the error chi-square for subject 3 was "only just" significant his results will be included in further analysis. The two main hypotheses that will be considered are the several additive and SEU models that were discussed in chapter 3. In analysis of variance terms, the additive model is that which predicts no interactions are significant and the SEU model predicts that the only significant interaction is the bilinear, SW x PW interaction. Bock and Jones' procedures were applied to the subjective values and variance-covariance matrix of the basic Thurstone model for each subject. This enabled estimation of the parameters of the information integration models to be carried out and also it was possible to partition the Thurstone model's estimation chi-squares as in table 5.14. The additive model is tested by the sum of the chi-squares due to interaction terms. The SEU model is tested by the chi-square due to all interactions except the bilinear

SW x PW one. It can be seen that both models can be rejected for subjects 3 and 4, at the 5% level of significance. For subject 3 these models are rejected in addition to the basic Thurstone model (see table 5.13). The component of the chi-square due to the bilinear SW x PW interaction has a single degree of freedom. Its significance leads one to accept the SEU model in favour of the additive. This was possible for 5 of the 7 subjects who remain to be considered. The results of the foregoing comparison thus lead to the conclusion that the general SEU model accounts for most subjects choices better than the additive model does.

As in experiment 1, however there may be simple linear models with only a few parameters based on the objective values of the gambles, which account adequately for peoples choices. Two such models were examined, corresponding to those considered in experiment I. One is Slovic's additive model and the other is the SEU model where the subjective value scales are linear functions of the objective values. As with the general models they are differentiated by the additive model having "main effects" only and the linear SEU model having an extra degree of freedom corresponding to the permissible SW x PW interaction. The partition of the basic Thurstone estimation chi-square for these models is shown in table 5.15. The table is analagous to that for the general models and conclusions are arrived at similarly. Both models are rejected at the 5% level for subjects 3 and 8. For 6 of the remaining 7 subjects, the linear SEU model is accepted and for the other subject. Slovic's additive model is accepted. This is based on at least a 5% significance level. The inadequacy of Slovic's model for most subjects is illustrated further in figure 5.6. Here the subjective values estimated directly from the data (observed) are compared to those predicted by Slovic's model. If the model fits well most points should lie along the main diagonal. This is not evident even for the subject for whom the model is accepted.

Table 5.8.

The Observed Number of Runs, with its Expectation and Standard Deviation under the Null Hypothesis for Each Subject.

Subject	Observed Runs, R	E(R)	SD(R)
1	67	65.10	2.92
2	85	84.43	4.10
3	63	65.08	3.01
4	92	98.57	4.68
5	52	64.43	2.72
6	72	78.14	3.70
7	71	75.29	3.42
8	70	76.43	3.60
9	82	78.78	3.53
10	115	114.73	5.48

Table 5.9.

Jonkheere and Bower's Heterogeneity of Trend Statistic, Z for Each Subject.

Subject	Z
1	-1.407
2	-0.279
3	-0.852
4	0.553
5	-1.708
6	-0.607
7	-0.701
8	1.085
9	-1.701
10	-

Table 5.10.

The Number of Times Each Subject Made Choice One for Each Gamble Pair.

Gamble	Subject								
Pair	1	2	3	4	5	6	7	8	9
1	7	2	3	6	7	1	2	2	3
2	1	0	0	1	0	0	3	1	1
3	5	6	7	5	7	7	5	2	7
4	5	7	6	6	7	7	7	6	7
5	0	0	0	1	1	1	6	5	5
6	7	7	7	6	7	5	2	6	3
7	1	5	6	6	1	7	4	7	3
8	2	4	0	3	1	0	2	2	2
9	7	2	4	2	7	3	1	3	5
10	0	7	7	2	0	6	4	4	1
11	0	6	7	3	6	7	7	7	6
12	0	3	7	4	2	7	6	6	6
13	2	5	0	4	6	1	6	5	7
14	7	3	1	7	2	3	0	1	1
15	0	3	7	3	1	7	6	5	1
16	3	3	0	3	6	2	6	2	6
17	6	2	3	5	2	1	0	0	2
18	1	1	0	0	5	2	5	7	1
19	5	7	7	7	7	5	7	7	7
20	2	1	2	2	0	2	1	0	1
21	1	5	6	2	6	6	6	7	6
22	7	6	0	5	7	1	2	2	1
23	6	2	6	3	1	2	1	0	0
24	0	2	2	2	3	4	6	7	3
25	6	7	1	5	7	3	7	2	6
26	7	6	6	7	7	7	6	7	7
27	1	1	6	3	1	6	7	5	3
28	7	3	3	6	1	1	0	1	1
29	6	7	5	3	6	5	7	6	7

Table 5.11.

The Likelihood Ratios and Degrees of Freedom for the EV, level and indifference models for each subject tested against the unrestricted model.

Subject	Indifference		Level		EV	
	Chi-square	DF	Chi-square	DF	Chi-square	DF
1	169.98	29	170.98	28	217.73	21
2	116.86	29	89.52	25	66.76	18
3	173.70	29	144.78	25	166.22	17
4	79.79	29	60.94	26	122.80	22
5	173.81	29	136.38	25	53.95	10
6	132.87	29	102.72	26	93.86	18
7	139.32	29	106.25	25	71.77	14
8	138.35	29	105.27	25	56.39	13
9	128.10	29	95.03	25	55.05	13
10	35.76[a]	29	-	-	-	-

[a] probability of chi-square value, p is in the range

$0.10 < p < 0.90$

All other chi-square values are significant at least at the 1% level.

Table 5.12.

Patterns of Preference Consistent with the predictions of the SEU model, the chi-square associated with it and its degrees of freedom.

Gamble Pairs		S1	S2	S3	S4	S5	S6	S7	S8	S9
24	18	1*	2	0	1	1	2	2	2	1(0)
18	26	2	2	2	2	2	2	2(0)	2	2
26	24	1	1	1	1	1	1	1	1(0)	1
26	15	1	1	1	1	1	1	1	1	1
15	25	2	2	2	2	2	2	1	1	1
25	17	1	1	1	1	1	1	2	1	0
17	21	2	1	1	1	2	1	1(0)	1	2(0)
21	23	2	0	2	2(0)	2	2	2	-1	2
23	9	1	2	1(0)	1(2)	1	2(0)	2	2(0)	1
9	15	2	1	1	2	2	1	0	0	2
9	14	2	1	1	1(0)	1	1	1	1	1
14	20	2	2(0)	1	2(0)	2	1	1	1	1
20	22	2	1	2	0	1	2	1	1	1
22	8	1	2(0)	2	1	2	2(0)	2	2	2
8	12	2	2(0)	1	2(0)	2	1	1	1	2
12	16	2(0)	1(0)	2	0	1	2	1	2	1
16	6	1	2	2(0)	1	2	2	2	2	2
6	14	2	2	2	2	1	2	1	1	2
6	7	1	1	1	1	1	1	1	1	1
7	5	2	2	2	2	2	2	2	2	2
5	19	2	1	1	2	1	1	1	1	1
19	13	1	1	2	1	1	2	2	2	2
13	3	1	2	1	1(0)	2	2	2	2	2
3	11	2	2	2	2	0	0	1	1	0
11	7	1	1	2	1	1	0	1	2	1
11	4	1	1	1	1	1	1	1	1	1
4	10	2	2	1	2(0)	2	1	1	1	0
10	2	1	0	2(0)	1	2	2	2	2	2
2	4	1	1	1	2(0)	1	1	1	1	1
Chi-square		3.388	12.930	6.666	32.73	1.317	6.102	5.665	13.276	7.845
Degrees of freedom		7	14	7	18	5	11	9	12	8

* If the pair is (a,b), 1 indicates a $>$ b, 0 indicates a $\sim$ b and 2 indicates a $<$ b. The numbers in brackets indicate the initial, ML preferences where these differed from the final ones.

Table 5.13.

Goodness of Fit of the Basic Thurstone Model.

Subject	Chi-Squares. Estimation	Error	Total
1	72.00**	2.70	74.70
2	57.86**	3.27	61.13
3	60.42**	14.34*	74.08
4	45.74**	1.32	47.06
5	84.07**	2.55	86.62
6	62.46**	1.65	64.11
7	73.55**	1.40	74.95
8	59.09**	11.01	70.10
9	65.37**	1.12	66.49
10	24.01	3.32	27.33
Degrees of Freedom	24	5	29

* indicates significance at 5% level,

** indicates significance at 1% level.

TABLE 5.14.

The Goodness of Fit of the SEU and General Additive Models: Partition of the Estimation Chi-Square.

Subject	Components of the Estimation Chi-squares		
	Main Effects	Bilinear SWxPW Interaction	Remainder of Interactions (Error)
1	50.17**	0.20	21.63
2	27.64**	4.00*	26.22
3	28.87**	0.30	31.25*
4	10.10	6.33*	29.30*
5	54.52**	5.83*	23.72
6	36.66**	10.56**	15.24
7	56.81**	3.03	13.76
8	34.91**	8.06**	16.11
9	41.17**	6.11*	18.21
Degrees of Freedom	6	1	17

* indicates significance at the 5% level

** indicates significance at the 1% level.

Table 5.15.

Goodness of Fit of Slovic's Additive Model and the Linear SEU Models: Partition of Estimation Chi-Square.

Subject	Components of the Estimation Chi-square.		
	Main effects	SWxPW Interaction	Remainder Interaction (Error)
1	40.49**	0.5	31.46*
2	25.35**	6.33*	26.08
3	16.08**	6.40*	37.94*
4	12.50**	4.88*	28.40
5	49.65**	12.90**	21.62
6	25.61**	18.46**	18.39
7	50.44**	6.78**	16.29
8	23.40**	0.04	35.65*
9	28.22**	14.13**	23.03
Degrees of Freedom	3	1	20

* indicates significance at the 5% level.

** indicates significance at the 1% level.

Figure 5.6.

Graph of subjective values estimated directly from the data against those predicted by Slovic's model.

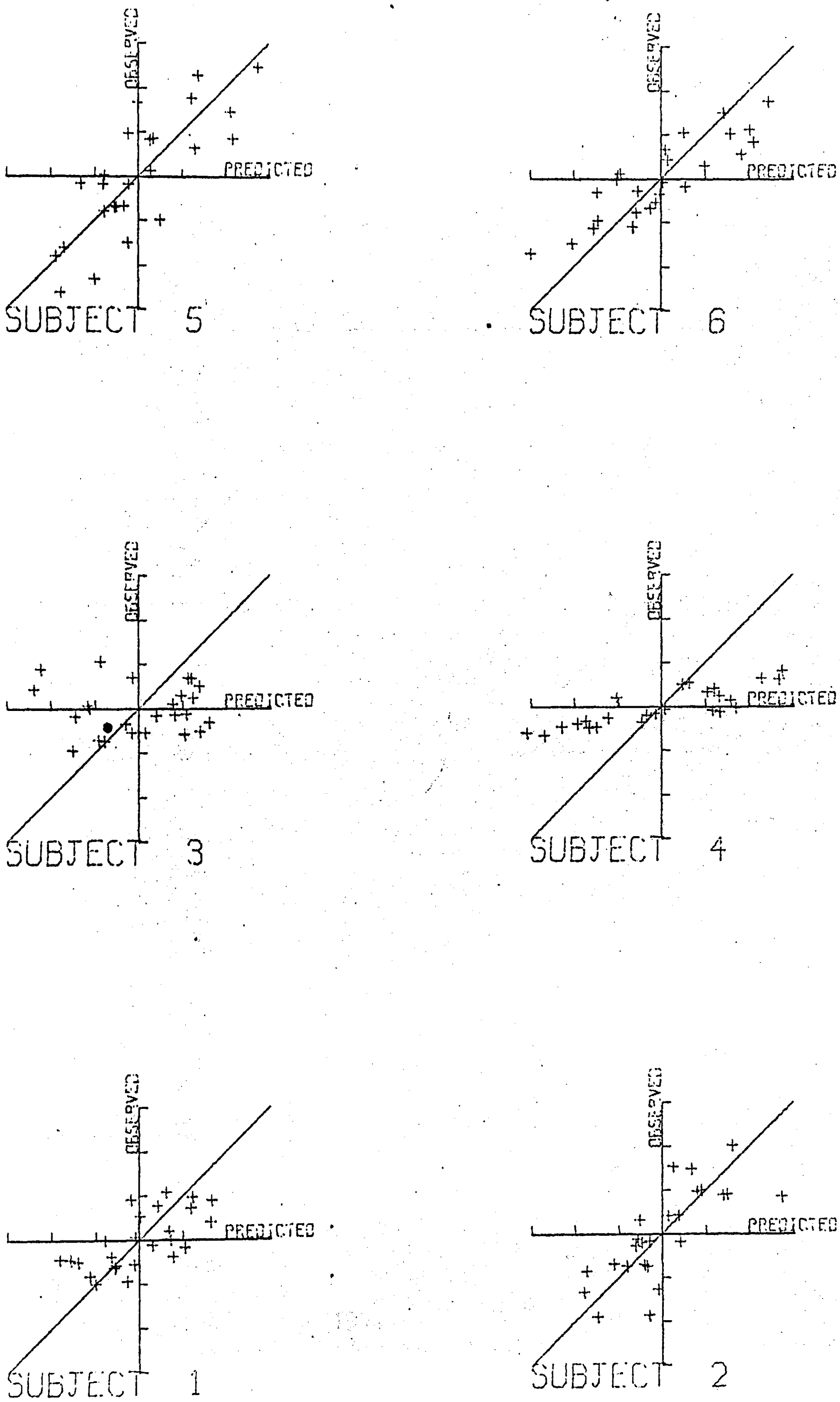
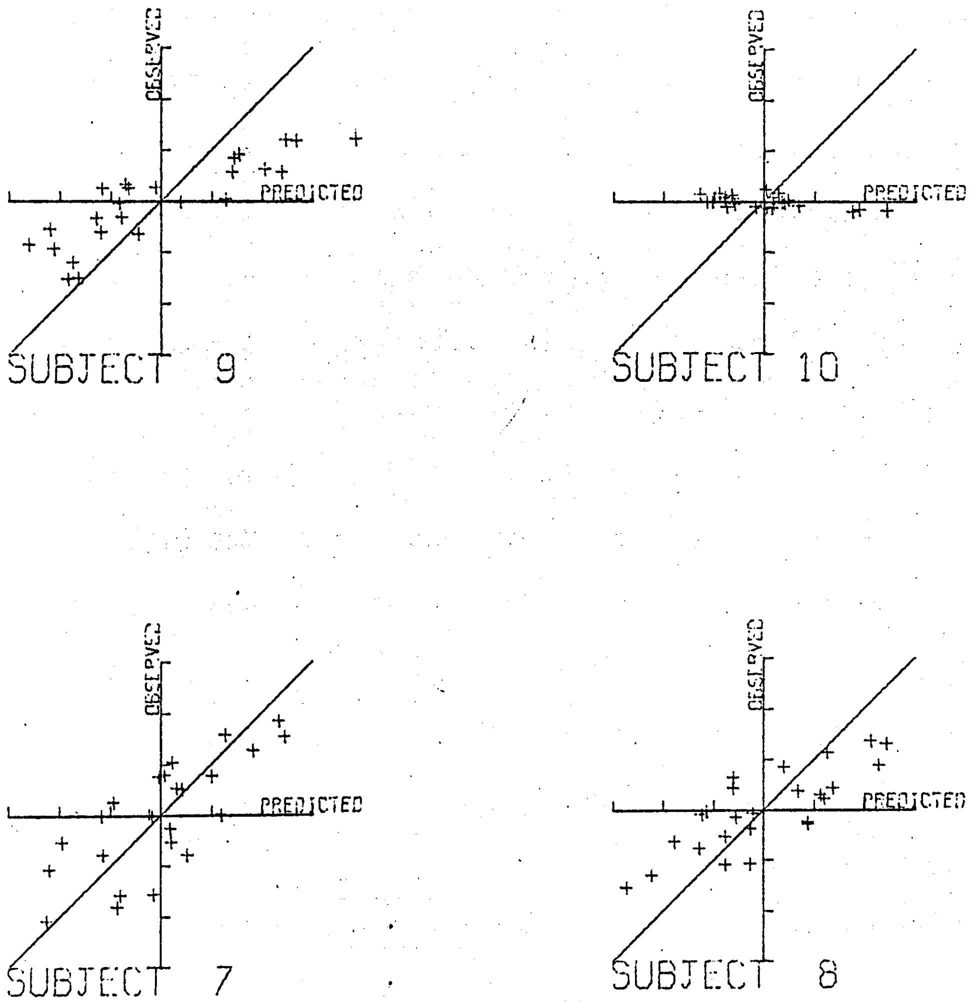


Figure 5.6 cont.



A Sampling Experiment to Examine the Assumptions Underlying the Functional Measurement Analysis.

The purpose of this simulation is to examine the test and scaling assumptions for the particular experimental design used in experiment 3. The method used is the same as for the sampling experiment reported in chapter 3. Sets of data were simulated from a process which generated data under the standard Thurstone case V assumptions with sampling standard error equal to one and known subjective values. The known subjective values were those that had been estimated in experiment 3 for subject 9.

250 sets of data in the 7- replicate, incomplete pair comparison situation of experiment 3 were simulated under the above conditions. The minimum normit chi-square method was then used to estimate the subjective values and derive the goodness of fit statistic, SSE. From the theory set out in chapter 2 the theoretical variances of the estimates and the theoretical sampling distribution of SSE under the null hypothesis are known. The expected values of the estimates are, of course the actual input values. There are 25 objects about which subjects had to choose in experiment 3, labelled arbitarily 1 to 25. The location of the estimated scale in the sampling experiment was fixed by letting object 25 be zero. The estimates obtained, therefore were of the remaining 24 objects relative to the last one.

The mean and sampling variance of the 250 estimates of each subjective value were computed and the 250 SSE values obtained were recorded in a histogram with lower limit zero and step size I. In figure 5.7 the mean observed subjective values are plotted against the actual subjective values. Of course, the expected plot should be along the main diagonal. It can be seen that the mean estimated scale of subjective values is somewhat compressed, relative to the actual scale. It does appear, though that the mean observed values at least retain the same order as the true values and apart from a scaling factor the interval properties appear similar to those of the true scale. Figure 5.8 plots the observed sampling variances against theoretical ones and reveals the estimation procedure to be somewhat erratic, the observed variances being in many cases far higher than expected. It is probable that this is due to the pair comparison design being very incomplete, observations from only 29 out of about 150 non-dominating pairs being taken, and also to n being only 7.

The theoretical sampling distribution of SSE is the chi-square with 5 degrees of freedom. It is plotted in figure 5.9 along with the histogram of the observed SSE values from the simulation experiment. It is clear that the actual sampling distribution of SSE does not, in the case of the design of experiment 3, correspond very closely to the theoretical one. However, the 95th percentile of the observed distribution was 19, far higher than that of the theoretical distribution. Thus, using the theoretical distribution and significance level 5% for the goodness of fit statistic is likely to be a conservative test of the model. That is, there will be a tendency to reject the model when it is true rather than accept it when it is false.

The results of the simulation experiment can be summarized as follows. When the design of experiment 3 is used there will be some bias in the estimation of subjective value. This is swamped, however by the variability of the estimates produced by the procedure, which makes it somewhat unreliable. The actual sampling distribution of SSE deviates considerably from that expected, but a 5% significance level based on the theoretical distribution can safely be used to give a rather conservative goodness of fit test.

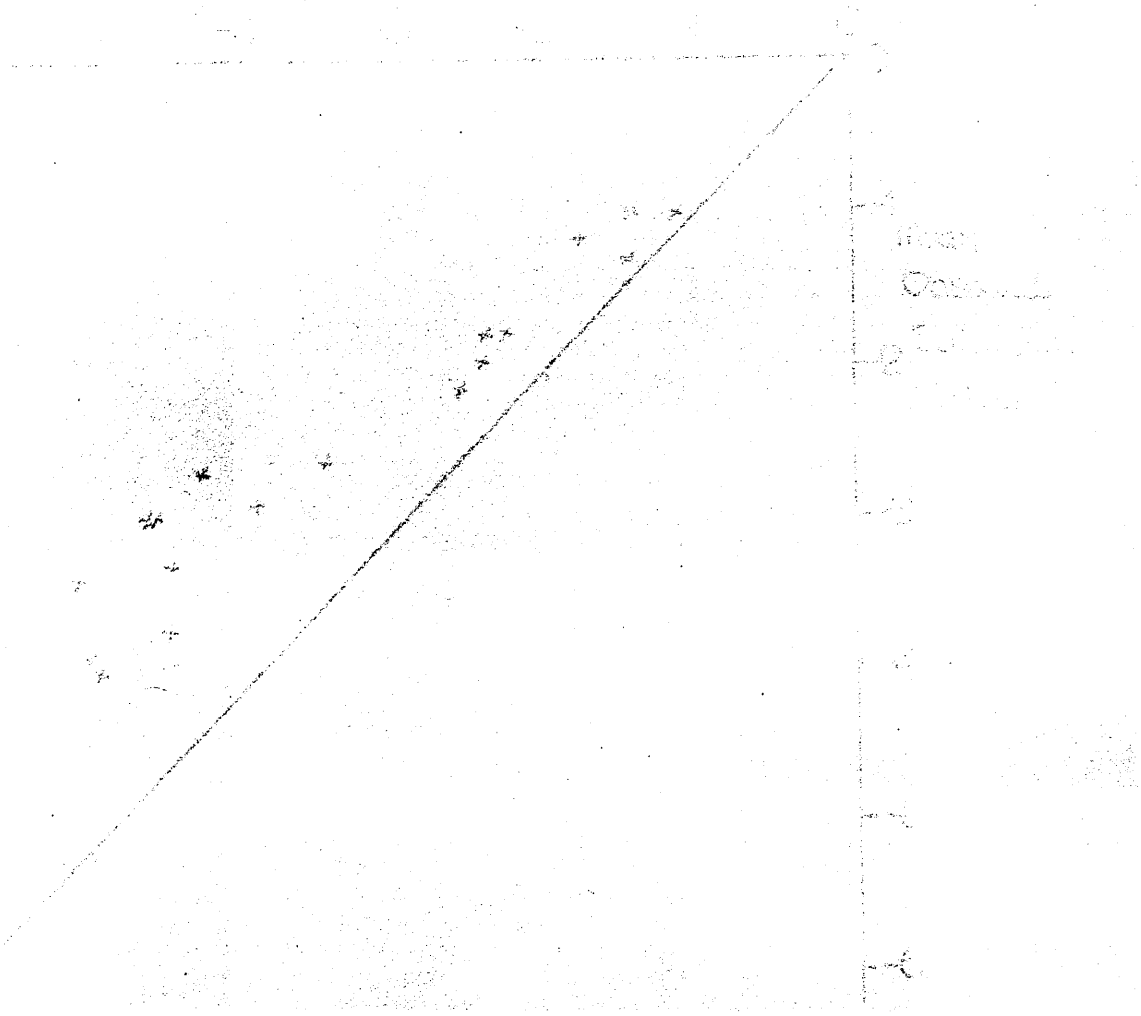


FIGURE 5·7

Mean Observed Subjective Values

Mean Observed Subjective Values against the True Subjective Values in the Simulation of Experiment 3.

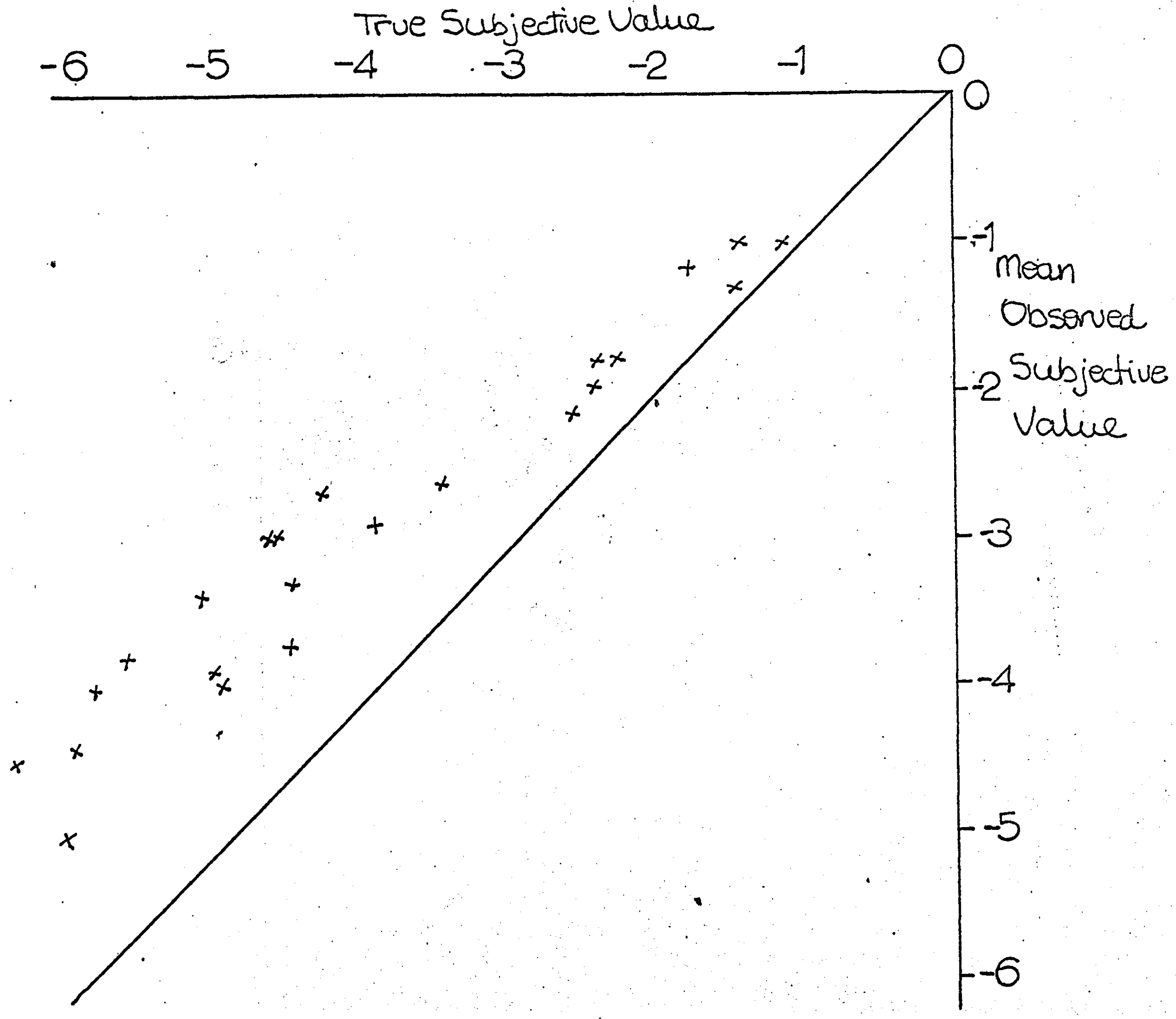


FIGURE 5.8

Observed Sampling Variance of Subjective Values against the Theoretical Variance in the Simulation of Experiment 3.

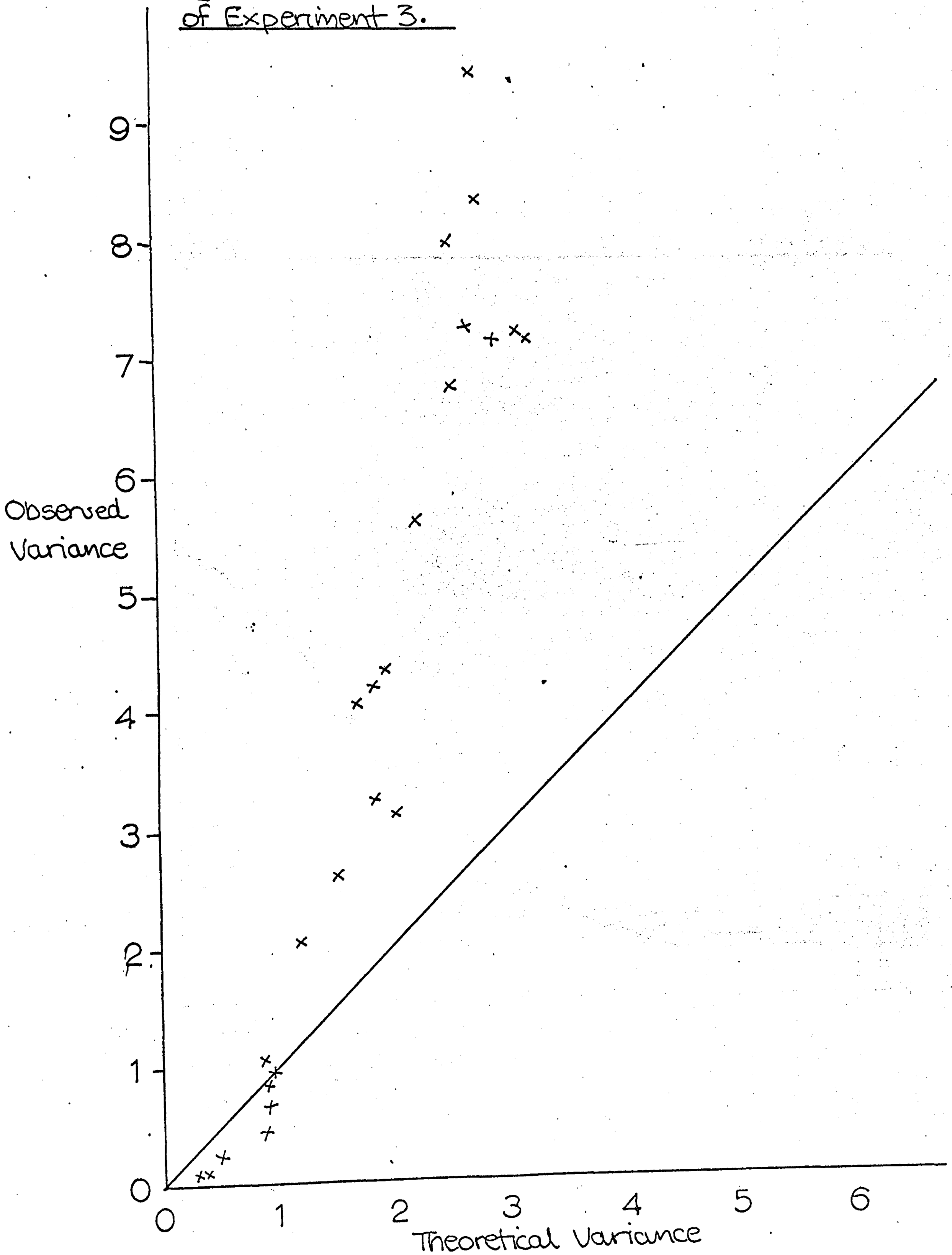
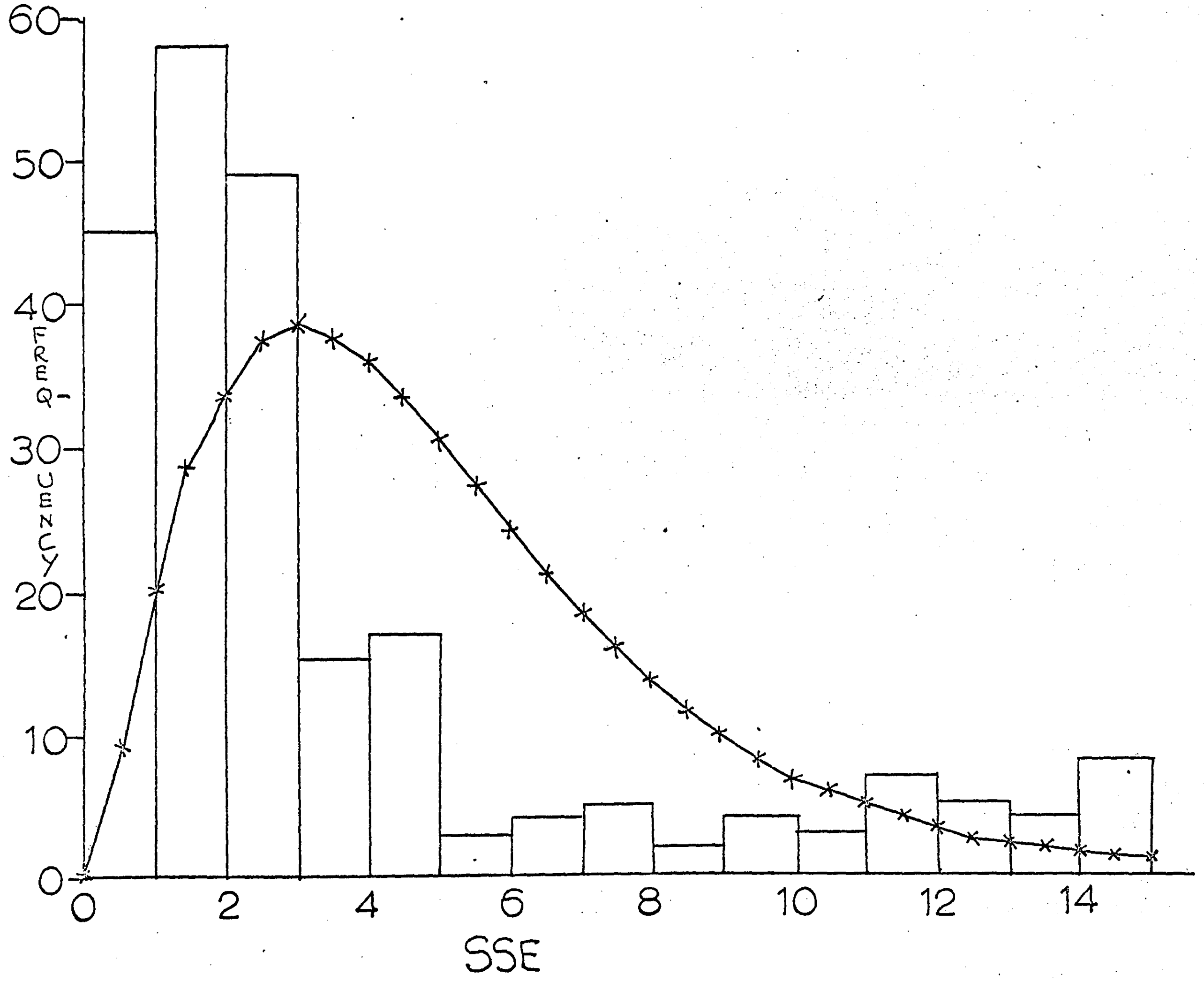


FIGURE 5.9

Observed and Theoretical Frequency Distribution of SSE in the Simulation of Experiment 3.



Discussion.

As in experiment I, results from QFA and functional measurement are more or less compatible with one another. In particular, both analyses concluded that one of the subjects, subject 10 chose randomly. QFA of the general SEU model led to its acceptance for all 9 subjects. Functional measurement showed that at least one of the additive or SEU models was compatible with all 9 "non random" subjects' choices with the possible exception of subject 3 for whom the basic Thurstone model appeared doubtful. Unfortunately the broad consistency between the two approaches is not surprising because of the lack of power of the QFA test. This will be discussed further in the final chapter.

The functional measurement showed that additive models can adequately account for the decision making of some subjects. For most of them, however multiplicative models were significantly better. The general SEU model accounted well for 5 subjects choices and the linear SEU model, despite the fact that it only has 4 parameters accounted adequately for 6 of the 9 "non-random" subjects' choices. That this simple model fitted the data is rather useful. It means that predictions of choice based on the objective values of any gambles considered, and estimates of only 4 parameters can be made with reasonable confidence.

The above conclusions, which are satisfactory from a theoretical point of view, are clouded by doubts about the assumptions underlying the statistical tests of the models. The chi-square values observed in the LR tests of SEU by QFA are in most cases considerably lower than those expected. This may be due to the affects of the LR test being a conditional one in this application, a point discussed in chapter 2. It may be quite wrong to assume that the appropriate degrees of freedom for the test is the number of actually restricted observations. However, it is not worth following up the suspicion of faulty test assumptions as there was a subjective element in the procedure used to estimate the maximum likelihood, SEU consistent preference pattern. The current methodological issues associated with QFA will be discussed in the final chapter.

The functional measurement proceeded by true algorithms and it was possible to test the assumptions underlying it empirically. The test showed that the estimation procedure was somewhat unreliable, in that two sets of data generated under the same assumptions could give widely differing subjective value estimates. Since the tests of functional relationships were based on such estimates their unreliability is the reason that doubt about the conclusions regarding the functional relationships is warranted. However, general optimism can be retained about the method of functional measurement used. Its faults may be rectified by a different choice of incomplete design or a larger n than 7. In the experiments to be reported in sequential and dynamic situations, this type of analysis rather than QFA will be applied as it has far less problems at present than the latter type of analysis.

Chapter 6.

Sequential Experiments.

Sequential decision making situations were discussed in the introduction and previous research on the effect of the situational factors, current capital (cc) and previous outcome (o) was reviewed in chapter 4. In the present chapter, 2 experiments are to be reported. In the first, experiment 4, subjects repeatedly make decisions between the same pair of duplex gambles. In the second, experiment 5 pairs of gambles, the same pairs as were used in experiment 3 are presented to subjects in an n-replicate, incomplete pair comparison experiment. Each gamble chosen is played before the next pair is presented. In experiment 4 many choices are obtained on a single pair and in experiment 5 a few choices are obtained on many pairs, hopefully enabling a comprehensive examination of the effects of the two sequential variables to be carried out.

If sequential variables have no, or only "second order" effects the same information integration models as applied in static situations, including additive and SEU models may account adequately for decision making in sequential situations. To this end, in experiment 5 a functional measurement analysis of information integration models is carried out. It was decided to use functional measurement, rather than QFA in view of the difficulties already encountered with the latter.

In chapter 4 certain methodological and statistical criticisms of previous research on the effect of sequential factors were made. The following recommendations for future experiments were suggested: i) the factors o and cc should not be confounded ii) sensitive tests of changes in responding should be used iii) DM's information should reflect the true probabilities and payoffs associated with alternatives and iv) the effect of response bias should be controlled. All these things were put into practice in experiments 4 and 5. Rappoport et al's (1971) solution to i) of blocking the sequence of choices and assigning different cc values at the start of each block was adopted. The factors cc and o can never be completely unconfounded but the blocking reduces the dependency a great deal. Jonkheere and Bower (1967) have developed a sensitive test of trend which will be applied in these experiments to examine changes in responding as a function of time, cc and o. The other experimental or statistical improvement which needs to be explained is how to deal with response bias. There were 2 physical responses possible on any trial. Which alternative corresponded to which response was randomized over trials. It is hoped that these practices will make the results obtained much less suspect.

Experiment 4.

An Investigation of the Effects of Previous Outcome and Current Capital on Repeated Choices from a Pair of Duplex Gambles.

The basic model underlying this experiment is a random utility model, similar to the Thurstonian information integration models previously discussed. Suppose the duplex gambles are g_1 and g_2. The decision maker makes a choice between them and the gamble chosen is then played. Let this continue through n stages, $t_1, t_2, \ldots t_n$ and let the utility of g_i at stage t_j be:

$$u(g_i, t_j) = a_{ij} + e_{ij},$$

where a_{ij} is a constant and e_{ij} is a random variable with $E(e_{ij}) = 0$ and e_{ij}, e_{ik} uncorrelated, $e_{ij} = e_{ik} = e_i$, $i = 1, 2$. Though other distributions of the error term would suffice, it is convenient to assume that it is distributed logistically.

The aims of the experiment are to test certain hypotheses about the difference in utility between g_1 and g_2, which can be expressed as

$$u(g_d, t_j) = u(g_1, t_j) - u(g_2, t_j)$$
$$= (a_{1j} - a_{2j}) + (e_{1j} - e_{2j}) = a_{dj} + e_d$$

where the d suffix denotes "difference". The assumption that e_d is distributed logistically follows from assuming the component errors are.

The first test is of the hypothesis that a_{dj} is not a function of t_j itself. That is, that there is no change in the choice probability simply due to time passing. The other basic hypothesis is that a_d is independent of the previous choice made.

If the above hypotheses are rejected it will then be possible to investigate the effects of current capital and previous outcomes, which is the major purpose of the experiment. This aim can be put in terms of the above basic model. (Reference to t_j in the utility difference equation can be omitted if no trend is found). Suppose the previous outcome was a change in capital of o pence and the current capital is c pence. The hypothesis that these affect the utility difference can be put in terms of the general linear model:

$$u(g_d) = a_d + l_o + m_c + (lm)_{oc} + e$$

where l, m and (lm) are the main effect and interaction parameters, and a_d becomes the general mean. Standard tests of whether o and cc affect the utility difference, and if so whether their effects are independent can be carried out. Many other hypotheses about how previous outcomes, choices and capital affect choices can be proposed and some of these will be discussed in the results section. That the previous ones have been selected as the "key" hypotheses is largely a result of following previous work. Utility differences have commonly been hypothesized to change with current capital and previous outcomes so this is what has been examined.

The Experimental Setting.

The duplex gamble is introduced to DM using the game for this purpose described in Experiment 3. In the actual experiment, though, wheels of fortune are not used. The usual diagramatic form of presenting the gamble alternatives is used, but actually playing a duplex gamble is mechanized. The basic information about the events in the n-stage sequential gambling game that DM is requested to play is given to him on a control panel which is represented in figure 6.1. At the start of the first trial of the game, DM is given a stake, in shillings. The amount is shown on a meter which is controlled manually by the experimenter. At the start of every trial, one of the blue lights A or B is on. This indicates which information card (A or B) should be used on this trial. These cards give the probability and payoff information about the gamble alternatives and are shown in figure 6.2. It can be seen that the only difference between them is that choice I is on the left on card B and on the right on card A.

The information card to be used is positioned below the control panel. If DM wishes to play the gamble on the left of the card he presses the button on the left of the control panel. Similarly for the gamble on the right. Which button he presses, and which light was on (A or B) at the start determines which pair of outcome lights are relevent on this trial. This relationship between outcome lights, initial lights and responses is shown in table 6.1. In this mechanized duplex gamble game, DM wins if the win outcome light comes on and he loses if the lose light comes on. The complementary events "not win" and "not lose" occur if the appropriate light does not come on. Thus a play is formally equivalent to a play of a duplex gamble on wheels of fortune if the probabilities of the various events correspond to the proportions given on the information card, and payoffs also correspond. After observing the outcome lights, DM and the experimenter calculate the net gain or loss, which is recorded on the "current capital" meter. This ends the trial. The start of the next trial is preceded by DM pressing the "next trial" button, after which light A or B comes on.

Obviously, the purpose of using the alternative information cards is to enable response biases to be taken account of. This usually will be done by letting the probability of light A coming on be 0.5 (and light B also). Thus, there are usually two sets of probabilistic events in the game: the onset of lights A and B and the onset of the outcome lights.

This mechanical gambling game is sequenced by alpha-numeric information on a punched paper tape which is read by a standard tape reader. The information is used in a logic circuit which directs the events to and from the control box. Thus, the whole play of the game is predetermined. The uncertainties of the game are introduced in the construction of the paper tape. An algol computer programme was written to make the tape and a reliable pseudo-random number procedure generated the symbols which signified the uncertain events with the appropriate probabilities. The sequence of events during an experimental session, including choices and outcomes was output on paper tape by a tape punch.

The main advantage of this mechanization is obviously speed. Its main disadvantage is that DM must take the experimenter's word that the game is fair. Unlike the wheels of fortune form of gamble, it is not seen to be fair.

Subjects.

Ten undergraduates from the University of Stirling, 6 male and 4 female took part in individual experimental sessions. They did not know when they volunteered that they would be paid for participating.

Experimental Design.

The first problem in the design of this experiment is to select a suitable pair of gambles. The main criterion is that there should be no obvious choice. The gambles should be quite close to each other in subjective value, so that both are chosen a reasonable number of times in the session. (If one of the gambles is chosen all the time this is not particularly informative).

Minor considerations are that both gambles should be reasonably favorable and both should give a reasonable number of favorable and unfavorable outcomes during a sequence of plays. Hopefully this will keep morale and motivation at a reasonable level. Bearing such things in mind, the pair of gambles shown in figure 6.2 were selected. These are two duplex gambles with positive expected value, one of which gives a small advantage with respect to payoffs while the other gives a small advantage with respect to probabilities.

The subject is required to play a series of sequential games, described in the previous section, with the selected pair of gambles. Over the whole series, the current capital before any choice and the outcomes of the previous one are not completely confounded as the stake given at the beginning of a game changes from game to game.

Some realism is injected into the situation by telling DM that after he has played all the games, one of them will be chosen at random and he will be paid the amount he won on that game. Should he be "ruined" in any game, that is, if his capital falls below zero then his winnings will be counted as zero and he will go on to the next game.

The next design problem is to decide how many trials per game and how many games to have. Preliminary experience led to the decision to use about 200 trials in total, as these could be comfortably played in about $\frac{3}{4}$ hour. If the number of trials in a game is small the sequential nature of the task is less important, and if it is too large the main variables will be more confounded. It was decided to play about 30 games of 7 trials in any session, bearing these points in mind.

The final decisions to be made are concerned with what stakes to use, and whether to select stakes randomly or according to some balanced procedure. It was decided to give DMs 3/-, 7/- or 11/- stake, and to attempt to control the effect of the order in which the stakes are given by a procedure due to H.Durap (1967). Now, the choice of stakes was partly determined by a desire to keep the probability of ruin (as previously defined) low. With 3/- stake the probability of ruin if gamble 1 is played 7 times is 0.164 and if gamble 2 is played 7 times it is 0.140. These are, perhaps large enough to affect DM's assessment of probabilities of entering future states which as discussed before, is undesirable. This point will be discussed more fully later. An advantage of using a stake as low as 3/-, though is that the subjective difference between winning about 3/- and winning about 11/- is likely to be quite large. Such differences in current capital are likely to reflect real subjective differences.

Suppose each of the 3 stakes begins a game 9 times, and each time it appears, each possible combination of stakes in the previous 2 games appears only once. To meet this condition 2 "dummy" games must be added at the start of the session. Within this constraint a random order of stakes was used. Subjects played 29 games in all, the first two being considered as practice games to be discarded in analysis. A set of stakes satisfying the above condition is shown below:

7 7 7 11 7 7 3 11 7 11 3 11 11 3 7 3 7 11
11 11 7 3 3 11 3 3 3 7 7

Random sequences of this type were produced by a computer programme of the algorithm due to Durap. The main purpose of using this type of sequence was to avoid a preponderence of any particular patterns of stakes which may occur if an ordinary random sequence was used and which may produce artefacts.

Procedure.

Most of the details of the procedure were given in the last two sections. In this section it is summarized and details not already mentioned are added.

Subjects were told that this was an experiment where they would have to make choices between gambles for small amounts of money. They would win money, but the amount depended to some extend on luck and to some extent on which gambles they chose to play. There was a small chance that they would win nothing, but they would not lose. (In fact, for the 10 subjects, the median winnings was 11/-, the lowest 7/- and the highest 18/-).

The duplex gamble was introduced to them with the apparatus used for this purpose in experiment 3. They played a sequential game of 5 stages, and it was explained that the result of the game was what they finished with, not the difference between initial and final amounts. They were told that the experiment consisted of playing a series of such games with 7 plays per game. Their stake would be 3/-, 7/- or 11/-, varying from game to game. If they fell below zero the game would be ended and they would go on to the next game. The results of all games would be recorded and at the end, one result would be picked at random which would be the subjects' wages for participation.

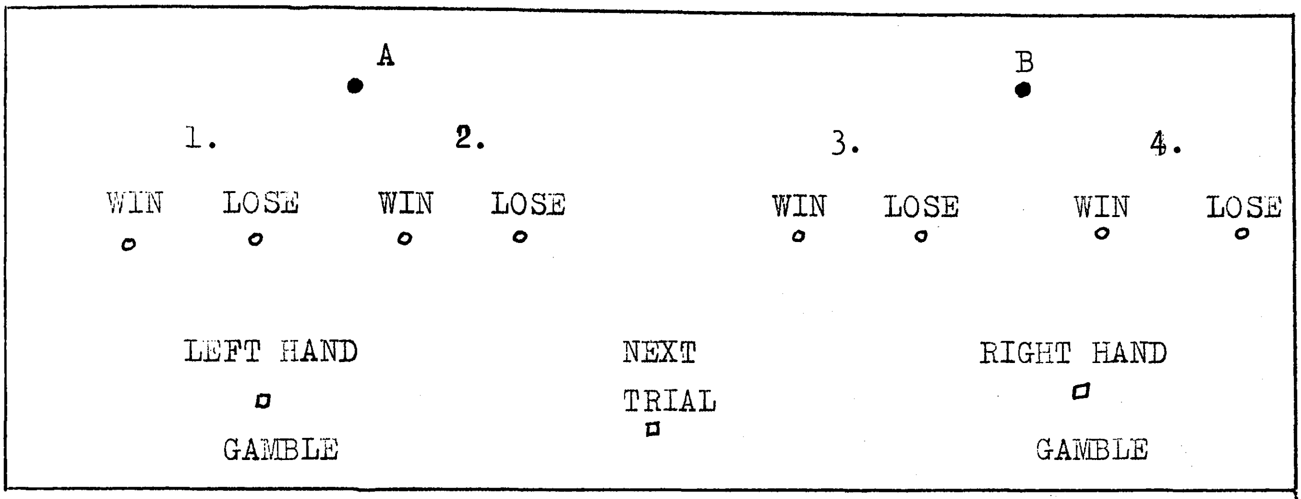
They were then told that to save time, and to enable the experimenter to keep an accurate record of the experiment, they would play the gambles with a machine. One of the gamble information cards was produced and they were told that these were the alternatives they would have to choose between. The way the machine worked, and how their current capital was recorded was explained to them. They played a trial game of 7 stages starting with 7 shillings. Probabilities of events in this trial bore no resemblance to actual ones. The trial was intended to familiarize subjects with the procedure. All combinations of win/not win and lose/not lose occurred. Subjects had to be told the significance of no light coming on after they responded (it meant they had neither won nor lost). Assurances were given that the game was not rigged. It was stressed that paying with the machine was "just the same as if the roulette wheels were spun on each play".

As a final check on whether they understood duplex gambles, 3 cards of pairs were shown to them such that one gamble dominated the other. All subjects at this stage chose the dominating gamble.

During the experimental session, the subject sat facing the control panel on the desk before him. The information cards were one on top of the other in front of the panel and the "current capital meter" was to its left in a prominent position. As the blue light changed from A to B, the subject placed the appropriate information card on the top. The experimenter sat opposite the subject during the session, checking the workings of the apparatus, setting the current capital before each game, altering the meter according to the outcomes and recording the results of each game. Unavoidably, therefore, he was an integral part of the experimental situation. Subjects invariably talked about how they were doing, and sometimes attempted to sound out about whether their choices were approved. The experimenter's general attitude was to indicate they were to decide according to their own wishes. Apart from this a low profile was maintained. Experimental sessions lasted about one hour, each subject receiving a different sequence of wins and losses. Subjects names were taken at the end, and the amount they won determined (in the way previously described). Their winnings were forwarded to them a few days after the whole experiment was finished.

Figure 6.1.

A Representation of The Control Panel for the Sequential Gambling Game.



▫ RESPONSE BUTTONS

• BLUE LIGHTS (information card cues)

∘ OUTCOME LIGHTS

The significance of the various lights and buttons is explained in the text.

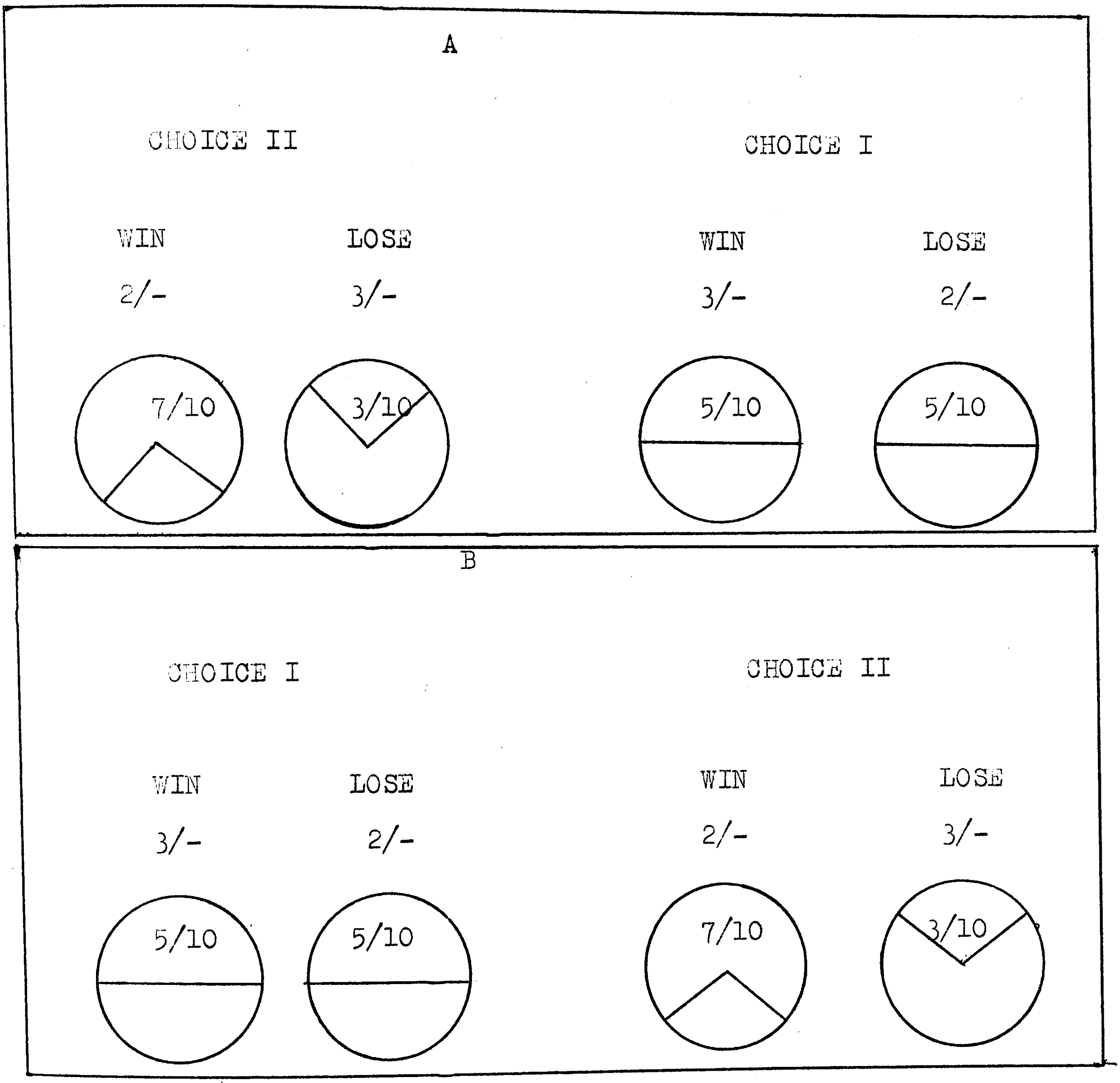
Table 6.1.

The connection between the information card used, the choice made and the relevant pair of outcome lights.

Information Card	Choice	Outcome Pair
A	gamble 2	1
A	gamble 1	2
B	gamble 2	4
B	gamble 1	3

Figure 6.2.

The Gamble Information Cards.



Results and Discussion.

In the experimental session, the games were logically, but not necessarily psychologically independent of one another. It is necessary, therefore to test whether the DM's preference probabilities changed as a function of the trial number in the whole session as well as whether they changed as a function of stage number within the games. In order to make the first test each subject's set of responses was considered as a whole sequence. Kendall's S statistic was calculated for each. Under the null hypothesis that the preference probability over the whole session is constant it is known that S has asymptotically the normal distribution with zero mean and known variance. The S-statistics, normalized to unit variance are shown in the "across games" column of table 6.3. On the basis of this test the null hypothesis cannot be rejected for any subject. To test that within each game, there is no trend in response probability the data can be considered as a game x stages, 29 x 7 matrix. Jonkheere and Bower's trend test can be applied to this. Tests for heterogeneous trend were not made as all 29 protocols consist of responses by a single individual. Jonkheere and Bower's S-statistic for each subject is shown in the last column of table 6.3. This different stationarity null hypothesis can be rejected at the 5% level for subject 8 and at the 1% level for subject 5. Subject 8 made very few '2' responses, nearly all of them early on in the games. Subject 5 chose both quite evently, but tended to make choice 2 more as the game progressed.

The next preliminary hypothesis to test is whether DM's responses are dependent on earlier ones. The whole sequence can be considered for this. On the binary sequence for each subject, one can test the restricted hypothesis that the sequence is an n-stage Markov chain against the more general hypothesis that it is an n + 1 stage Markov chain. Anderson & Goodman (1957) give asymptotic likelihood ratio tests for such alternative hypotheses. The likelihood ratio is transformed to a statistic which, under the null hypothesis that the higher-order Markovian model is no better than the lower varies asymptotically as chi-square with degrees of freedom equal to the difference between the number of parameters in each model. In table 6.4, H_2 represents the second order, H_1 the first order Markov model and H_0 the independent bernoulli model. The likelihood ratio tests show that H_2 is accepted over H_1 at the 1% level for subject 4 and that H_1 is accepted over H_0 at various levels for subjects 4, 5 and 8. For the remaining subjects, the simple bernoulli model H_0 can be accepted. Inspection of subject 4's responses reveals that his strategy was to decide which gamble to choose at the beginning of each game and stick to it throughout that game. Such a strategy obviously will lead to very marked response dependencies.

The preliminary analysis has shown that the basic bernoulli model can be accepted for all subjects except 4, 5 and 8, whose responses were sequentially dependent. (Subjects 5 and 8's responses were non-stationary as well as sequentially dependent). The latter group should therefore be excluded from the analysis of the affect of current capital and previous outcomes which is based on the bernoulli assumptions.

Each subjects' responses were put into a frequency table such as that shown in table 6.5 for subject 10. The

Current Capital x Previous Outcome, CC x O division is the basic one. Several S's protocols were examined, and it was decided to define the classes low, medium and high current capital as 6/- or less, between 7/- and 9/- inclusive, and 10/- or more respectively. Previous outcomes were divided according to whether overall DM had won, neither won nor lost, or lost. The nine cells of these tables all contained a reasonable number (more than 10) of observations. Let us now denote the levels of the factors by the suffices 0, 1 and 2. In table 6.6 the transformations of the frequencies required to make the analysis outlined in the introduction are shown for subject 10. The first two rows are the observed proportions of 1's and total number of observations in each cell. The third and fourth rows are the empirical logistic transform of the proportions and estimates of their variances.

Following Cox (1970), pp 30-40, an unweighted least squares analysis of the z's was carried out. The main effects and interactions were partitioned into components, each with a single degree of freedom, by the method of orthogonal polynomials. The main effects were partitioned into linear (L_{CC} and L_O) and quadratic (Q_{CC} and Q_O) components and the remaining effects were partitioned into their interactions ($L_{CC} \times L_O$, $L_{CC} \times Q_O$, $Q_{CC} \times L_O$, $Q_{CC} \times Q_O$). When these are normalized, under the hypothesis that all effects are zero, the 8 components for each subject should be random samples from the unit normal distribution. They are shown in table 6.7. One way to test the null hypothesis is to examine the extreme values of the set of contrasts for each subject. Only one of these, the L_{CC}

contrast for subject 10 is significant at the 5% level. For this subject, one can consider the remaining 7 values and test the extreme value in the same way. It is not significant. Tables for the above tests are given in Biometrika Tables for statisticions, Vol. 1, Pearson & Hartley (1958). Also for each subject, the absolute values of the contrasts can be ordered and plotted against the expected values of the order statistics. Such graphs are shown in figure 6.3. Subjects 4, 5 and 8 are included in this. Like most of the other subjects, the observed contrasts fall rather close to the expected values under the null hypothesis. The L_{cc} contrast for subject 10 is the only point which deviates markedly from the expected line.

Because of the above rather negative finding all manner of ways to check that O and CC did not markedly affect choices were tried. In particular, choices were ordered over the whole sequence according to what current capital DM had at the time of choice. Then, Kendall's S was calculated for the resulting sequences. Some other operations made on the data were as follows: Frequency tables of responses made for different categories of aggregate outcomes over the previous few trials were constructed. Frequency tables for the first response in each game, classified according to stake, and also for the last response in each game, classified according to outcomes in the game were both constructed. These manipulations of the data all failed to reveal any further positive effects of current capital or previous outcomes.

Thus it appears that the basic bernoulli model adequately accounts for most subjects' choices. The single parameter of this model is the preference probability for choice one over choice two. In table 6.8 estimates of it are shown for subjects except 4, 5 and 8. A Z-test, using the binomial approximateion to normality, of whether the values were significantly different from 0.5 was carried out. It can be seen that the hypothesis could not be rejected at the 5% level for subjects 1, 2 and 9. For these subjects it can be concluded that they were choosing randomly. Subjects 3, 6, 7 and 10 all had a preference for choice 2, indicating that they weighted probabilities more than payoffs. The results which support the bernoulli model also support the idea that static information integration models, such as the additive and SEU models can account for choice behaviour in sequential situations. All such models require that choice probabilities are constant, despite changes in sequential variables.

The results for the 10 subjects can be summarized as follows:

i) the choices of three subjects (4, 5 and 8) were dependent on trial number and/or the previous choice made.

ii) only one of the remaining subjects (10) was affected by either cc or o, and he appeared to be affected by cc.

iii) the remaining subjects' choices could be explained by the bernoulli model, three of them (1, 2 and 9) appeared to choose randomly and the other three (3, 6 and 7) showed a preference for choice 2.

Table 6.3.

Tests of Response Trends Both Within and Across Games.

Subject	Across Games, Z-scores	Within Games Jonkheere's z-statistic
1	-0.958	0.206
2	0.838	-0.498
3	-1.318	-0.229
4	-0.394	0.179
5	0.078	-2.663**
6	-0.193	-0.642
7	1.214	0.961
8	0.084	-2.450*
9	-0.481	1.149
10	-0.057	0.446

*Significant at the 5% level,

**Significant at the 1% level, 2 tailed.

Table 6.4.

Likelihood Ratio Tests for First and Second Order Response Dependencies, Across Games.

Subject	$X^2(H_2,H_1)$	$X^2(H_1,H_o)$
1	0.31	3.10
2	0.40	2.36
3	2.48	3.65
4	13.20**	97.37***
5	0.43	5.13*
6	5.06	0.06
7	2.71	0.03
8	0.36	9.67**
9	2.06	0.11
10	2.71	1.23
Degrees of Freedom	2	1

***) indicates the .1% level.
**) indicates the 1% level.
*) indicates the 5% level.

Table 6.5.

The Frequency of Subject 10's Responses Classified According to Response, Current Capital and Previous Outcome.

Response	Current Capital: Low			Medium			High		
	Previous Outcome								
	Loss	Zero	Win	Loss	Zero	Win	Loss	Zero	Win
1	4	2	2	2	2	1	8	13	24
2	12	10	20	6	6	17	9	12	35

Table 5.6.

Transformations of Subject 10's Data for the Unweighted Least Squares Analysis.

Transformation	Factor Combination								
	CC_0O_0	CC_0O_1	CC_0O_2	CC_1O_0	CC_1O_1	CC_1O_2	CC_2O_0	CC_2O_1	CC_2O_2
$p = n_1/n_1+n_2$	0.250	0.167	0.091	0.250	0.250	0.056	0.471	0.520	0.686
$n = n_1+n_2$	16	12	22	8	8	18	17	25	59
$z = \log\left\{\frac{n_1+.5}{n-n_1+.5}\right\}$	-1.02	-1.44	-2.10	-.96	-.96	-2.46	-0.111	0.07	-.37
$v = \frac{(n+1)(n+2)}{n(n_1+1)(n-n_1+1)}$	0.29	0.46	0.40	0.54	0.54	0.59	0.22	0.15	0.07

Table 6.7.

Results of the Unweighted Least Squares Analysis.

Contrast	Subject 1	2	3	6	7	9	10
L_{cc}	0.10	0.33	-1.54	-0.40	-0.23	0.10	-3.29*
Q_{cc}	0.68	-0.06	1.65	0.87	-0.31	0.45	1.31
L_o	-0.60	0.17	0.01	0.58	-0.13	-1.52	1.96
Q_o	0.44	0.62	1.80	2.36	-0.12	0.64	-0.92
$L_{cc} \times L_o$	0.87	-1.21	0.42	1.61	-0.69	0.18	0.83
$L_{cc} \times Q_o$	-0.32	-0.04	0.74	0.88	0.48	-0.95	0.20
$Q_{cc} \times L_o$	0.29	0.01	0.24	-0.13	0.68	1.44	-0.71
$Q_{cc} \times Q_o$	0.04	-0.69	-1.43	-2.22	0.06	-0.75	0.51

* Significant at the 5% level.

Table 6.8.

Bernoulli Model Parameters.

Subject	Estimate of p(1, 2)
1	0.441
2	0.568
3	0.351**
4	-
5	-
6	0.277***
7	0.360**
8	-
9	0.548
10	0.314***

***) Indicates that the parameter is significantly
**) different from 0.5 at the (.1% (1% level.

Figure 6.3.

Graphs of absolute standardized logistic factorial contrasts against the expected semi-normal order statistics.

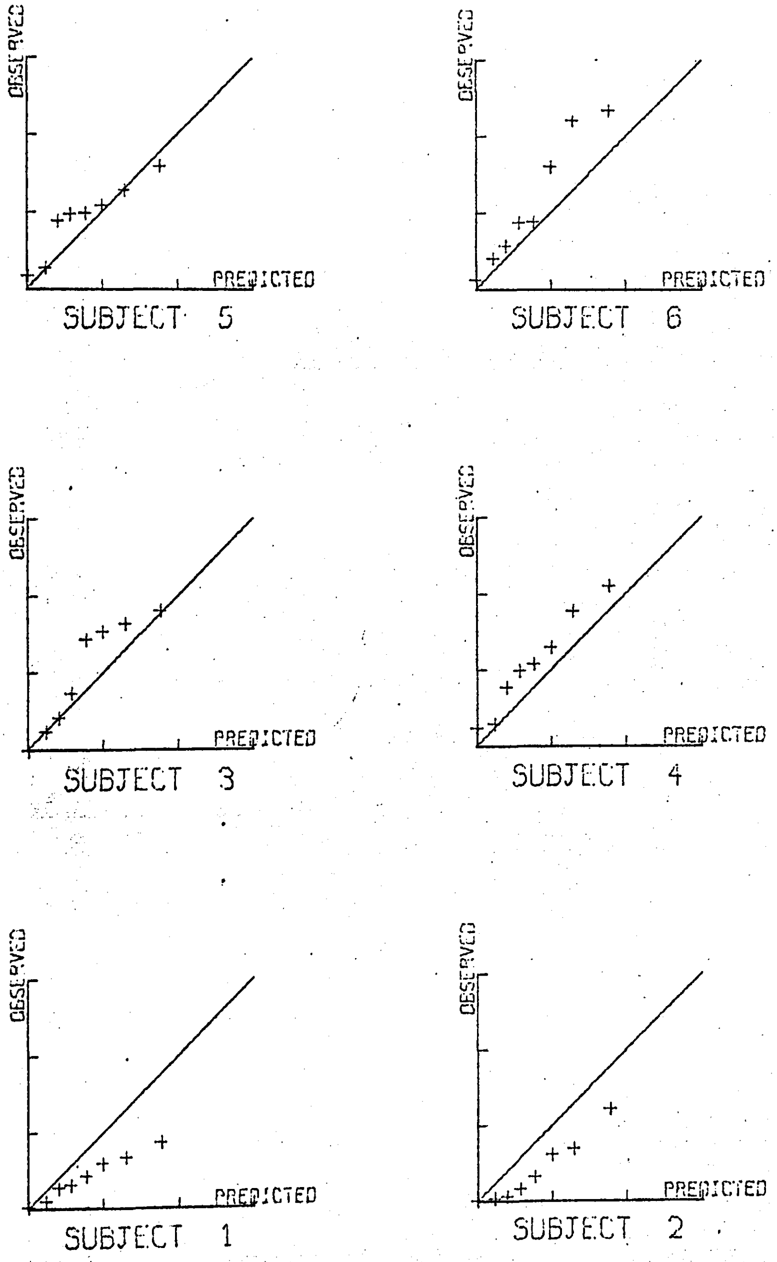
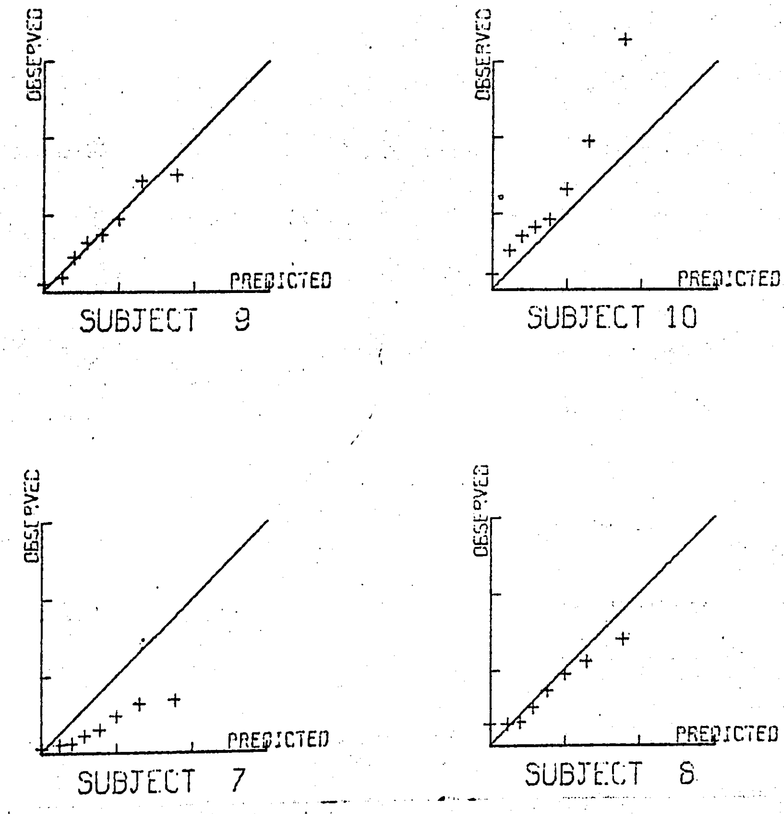


Figure 6.3. cont.



Experiment 5.

A Study of Information Integration Models in a Sequential Decision Making Situation : the effects of previous outcomes and current capital.

In this experiment we return to the incomplete, n-replicate pair comparison design, in order to study information integration models by functional measurement under "real play" conditions. The simplest way to do this would be to repeat experiment 1 or 3, but allow each gamble chosen to be played before the next choice is made. The subject could be given a stake at the beginning of the experiment and be allowed to take away with him what he ends up with after playing all the chosen gambles. This, essentially is what will be done. The gamble pairs of experiment 3 will be presented to the subject in real play situation. But, the change from the static to the sequential environment makes it preferable to change other features of the experiment. The sequential environment is that much more interesting for the subject, so it is reasonable to expect him to make more choices before becoming bored. So, a larger n in the n-replicate pair comparison experiment can be used, making the parameter estimates more reliable. Also suppose assumptions of the Thurstonian model do not hold in a sequential situation. In particular suppose the response probabilities change as a function of trial number, current capital, previous outcomes or some function of all three. In this case it would be inappropriate to attempt functional measurement, or any analysis based on an assumption of stable preference probabilities. Tests of whether systematic trends in preference probabilities occur, as trial number, previous outcome or current capital increases must be carried out. Thus the argument which led to the "blocking" of trials in experiment 4 applies. After blocking, previous outcomes and current capital will not be so highly correlated. Tests of their effect on choice probabilities can therefore be made more or less independently.

Experiment 3 gave encouraging results supporting information integration models in a static situation. In experiment 4 there was little evidence that the utility difference between a particular pair of gambles changed when previous outcomes or current capital did. The first aim of the present experiment is to see if this finding of experiment 4 can be extended to a wider set of pairs of gambles. If it can, it will then be possible to test the information integration models in sequential situations.

Experimental Design.

The design of this experiment can be nicely summarized. The design of the sequential decision making situation of experiment 4 is used, the difference being that the pair of gambles available to DM changes from trial to trial. The pairs used follow the incomplete n-replicate pair comparison design of experiment 3. The pairs are the 29 from the same set of duplex gambles with the probability of losing fixed at 0.5 that were studied in that experiment.

The experimental setting is identical to that of experiment 4, i.e. DM plays the mechanical version of the sequential gambling game. DM plays 58 sequential gambling games of seven stages per game. Instead of just two alternative gamble information cards being available, as in experiment 4, there are booklets of gamble information cards.

The booklets of gamble information cards are made up of the stimulus cards from experiment 3. It will be recalled that in experiment 3 there were seven cards for each of the 29 pairs, the 203 cards being in random order, except that no pair appeared if it had done so in the previous 10 trials. There was a different randomization for each subject, and for any pair, the juxtaposition of the gambles was random. In the present experiment, two of these sets of 203 stimulus cards provide the gamble information cards for each subject. Thus, DM chooses between each pair 14 times. The order of the 406 information cards is the same as in experiment 3, one set of 203 following the other. A slight reordering was made around the middle of the sequence to make sure that no pair appeared if it had done so in the previous 10 trials. With this order, the cards were split into sets of seven, giving the basis of the 58 gamble information booklets. To each booklet a page was added at the beginning giving the stake with which DM would start that game, and the number of the game.

Three stakes were used, 8/-, 13/- and 18/-. The same algorithm was used to give the order of stakes, balancing for sequence effects, as was used in experiment 4. Two orders, such that each stake was preceded by every combination of stakes in the previous two games were used. When the "dummy pair" of stakes are included at the beginning of each order this gives 2 sequences of 29 stakes. As this device was not crucial to the experiment - merely spreading the stakes a little better than complete randomization - the games included simply to make the balancing for sequence effects complete were not ignored in the analysis. The particular levels of stake were selected to give a reasonable spread of current capital and a small probability of ruin. As before, ruin is defined as occurring when ones capital falls below zero. Thus with 8/- stake, one will always play at least 3 of the 7 gambles in a game. To play fewer than four gambles in any game one would have to be quite unlucky. Hopefully, then, DM does not pay too much attention to the possibility of ruin and we shall not lose much data.

Subjects.

Five undergraduates at Stirling University, volunteers who were told they would be paid took part in individual experimental sessions. There were 3 female and 2 male students, between 18 and 22 years.

Procedure.

Subjects were introduced to duplex gambles in the usual way. Then, a booklet of 5 duplex gamble pairs, where one dominated the other was produced. On the first page was written "practice game", and "stake 8/-". DM played the sequential game, working through the booklet choosing the dominating gamble of each pair, which was then played using the roulette wheels. If the non-dominating gamble was ever chosen more explanation about duplex gambles was given and the game was played again. It was explained that the result of a game was what DM ended up with, not the difference between final and initial cash.

Then DM was introduced to the mechanical gambling apparatus. He played the practice game on it. The experimenter set the current capital meter to 8/- before he started. The blue button A was on at the start of the game, and the tape which controlled the apparatus was such that light A would always come on when DM pressed the "next trial" button. Its' significance in this experiment, then is to inform DM that he can make the next choice whenever he is ready. DM put the booklet on the desk below the control panel, open at the first gamble pair. To choose the gamble on the left, he pressed the left hand button, when the pair one outcome lights are relevant. Pressing the right hand button signified that he chooses the right hand gamble and pair two outcome lights become relevant. Blue light B and pairs 3 and 4 of outcome lights are redundant in this experiment. In the practice game, the controlling tape was written such that DM would experience all kinds of outcomes. Misunderstandings about the situation were cleared up and DM was ready to start the experiment proper after this.

The experiment proper was carried out in exactly the same way as experiment 4, except that at the start of each game, DM placed a new booklet below the control panel and called out the initial stake for that game. The experimenter set the meter, and an additional duty was to check that the next pages of the booklet was turned up prior to DM choosing. Since the order of gamble pairs was known it was relatively straightforward to sequence the control tape for the apparatus so that the outcomes occurred with the probabilities given on the card.

In this experiment, sessions lasted about 90 minutes. Subjects' winnings were determined in the same way as in experiment 4. They collected them a few days after the completion of the experiment.

Results and Discussion.

Each subject chose between each of the 29 pairs of gambles 12 to 14 times. Subjects were ruined only occasionally. In all, only 9 observations were lost from the 5 14-replicate pair comparison experiments. The first task was to make 3 data matrices for each subject. His responses (whether he made choice one or two, as understood in experiment 3) to each pair were ordered in different ways in each matrix: with respect to trial number, magnitude of previous outcome and current capital. This gave 3, 29 x (approximately) 14 data matrices. If the 29 pairs of gambles label the rows of the matrix the problem is to discover whether, as one moves across the rows, there is a tendency for the probabilities of "one" responses to change. That is, to discover whether response probabilities change as trial number, magnitude of previous outcome or current capital increases. Now, direction of change is of no interest since the labelling of responses as "one" or "two" is arbitary. Thus, Jonkheere and Bower's heterogeneity of trend statistic, W is more appropriate than their trend statistic Z. The assumptions under which W approximates the chi-square distribution when the null hypothesis holds are not met. Therefore, for all 3 data matrices for each subject the transformation of W suggested by Jonkheere and Bower was calculated. Since, under the null hypothesis that response probabilities are constant within rows this is distributed as the unit normal it can be denoted by $\mathcal{Z}$. These $\mathcal{Z}$ - scores are tabulated in table 6.9. None reach significance at the 5% level (two-tailed). For certain $\mathcal{Z}$'s the probability that a value as extreme as the $\mathcal{Z}$ would occur under the null hypothesis are given. It can be seen that the most doubtful Z's under the null hypothesis are subject 1's value for current capital ($p = .0758$) and subject 2's value for previous outcome ($p = 0.0644$). According to this test, then, the above subjects may have been influenced by the sequential factors indicated, though the evidence is not too conclusive. There are no grounds for supposing that response probabilities changed with time, or that the sequential factors affected any other subjects.

For all subjects, then, it is reasonable to procede with the functional measurement analysis. The goodness of fit of the basic Thurstone model is shown in table 6.10. Since none of the error chi-squares reach significance it can be concluded that the basic Thurstonian model accounts for all 5 subjects responses. Somewhat surprisingly, the estimation chi-squares also fail to reach significance for 4 of the 5 subjects. For these subjects the hypothesis that they are choosing randomly cannot be rejected. Subject 5 was the only one who appeared to show preferences for some gambles over others. This is the only subject for whom it is necessary to consider which multifactor submodels best account for his preferences. With his data, let us consider certain additive and SEU information integration models. This parallels the analysis carried out in experiment 3. Following this, the random bernoulli model, which accounts adequately for the remaining subjects indifferences will be discussed.

As before, the two main hypotheses to be considered are the general additive and SEU models. The former predicts that only main effects are significant and the latter predicts that the bilinear, SW x PW interaction is the only significant interaction. After Bock and Jones' procedures are applied to the Thurstonian subjective value estimates and their variance-covariance matrix, the Thurstonian estimation chi-square can be partitioned. The partition for the general additive and SEU models is shown in table 6.11, for subject 5. The main effects, bilinear SW x PW interaction and remainder interaction (error) chi-squares are all significant at least at the 1% level. The significance of the sum of the interactions shows that the additive model can be rejected and the size of the bilinear SW x PW interaction shows that the SEU model is significantly better than it. Unfortunately, the remainder of the interaction is significant and the SEU model, on this basis should also be rejected. But, it is clear that the SEU model, with only seven parameters accounts for most of the variation in the estimated subjective values. With the two step procedure adopted there are two chi-square components due to error. The first was obtained during the estimation of the subjective values and the second is what was called the remainder interaction above. An F-test of whether these two chi-squares could have come from the same population can be carried out. The value obtained, $F = 4.58$ is about at the 5% significance level (degrees of freedom 17 and 5) which is not conclusive evidence against the hypothesis. Thus, if the basic Thurstonian scaling model is accepted there is no conclusive evidence that the remainder interaction chi-square did not occur at random. It is reasonable, then, to accept the SEU model, especially as no more general information integration models have been considered.

Further support for the model comes from figure 6.4. Here subjective values predicted by the SEU model are compared to those estimated directly from the data. It can be seen that the points lie along the expected line.

Also as in experiment 3, Slovic's additive model was compared to the simple linear SEU model. The latter has an additional parameter, which can be referred to as the SW x PW interaction. The partition of the Thurstonian estimation chi-square with respect to these two models is shown for subject 5 in table 6.12. The main point of interest here is that the linear SEU model is significantly better than Slovic's additive model, though neither can be accepted. The latter part of this conclusion is made on the basis of the significant error chi-squares for both models and the former part on the basis of the significance of the SW x PW interaction chi-square. The linear SEU model has never seriously been proposed so it is not necessary to defend or attack it. The analysis of the two models based on the objective values can rest here, therefore, with Slovic's additive model rejected.

As well as the preceding findings for subject 5 it has been shown that the other 4 subjects apparently chose randomly. This is not the first time such a result has been found: the same was true for 1 subject in experiment 3 and 3 subjects in experiment 4. "Random" results in sequential situations have served to emphasize the negative findings regarding sequential factors. In the present experiment the gamble information should be the main determiner of choice. The single subject who was influenced by it was not influenced by the sequential factors. Also, the four subjects who appeared to be indifferent to changes in gamble information also appeared to be influenced little by previous outcomes or current capital. Thus, when the primary factors under investigation had little effect it was not because of the effect of the secondary, sequential factors. These results all support the bernoulli model and thereby the idea that static information integration models can account for choice behaviour in sequential situations.

Before leaving experiment 5, the extent of bias in the estimates of the subjective values of the alternatives will be examined for subject 5 in a short Monte Carlo study.

Table 6.9.

Jonkheere and Bower's Heterogeneity of Trend Statistic, Z calculated for each subject, to determine heterogeneous trends with respect to trial number, current capital and previous outcomes.

Subject	Trial Number		Current Capital		Previous Outcome	
	z	p	z	p	z	p
1	0.548	-	-1.844	.0758	-0.388	-
2	0.200	-	0.054	-	1.848	.0644
3	-1.260	-	1.438	.1498	1.159	-
4	1.264	.2076	-1.257	-	0.004	-
5	0.588	-	-0.649	-	1.536	0.126

The p columns give, for selected z's the probability that a value as extreme as the z would occur under the null hypothesis.

Table 6.10.

Goodness of Fit of the Basic Thurstone Model.

Subject	Chi-Squares		
	Estimates	Error	Total
1	23.25	2.86	26.11
2	18.50	9.41	27.91
3	25.87	2.28	28.15
4	32.59	7.74	40.33
5	122.77***	2.28	125.05
Degrees of Freedom	24	5	29

*** indicates significance at the .1% level.

Table 6.11.

Goodness of Fit of the General SEU and Additive Models for Subject 5 : Partition of the Estimation Chi-square.

	Main Effects	Bilinear SWxPW interaction	Remainder interaction (error)
Component of Chi-square	68.86**	18.40**	35.56**
Degrees of Freedom	6	1	17

** indicates significance at the 1% level.

Table 6.12.

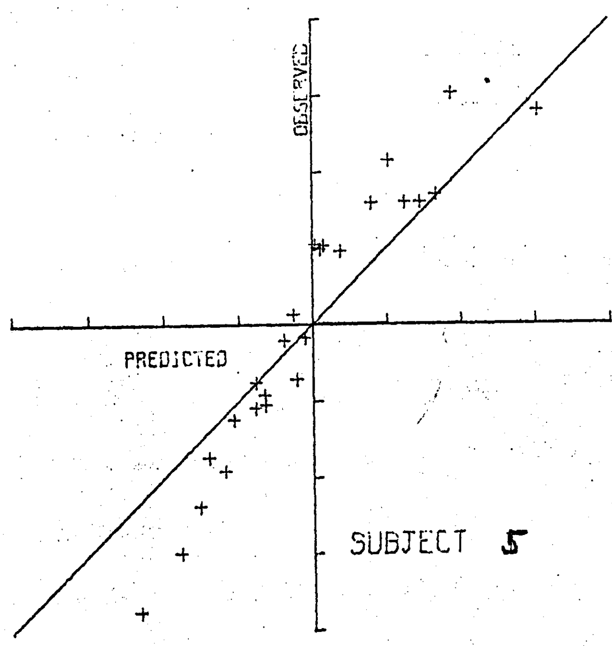
Goodness of Fit of the Linear SEU and Slovic's Additive Model: Partition of Thurstonian Estimation Chi-square.

	Main Effects	SWxPW interaction	Remainder interaction (error)
Component of Chi-square	52.10**	25.19**	45.48**
Degrees of Freedom	3	1	20

** indicates significance at the 1% level.

Figure 6.4

Graph of subjective values estimated directly from the data against those predicted by the general SEU model for subject 5.



A Sampling Experiment to Examine the Assumptions Underlying the Analysis of Experiment 5.

The purpose, and method of execution of this sampling experiment were the same as those of the simulation of experiment 3. It was hoped to test whether the bias in estimates of subjective value was small compared to the sampling variance and if the distribution of the goodness of fit statistic, SSE approximated the theoretical distribution. Data in the 14-replicate pair comparison situation of experiment 5 was generated under Thurstone case V assumptions. The assumed subjective values of the processes were those estimated for subject 5 in the experiment. As with the previous sampling experiment mean estimates and variances of 250 simulations were calculated. The 250 values of SSE were recorded in a histogram.

Figure 6.5 shows the mean estimates plotted against the actual subjective values and figure 6.6 shows the sampling variances plotted against the theoretical ones. Compared to the 7-replicate situation of experiment 3 the extent of bias in estimates was much less and the estimates themselves were more reliable. Somewhat surprisingly, however the sampling distribution of the goodness of fit statistic, SSE approximated the theoretical distribution worse than was the case in the earlier simulation. Most observed values were between 0 and 3 and the distribution had a large "tail" area, 10% of values being greater than 18. The goodness of fit test, then appears to be an even more conservative test than was the case in experiment 3.

Apart from the poor approximation of the sampling distribution of SSE to the theoretical the design with n = 14 seems to enable a far more reliable functional measurement to be carried out. Where it is possible, then it would seem to be desirable to use an n of at least 14 in incomplete, pair comparison designs where Bock & Jones' functional measurement procedures are to be applied.

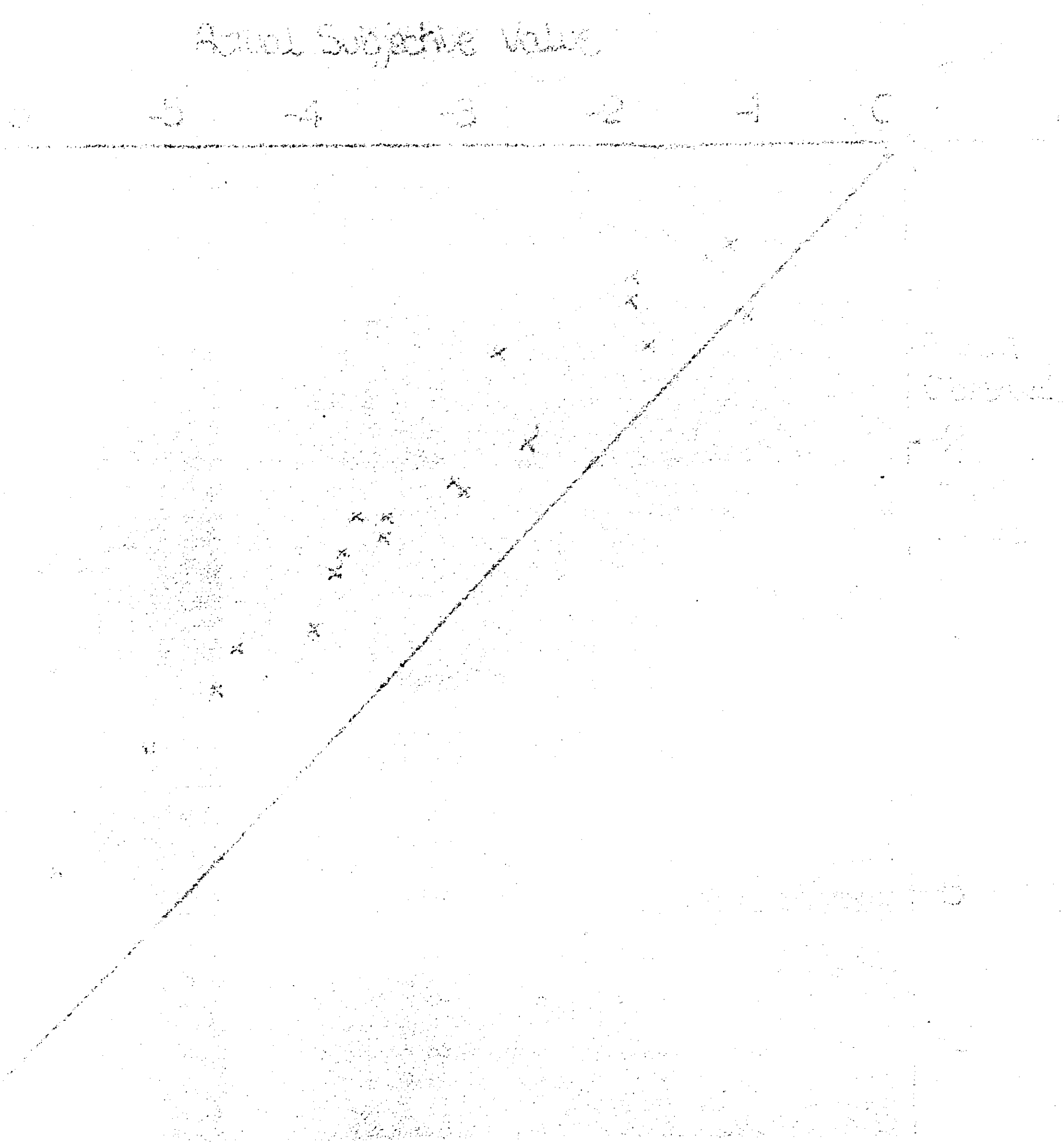


FIGURE 6.5

Graph of the Mean Observed Subjective Values against the Actual Values in the Simulation of Experiment 5.

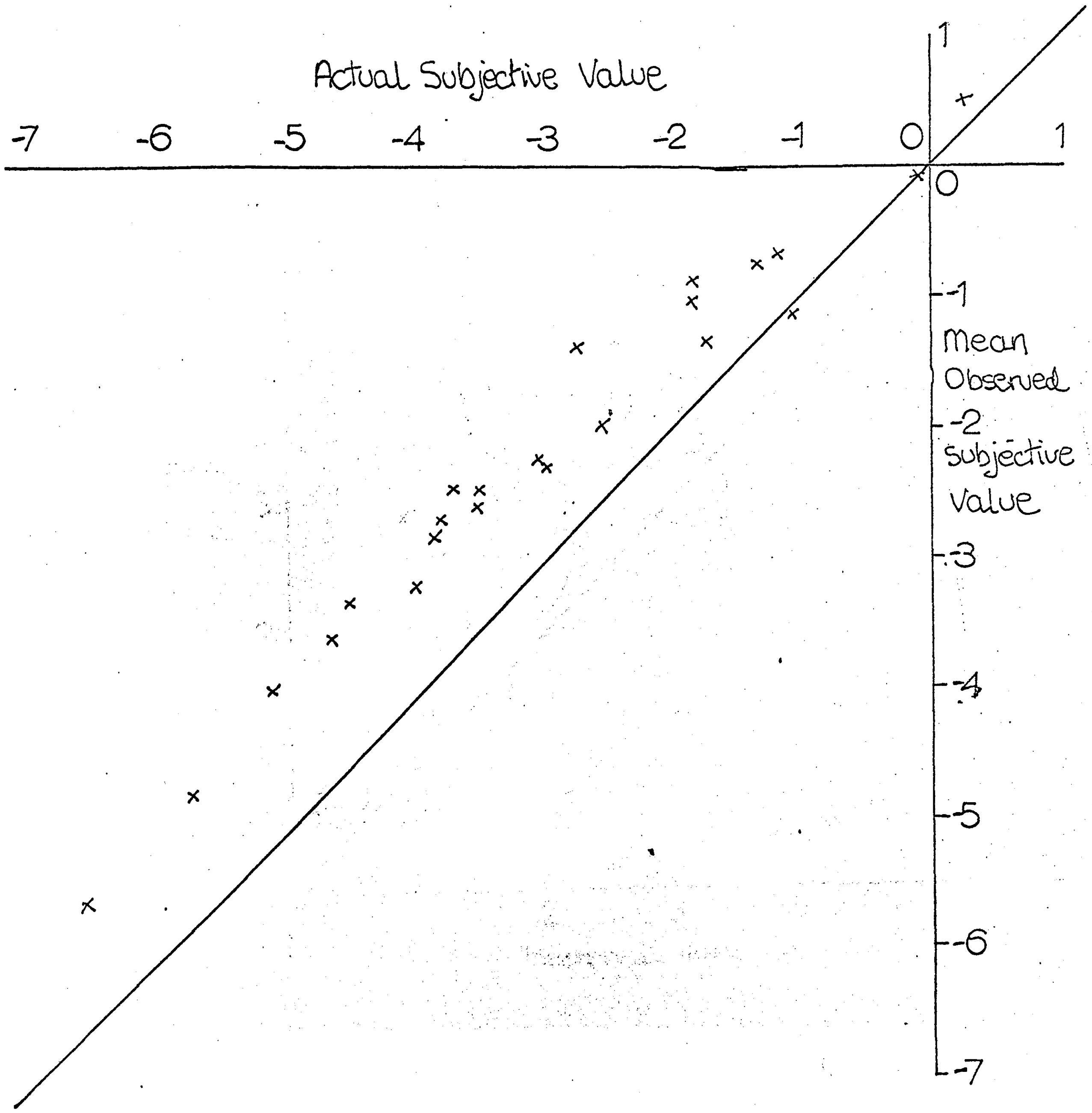


FIGURE 6·6

Observed Sampling Variance of Subjective Values against Theoretical Variance in the Simulation of Experiment 5.

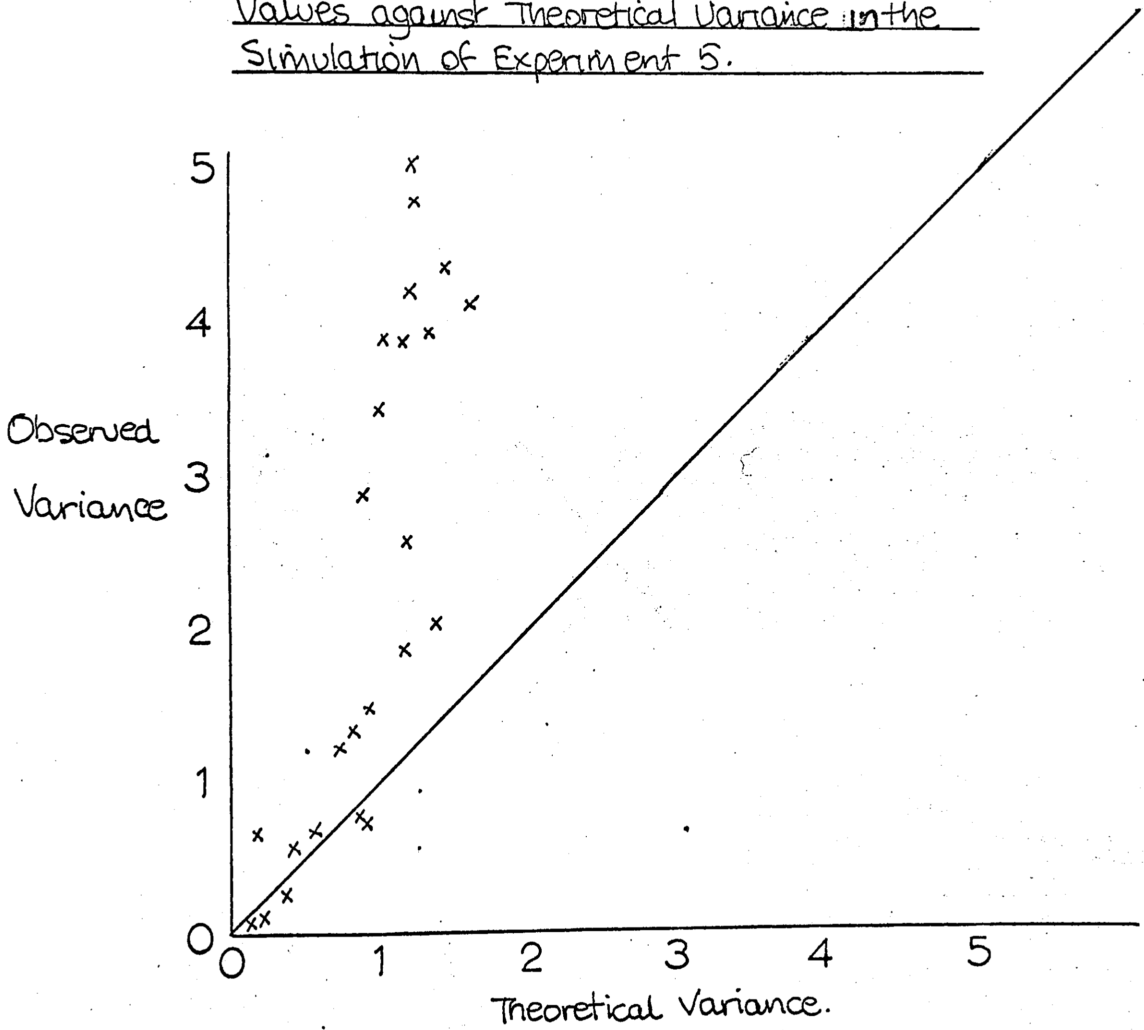
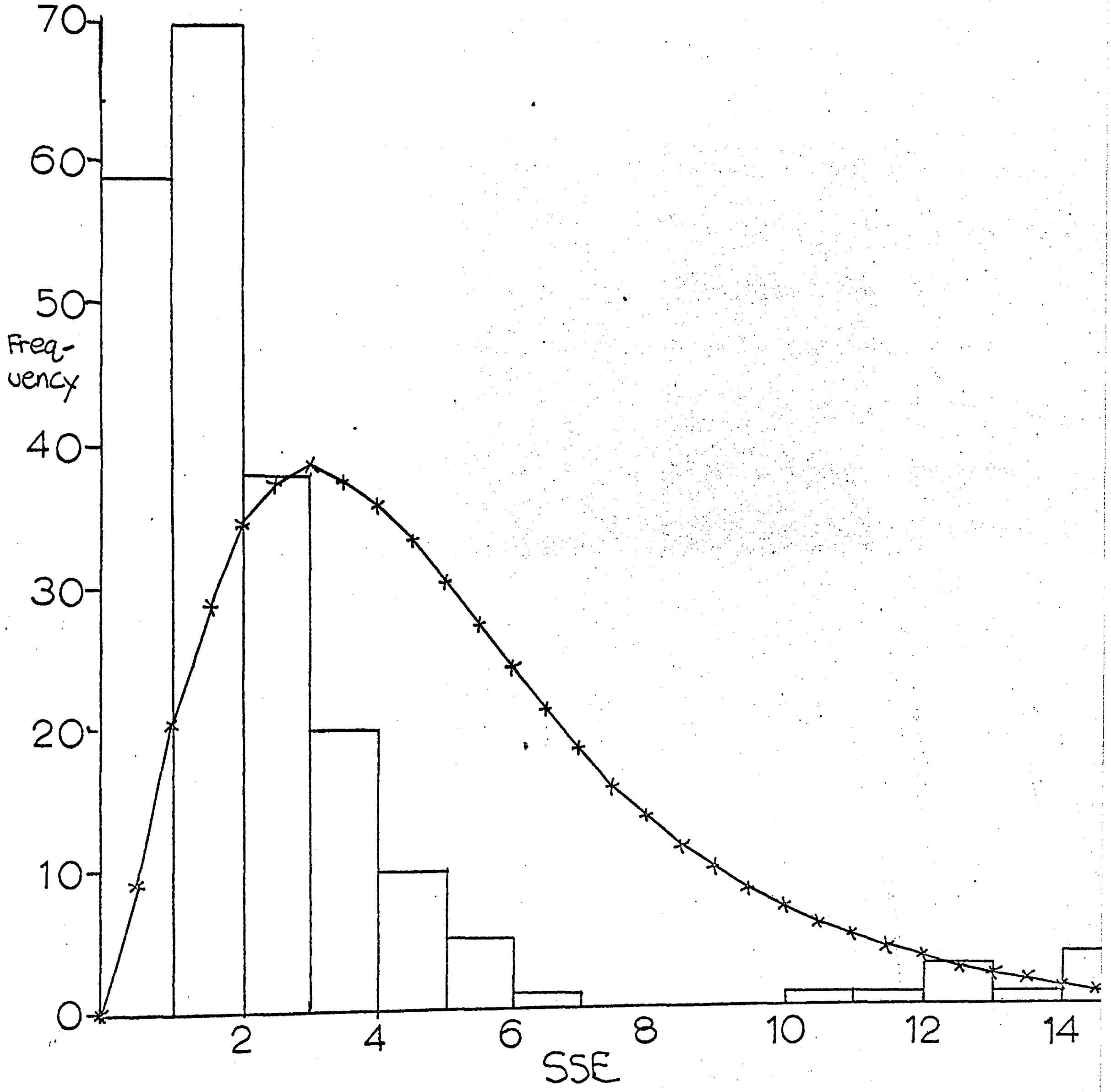


FIGURE 6.7

Graph of the Theoretical Frequency Distribution of SSE and the Histogram of SSE in the Simulation of Experiment 5.



Chapter 7.

A Dynamic Experiment.

The discussion so far has centred around simple information integration models of decision making in static situations and examining them by QFA and functional measurement. When sequential decision making situations were studied these same information integration models applied, because the alternatives themselves could still be described by the same information dimensions. A "dynamic" alternative needs additional dimensions to complete its description, however. In fact, for an n-stage dynamic decision making situation, the number of dimensions required to describe the simplest alternative is rather large. It would be difficult to give a comprehensive examination of additive or SEU information integration models in the dynamic case. The more realistic aim of the experiment to be described is to consider a 2-stage situation and to examine how some of the information dimensions are integrated while others remain fixed.

Consider the career choice problem discussed in the introduction. The information that DM considers initially is about the payoffs and probabilites of the first and the second stages and also the probabilities of being in the different second stage states. A betting game, with the same basic characteristics could be constructed. The time scale would be much reduced, of course, and the payoffs would be on a single dimension. Nevertheless, interesting conflicts between immediate advantages and temporally more distant ones or between immediate advantages and the probabilities of going to favorable decision states could be constructed. Such a game is represented in figure 7.1.

This is a two stage, decision dependent betting game in which DM has complete information about all aspects. He begins the game at A with a certain stake, in new pence. He must choose between the gambles A1 and A2. Both are bets of the (x, 1/2, y) type, where x pence is won with a probability 1/2, otherwise y pence is lost. When he chooses A1 or A2, he also chooses the wheel of fortune associated which determines whether he goes to B or C in the second stage. In B his choice is between the bets B1 and B2, in C it is between C1 and C2. These, in general are bets of the type (x,p,y) where x is won with probability p, otherwise y is lost with probability (1 - p). Initial conflicts between proximal advantages (payoffs) and distal advantages (probabilities and payoffs) are determined by the parameter values of the gambles at both stages. The degree of dependence of later choices on earlier ones is determined by the probability of going to B (and conversely C), the parameters of the wheels of fortune associated with the alternatives in A.

In this game, although choices and events occur in quick succession, the basic conflicts common to all dynamic decision situations are present. A rough information processing model can be proposed. It seems likely that people will start at the final stages and work backwards when considering the initial choice. That is, they will first decide which second stage they prefer to be in. Then, they decide which first stage alternative they prefer. This decision will involve the first stage information, with the probability of going to their preferred second stage state included as a dimension of the alternatives to be weighed up. Models of decision making for such alternatives, similar to the information integration models for static alternatives can be considered. It would be possible to consider them as representation hypotheses and study them by QFA, or as multifactor submodels of basic Thurstonian models, to be studied by functional measurement. At the present time the latter is more feasible.

Let us consider again the game depicted in figure 7.1. Suppose different games are constructed in which the second stage states all remain fixed, and as shown in the figure. Suppose also the probability of a win is 0.5 in stage A for both alternatives. Now, subjects are invited to play a series of games in which the three parameters of each alternative in A, amount won (SW), amount lost (SL) and chance of going to A(PA) all vary from game to game. The SEU model, in this case, corresponds to an additive one. If the subjective value of a gamble alternative in A is denoted by S(G) and subjective scales of the three variable parameters by U_i, i = 1, 2, 3 then the general SEU, or additive model states that: $S(G) = U_1(SW) + U_2(SL) + U_3(PA)$. The simple linear SEU model also corresponds to an additive model and states that: $S(G) = W_1.SW + W_2.SL + W_3.PA$, where the W_i's are parameters of the model and SW, SL and PA are the objective values of the gamble parameters. They can be put in this simple way, of course, only because other features are fixed. The aim of the experiment to be reported is to test the above models by Bock and Jones method of functional measurement. It is intended as a small step towards the descriptive study of dynamic decision making.

Experimental Setting.

It is clear that to test the models discussed it will be necessary to have an efficient way of playing the games, if a real play situation is contemplated. Of course, it would be possible simply to describe the game, and then show subjects information sheets about which to make decisions, in the manner of experiment 3. It seems more likely that subjects will understand the alternatives and will give highly motivated responses if they are involved in a real play situation, however. The most efficient way available for this was to automate the game using the Elliott 903 computer of the Department of Psychology, University of Hull.

To explain the automated game, the following is a description of the events in a practice session which was used to introduce the experiment to the subjects. The subject sat before a teletype which was on-line to the computer. On a desk to his left were information sheets for 5 games. They were in a stack, numbered in the top right hand corner. It was explained to the subject that in the experiment he would be asked to play a series of betting games. Before each game he was to suppose that he had a stake of 105p. The significance of the information sheet was explained to him, and the computer was programmed. A flow diagram of the events in the game is shown in figure 7.2. A final printout of a game is illustrated in figure 7.3. After any game, as shown in the flow diagram, the next game begins. The subject removes the previous information sheet to reveal the next relevant one.

Now, the computer programme controlling the sequence of events shown in figure 7.2 was written in algol, and an important feature of it is the way the computer "played" the gambles. Uncertain events were determined as follows: the probabilities and payoffs were fed in as data via the tape reader before every game. Before any uncertain event a random number in the range of 0 - 1 was selected and compared to the probability of one of the two possible events. If the number was less than this probability then that event occurred, otherwise the complementary one occurred. The random number procedure was a standard Elliot one, based on the Lehmer congruance method. By a parameter read in at the start of the programme it was possible to select one of 64 strings of random numbers. Thus, sessions with the same order of games could be made to have different outcomes. At the end of the session, the subjects' first-stage responses are output to the computer tape punch. The programme controlling the experiment carries out some re-ordering of this data to enable the experimenter to quickly assess the data. After the data has been output in this way "FINISH" is printed on the teletype. The subject knows the session has ended as (hopefully) he has no more information sheets before him. This then is the setting in which the experiment takes place. It has a great advantage over the real play conditions of the last two experiments in that it is fully automated and the experimenter need not be present. As with the earlier real play settings, however it is essential that subjects accept the veracity of the data on the information sheets.

Experimental Design.

The stimuli were roughly described earlier. DM's play two-stage games which involve choices between certain gambles. The second-stage gambles are always the same, those shown in figure 7.1, and the information integration models of interest are relevant to the individual's initial decisions in the games. The Bock and Jones' analysis requires a pair-comparison experiment where pairs of alternatives from some set are repeatedly presented. Now, any alternative in A can be fully characterized by the values of the parameters SW, SL and PA when other features of the game remain fixed. It was decided to consider the set of alternatives which consisted of every combination of: SW, SL = $\{65p, 40p, 15p\}$ and PA = $\{.75, .50, .25\}$. This experiment resembles experiments 3 and 5, therefore, in that alternatives are from a 3 x 3 x 3 product set. The problem of choosing a subset of pairs in the incomplete pair-comparison design is therefore similar. The solution decided on, for reasons, discussed in the earlier experiments was to use a cyclic chain design. That is, within each level, as previously understood, a cycle of pairs was selected, each cycle linked by another pair. In the light of experience with this design it was decided to include an extra pair between each cycle, thus strengthening the links between them. Hopefully, this will make the scale estimation more reliable. It also gives more degrees of freedom for testing the basic Thurstone scale. The 33 pairs to be used are illustrated in figure 7.4.

Also in the hope of improving the reliability of the estimation from that of experiment 3, an n of 12 in the n-replicate experiment was decided upon. This means the experiment involves each DM in playing 12 x 33 = 396 games. It was felt that these could be played in 3 sessions of 132 games. An additional methodological point of interest, therefore will be to see if trends appear over the 3 sessions.

In each session, the same order of games was used. The set of 33 pairs was presented four times in order. Randomization for order was within each set of 33 pairs of alternatives, unlike earlier experiments. Between replications, care was taken to ensure that no game was played on any trial if it had been played in the previous 10 trials. Subjects were told that they had 105p stake with which to play each game (note they were not actually given it). Current capital was therefore controlled by keeping it constant for each game. With this stake there was no possibility of ruin. Also each game was logically if not psychologically independent.

The procedure used so that there were real consequences of the decisions was as follows: After each session a game was picked at random and the result of this game recorded. Subjects were paid the average of these three results. Thus, payoffs could vary between 0 and 210p, but because of the averaging were unlikely to approach the extremes. The computer used a different string of random numbers to generate events in each session of the experiment. Thus, in each session outcomes would be different even if subjects made identical sequences of choices.

Prior to each subject's first session, a preliminary set of five games was played. This was to familiarize subjects with the games and the structure of a session.

Subjects

Five post graduate students, all from Hull University took part. They were all volunteers who were told the nature of the experiment, its probable duration and payoff. They were all male, between 21 and 24 years.

FIGURE 7.1

Information Card for a 2 - Stage Betting Game.

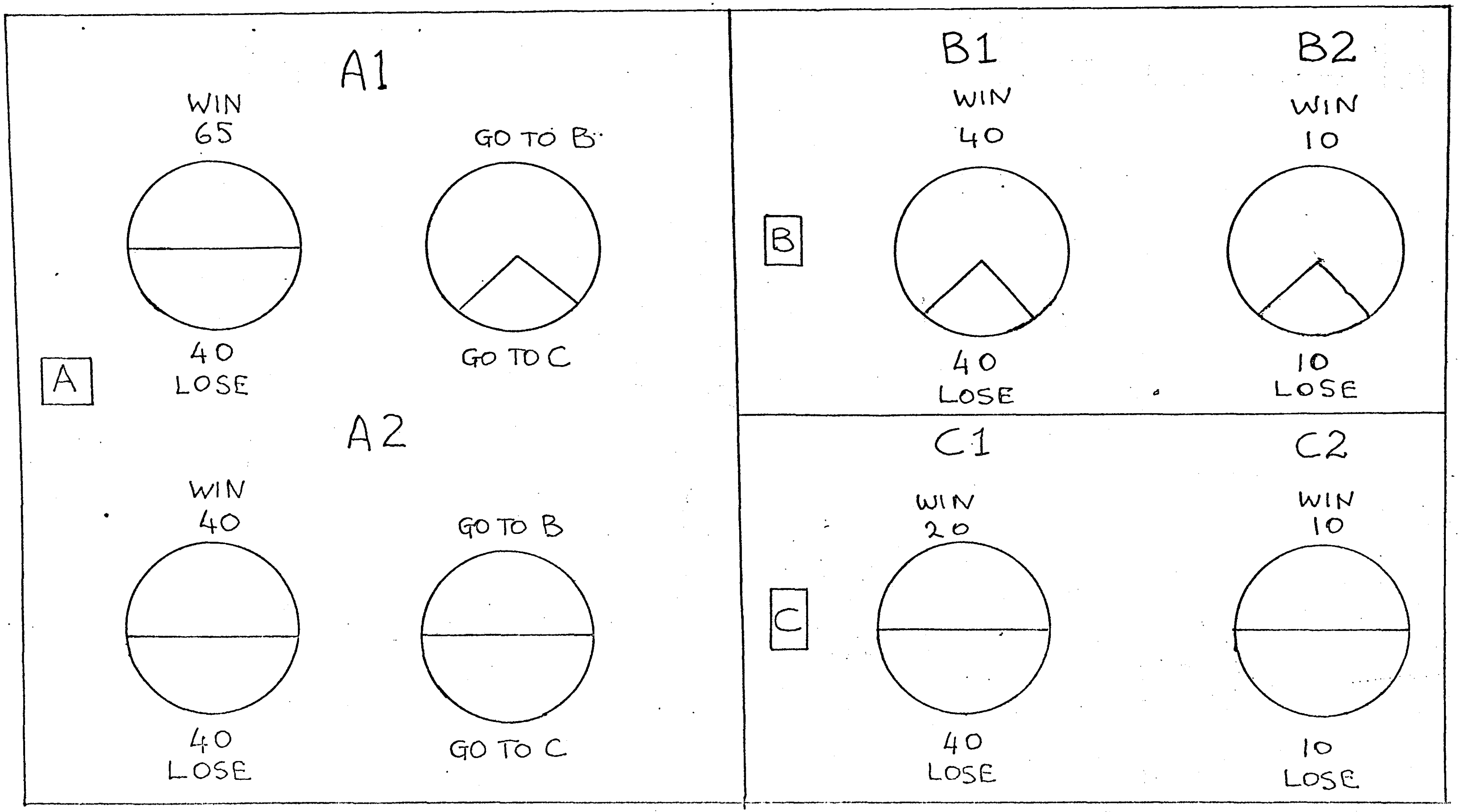


Figure 7.2.

Flow chart of the events in the two-stage betting game.

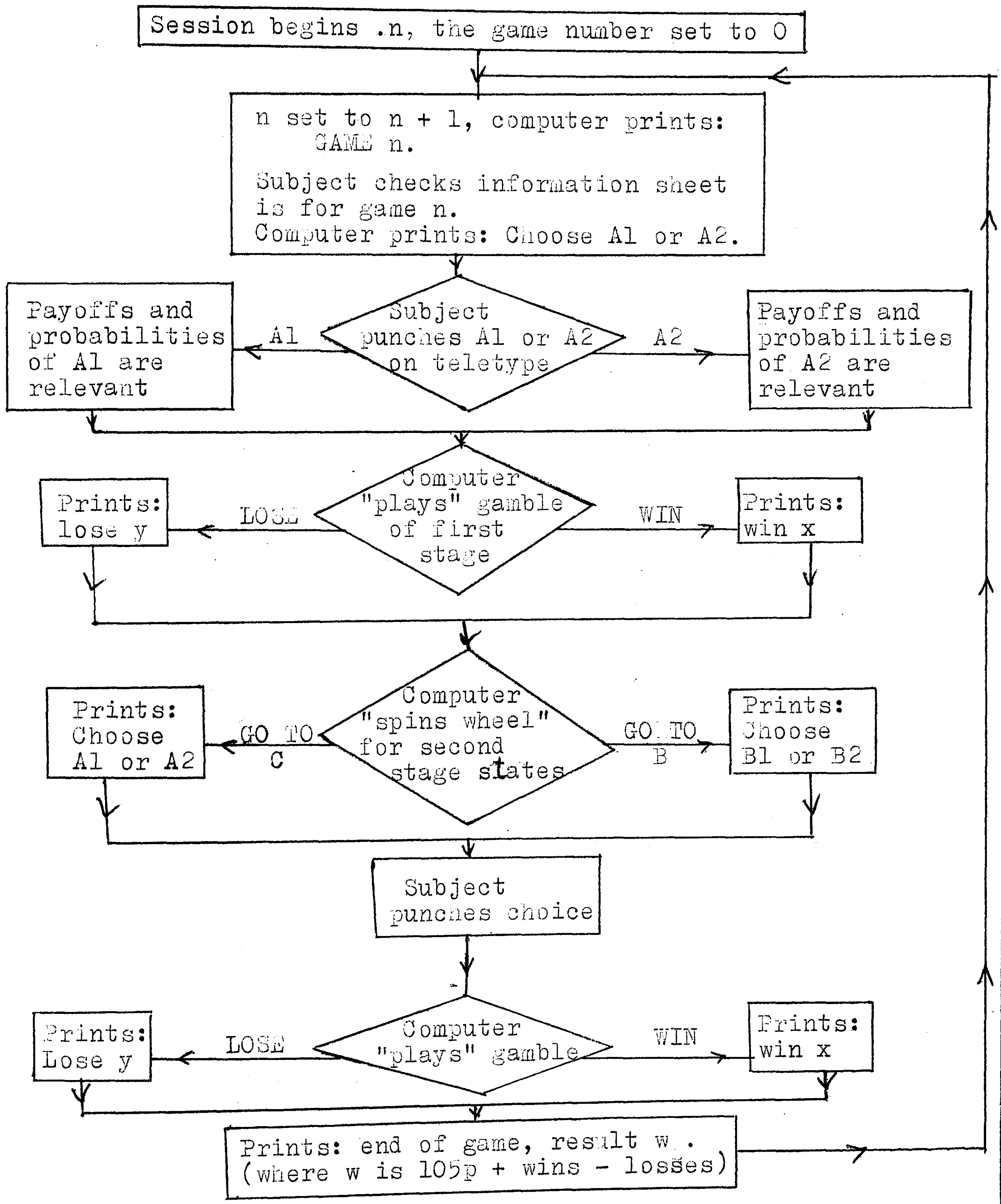


Figure 7.3.

Typical Teletype Printout Resulting from 3 Games.

GAME 1

STAKE 105

CHOOSE A1 OR A2

A1 WIN 40

CHOOSE B1 OR B2

B1

WIN 40

END OF GAME

RESULT 185

GAME 2

STAKE 105

CHOOSE A1 OR A2

A1 LOSE 15

CHOOSE C1 OR C2

C2

LOSE 10

END OF GAME

RESULT 80

GAME 3

STAKE 105

CHOOSE A1 OR A2

A1 WIN 40

CHOOSE C1 OR C2

C2

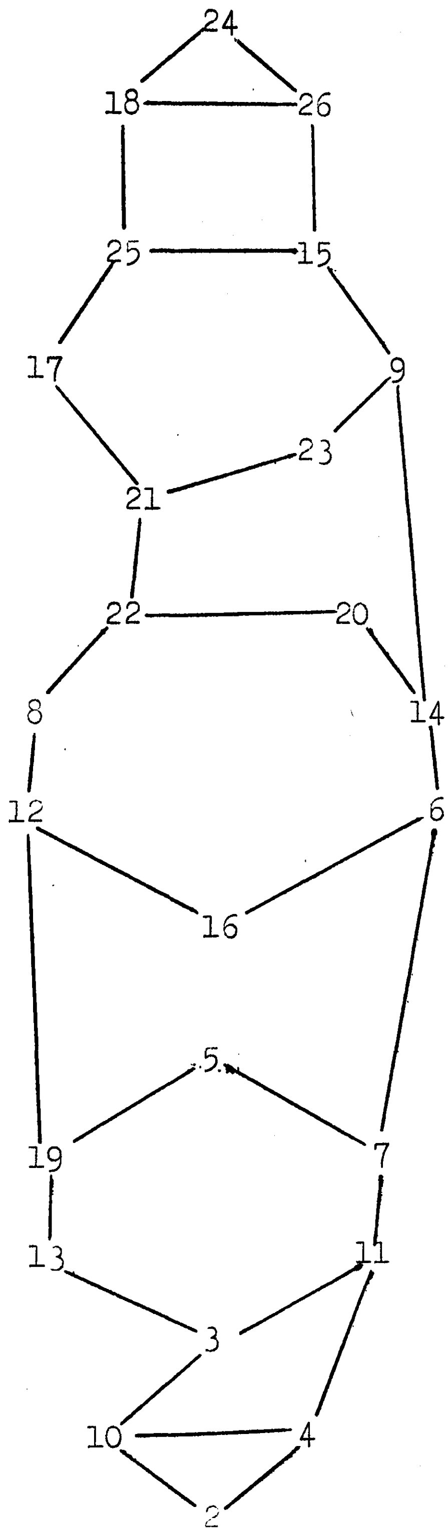
LOSE 10

END OF GAME

RESULT 135.

Figure 7.4.

The Cyclic-chain Design for Experiment 6.



Key

Gambles Code Number	Parameter Values P	W	L
26	.75	65	40
25	.75	65	65
24	.75	40	15
23	.75	40	40
22	.75	40	65
21	.75	15	15
20	.75	15	40
19	.50	15	65
18	.50	65	15
17	.50	65	40
16	.50	65	65
15	.50	40	15
14	.50	40	40
13	.50	40	65
12	.50	15	15
11	.50	15	40
10	.25	15	65
9	.25	65	15
8	.25	65	40
7	.25	65	65
6	.25	40	15
5	.25	40	40
4	.25	40	65
3	.25	15	15
2	.25	15	40

A line between two gambles indicates that this pair were presented.

Procedure

Subjects sat in front of the on-line teletype which was in a quiet room away from the actual computer. The experiment was introduced to them at the start of the first session with the following words, after the preliminary session information sheets had been indicated to them:

"This experiment involves playing a series of betting games, which have 2 stages.

At the start of each game you are given a stake of 105p with which to bet. In the first stage of the game you are in A, and you must choose between the bets A1 and A2 which are represented here. (state A is indicated). It is easiest to think of the circles as wheels of fortune, which have a pointer. To play a bet, the pointer is spun, and if it lands in the top sector you win, and if not you lose. So, the size of the top sector indicates your chance of winning and the size of the bottom sector your chance of losing. In these games, the bets in A give an equal chance of winning or losing. When you are in A, you must take the second wheel associated with the bet chosen. The pointer on this second wheel is spun, and where it lands decides where you go for the second stage. If it is the top sector you go to B, and if it is the bottom one you go to C.

At the second stage, you are either in B or C, and you simply choose between the bets available. The result of the game is your stake, plus or minus your total winnings or losses.

You play the game with the computer. It gives you the basic information for you to play the game. You type in your choices on the teletype and the computer types back how much you win or lose and which state you go to. It plays fairly, and when an event occurs it does so with the chance indicated to you."

The 5 demonstration games were then played by the subject. It was explained that the state B was more favorable than C and these always remained the same. They were told that the chances of going to B and C, the amount that could be won or lost all changed, but the chance of winning or losing in A did not change. As the subject played the demonstration games, any other points were cleared up. Also, his choices were observed to make sure he always chose the dominant one, where one choice in A was dominant. The instructions ended as follows.

"What you have just played is a miniature version of an experimental session. I want you to play 3 sessions, playing 132 games in each. Each should last about 1 hour. At the end of each session, we will pick out one of the games at random and record the result. You will be paid the average result of these three games. This is to give you an incentive to consider your choices carefully. Remember, each game begins with a stake of 105p."

The subject then began the experimental session and the experimenter left the room. At the end of the session he returned and the payoff game was selected by a coin tossing procedure. For the second and third sessions, subjects were merely reminded of the payoff conditions and the general structure of the game. The three sessions for each subject were played at the mutual convenience of the subject and experimenter. Most were carried out on successive days and none were more than three days apart. After the final session, subjects were paid and questioned about their strategies and any other points about the experiment. In particular, whether they thought the computer had played "fairly."

Results and Discussion.

It is more important to test for trends in responding in this experiment than it was in the previous ones because here the data was collected over three sessions. Intuitively it seems more likely that more changes will occur over a longer time span. As in experiments 3 and 5, therefore, each subject's data was arranged into a gamble pair x replications matrix. Jonkheere and Bower's heterogeneity of trend test was carried out on each 33 x 12 data matrix so formed. Jonkheere and Bowers' transformed W statistic is approximately unit normal under the bernoulli null hypothesis. Their values, denoted by Z are shown for each subject in table 7.1. It can be seen that the null hypothesis can be rejected at the 1% level for subjects 2 and 5 and the alternative hypothesis that changes in response probability occur is accepted for these subjects. For subject 3 also the z value is unlikely under the null hypothesis ($p = 0.0672$, two tailed).

Now, the conclusion from the above is that for 3 subjects any analysis based on an assumption of constant preference probabilities is dubious as this assumption is demonstrably not met. However, rather than waste this data it was decided to proceed with further analysis and consider the results to be undermined by the lack of stability of the response probabilities. This further analysis, as described in the introduction is by functional measurement to investigate two additive information integration models. The second column of table 7.1 shows the number of gamble pairs for each subject whose preference frequencies were not zero or one. Since the number of pairs in each case is 33 it is clear that zero and one preference frequencies did not dominate. Thus, parameter estimation by Bock and Jones' methods is not likely to be too biased and the goodness of fit statistics will probably approximate the asymptotic ones quite well.

The results of the Bock and Jones' functional measurement analysis are presented in the usual way. First the goodness of fit of the basic Thurstone model is shown in table 7.2. From the estimation chi-squares it is clear that the null hypothesis of all equal subjective values can be rejected at the .1% level for all subjects. The error chi-squares are all non-significant showing that the fit of model to data is good. The same cannot be said of the additive submodels that were considered. The goodness of fit of the general additive and linear additive models defined in the introduction is shown in table 7.3. All estimation chi-squares are significant, showing that both models account for more variance than expected by chance. Most error chi-squares are significant also, however showing the models do not fit too well. In particular, the error chi-squares associated with the linear additive model are significant for 4 subjects and those associated with the general additive model are significant for 2 subjects. The more successful model is the general additive one, and in figure 7.5 subjective values predicted by it are plotted against those obtained directly from the data for each subject. The graph for subject 3 indicates that there is a single point well away from the observed = predicted line. Possibly the bad fit of the model for this subject can largely be attributed to this.

The graphs of figure 7.6 are plots of estimated subjective values of alternatives against one or other of their parameter values, for subjects 5 and 3. For subject 5 it can be seen that all 9 graphs are roughly parallel. Anderson (1970) discusses graphical tests of goodness of fit for additive functional relationships. The condition that must be met for such relationships is parallelism. Subject 5's graphs therefore support the other results for the general additive model. Subject 3's graphs also support the previous results as parallelism clearly does not hold. The departure from parallelism may help to detect the source of failure of the additive model. This failure seems to be due to some alternatives with PA = .50 not being valued as much as expected in comparison with the other alternatives, especially when W = 65p or 15 p. The alternative with W = 65p, L = 65p, and PA = .50 is particularly lower than expected and is the notably deviant one in figure 7.5. This seems to be the source of departure from additivity - a W x PA interaction. The rough graphical method to detect the departure from additivity did not yield clear results for subject 1 so it is not reported. More refined methods were not warranted for either subject as no alternative information integration model to substitute for additive ones has emerged from theoretical discussions.

In conclusion, then it can be said that the general additive model predicted the decisions of 3 subjects quite well, but significant interactions between the alternative's parameters were found for the other 2 subjects. For one of these subjects the source of interactions was revealed quite clearly by graphical means but for the other it was not. Doubt about the positive results for the general additive model is warranted because the response probabilities for 2 of the 3 subjects for whom the model fitted were unstable. This attempt to extend the domain of simple information integration models has not, therefore been unanimously successful. The implications of this are discussed in the final chapter.

Table 7.1.

Jonkheere and Bower's Heterogeneity of Trend Test.

Subject	Z	N,*
1	-0.844	32
2	2.909**	15
3	1.838	18
4	1.517	31
5	3.136**	30

*N is the number of non-zero or one preference frequencies

** indicates significance at the 1% level.

Table 7.2.

The Goodness of Fit of the basic Thurstonian Model.

Subject	Chi-squares		
	Estimation	Error	Total
1	102.082***	10.062	112.144
2	167.362***	5.292	172.654
3	177.216***	6.977	184.193
4	95.613***	7.234	102.847
5	98.518***	6.492	105.010
Degrees of Freedom	24	9	33

***indicates significance at the .1% level.

Significances of total chi-squares are not shown, though all are significant at the .1% level.

Table 7.3.

Goodness of Fit of the General and Linear Additive Models.

Subject	Chi-squares			
	General Additive		Linear Additive	
	Estimation	Error	Estimation	Error
1	55.37***	46.71***	49.81***	52.26***
2	142.70***	24.66	141.40***	25.96
3	133.62***	43.60***	122.54***	54.66***
4	78.10***	17.51	43.67***	51.95***
5	75.15***	23.37	60.76***	37.76*
Degrees of Freedom	6	18	3	21

*** indicates significance at the .1% level

* indicates significance at the 5% level.

Figure 7.5.

Graphs of subjective values estimated directly from the data against those predicted by the general additive model.

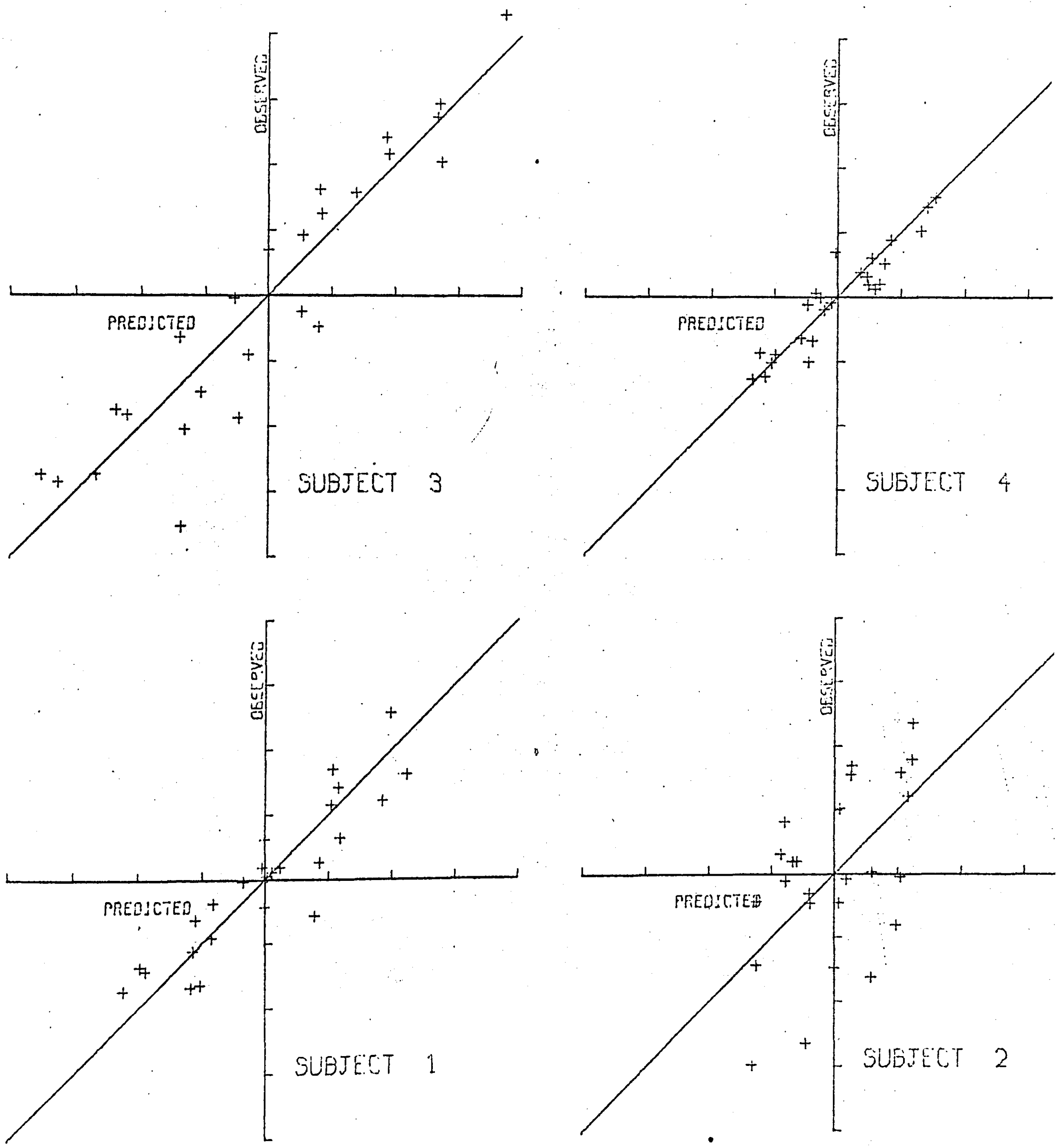


Figure 7.5 cont.

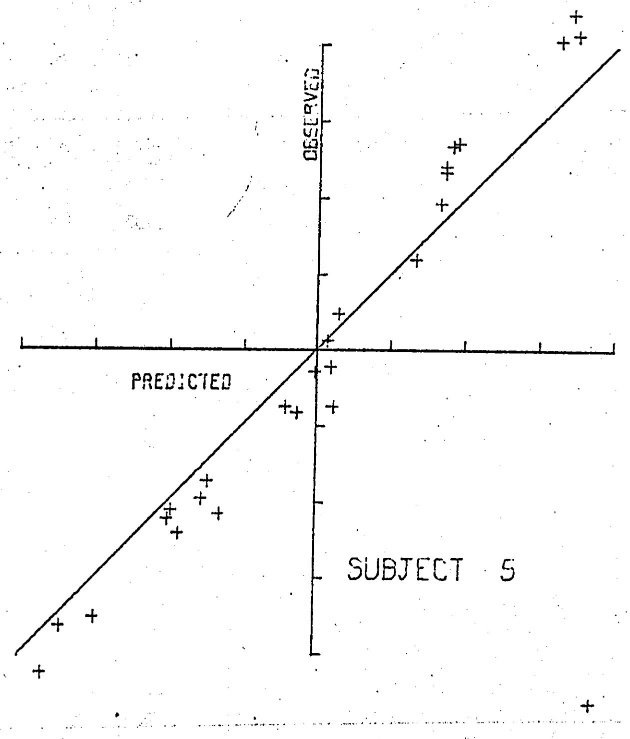
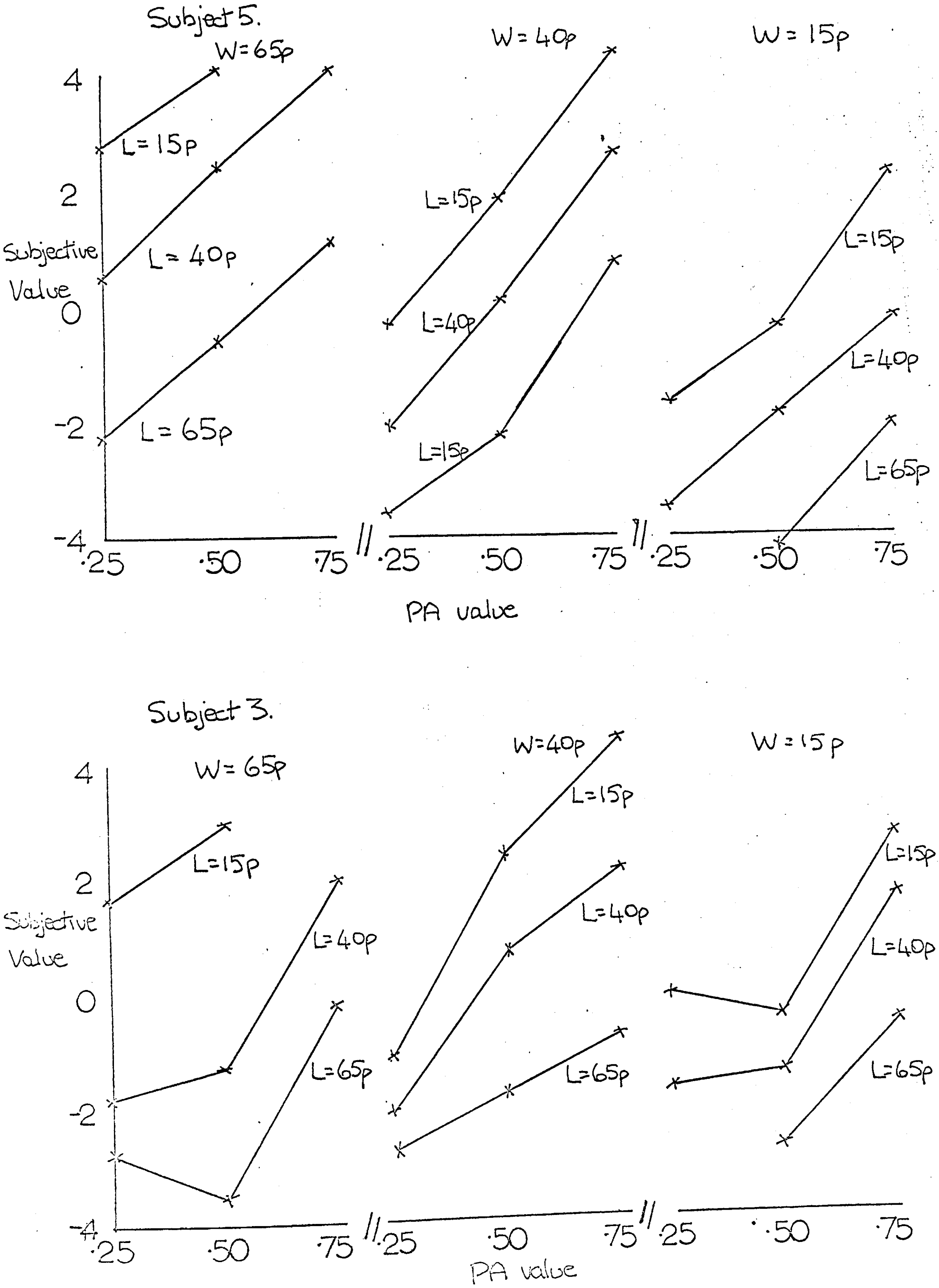


FIGURE 7.6

Subjective Values of the Alternatives Plotted Against Their Parameter Values, W L and PA.



Chapter 8.

Conclusions.

Qualitative Functional Analysis.

In this section the QFA results of chapter 5's experiments will be discussed in relation to the previous QFA results described in chapter 4. As well as this substantive discussion an attempt will be made to evaluate the methodology developed in chapter 2 in the light of its application in chapter 5. The substantive and methodological discussions will be kept separate. Some of the aims of the discussions will be to suggest how and where QFA might be best employed in the future study of individual decision making.

The main substantive issues raised in experiments 1 and 3 were whether preferences for simple gambles could be described by the multiplicative representation for $\mathcal{A}_2 = \langle P \times S, \leqslant \rangle$ and whether those for 3-parameter duplex gambles could be described by the dual-distributive representation for $\mathcal{A}_3 = \langle PW \times SW \times SL, \leqslant \rangle$. Other representations were rejected by virtue of their consequences violating the sure thing principle.

First, let us consider the choices made for simple gambles by the single subject in experiment 1. Any experiment using pair comparisons can only consider a relatively small number of gamble pairs. Thus, in the two previous studies where choices for simple gambles were observed the pairs used were chosen with the test of a particular condition in mind. Coombs et al (1967) hoped to test triple cancellation and

Tversky (1969) wanted to test transitivity of preference. The gambles used in experiment 1 were chosen because it was hoped to test the transitivity and double cancellation of indifference. However indifference did not appear to play a major role in the subjects' choices. In fact many of the observed preference frequencies were close to zero or one. If this subject's choices between simple gambles were going to be determined by indifference to any great extent it is reasonable to suppose it would have been apparent in this experiment, since the gambles differed little in payoffs and probabilities.

Thus, the broad conclusion that the data satisfied the SEU model is the only one that can be drawn, because no particular consequence was subjected to rigorous test. This agrees with support for the SEU model found by Coombs et al and with certain studies applying QFA to simple gambles which did not use choice data. Among these are Tversky (1967a)b) and Wallesten (1971). It is clear that choice data from more subjects should be collected to see if there are people for whom a fair number of choices are determined by indifference. Then the rigorous tests of indifference could be carried out. When the experiments were being planned however, problems existed which would have made this difficult. Such difficulties did not exist for statements of preference. Thus a further probe into the nature of indifference was carried out by collecting statements of preference in experiment 2.

This experiment put the indifference problem in perspective. It showed that indifference did appear to be structureless, but it did not appear to play an important part in the way people considered simple gambles. Again, general support for the SEU model was found. Experiment 2 vindicates to some extent earlier studies (e.g. Coombs et al (1967), Fagot (1959)) which explicitly excluded indifference from their models. Probably, strict preference models with errors defined simply will be adequate in future, though further choice data to check the minor role of indifference should be sought when methodological problems preventing this have been eliminated. In addition Tversky's experiment should be replicated in such a way that both probability and payoff dimensions are presented explicitly. The work still to be done with QFA of simple gambles, then is to give rigorous tests of each of the 4 main consequences of SEU using gamble pairs selected specifically with this in mind, and assuming an appropriate error model. Binary choices should be the main data but it should be supported with statements data from larger numbers of subjects.

Similar conclusions are to be drawn from experiment 3, in which the QFA of 3-parameter duplex gambles was carried out. In this experiment the gamble pairs were not selected with any particular consequence of SEU in mind. Rather, since there are no other QFA studies of decision making under uncertainty using 3 dimensional alternatives it was hoped to carry out a broad "goodness of fit" type test of representative pairs of duplex gambles. Unfortunately, the more representative the basic set of gambles the less rigorous is the testing of SEU's consequences. The test of SEU in experiment 3, then was not particularly powerful. Nevertheless, the SEU model was supported by these results. The next step, again after solving methodological problems should be to test each main consequence rigorously by observing choices from carefully selected gamble pairs.

Some of the unresolved methodological problems of QFA stem from the necessity to impose a statistical model on the algebraic structures which underly the representation hypotheses. For statements of preference data the statistical models (e.g. the simple error model) and algebraic structures fit together quite neatly. There is only one problem: how to maximize the likelihood of the data under the restriction of the qualitative conditions. This problem is all but solved. The appropriate algorithm is known but not programmed. The interim solution adopted in experiment 2 was actually quite reasonable and straight forward to carry out. Comparing observations (in this case percentage violations of conditions) with Monte Carlo results of "statistical subjects" behaving strictly in accordance with the model is an orthodox solution to model testing when analytic results are unknown.

The same satisfactory picture has not emerged with regard to binary choice data. There were 2 main problems: i) which probabilistic model to choose and ii) how to maximize the restricted likelihood of the data when some probabilistic model is assumed. If satisfactory solutions to these problems could be found likelihood ratio tests of the consequences of information integration models, parallel to those for statements of preference data, could be carried out. The solution proposed in chapter 2 was not particularly satisfactory. First of all consider the choice of probabilistic model. The choice was narrowed down to probabilistic models 1 and 2 on grounds that were discussed at length and still hold. But, when probabilistic model 1 is actually applied it is clear that the value of its parameter, α which fits the data best in general is $\alpha = 0.5$, which is not acceptable. This should lead us to reject model 1 and assume probabilistic model 2 in order to make the tests. Unfortunately this was not possible as it was not known how to solve problem ii) above for probabilistic model 2. Thus, it was necessary to assume model 1 and not accept $\alpha = 0.5$ which led to all kinds of difficulties, the most prominent being: a) the arbitariness of the selection of α and b) the conditionality of the likelihood ratio test when model 1 is assumed.

This position changes somewhat if the implications of the substantive findings about indifference, previously discussed are followed up. If indifference plays a minor role in human decision making, it need be considered in neither representation hypothesis nor probabilistic model. First of all, if indifference is not considered an α value of 0.5 in probabilistic model 1 does not appear to be so objectionable, which removes the problem a) above. However, more significantly all objections save one disappear for probabilistic model 2.

This comes about because a linear programming method of maximizing the restricted likelihood in the general case becomes possible. The maximum likelihood pattern of preferences restricted by the qualitative conditions considered and assuming probabilistic model 2 is simply that pattern with minimum reversals of the observed preferences which satisfies the conditions. The algorithm to do this is that given in appendix 2 with indifference omitted and the coefficients of the binary variables in the objective function changed as follows: for variable x_{ij} the coefficient becomes the number of times a_i was chosen over a_j. The notation above is different to that used in the appendix but the meaning should be clear. Now, the best restricted estimate of $1-\alpha$, the probability that observed and theoretical preferences coincide becomes the proportion of observed and theoretical preferences which coincide. The parallel of this case with that of statements data and the simple error model should be obvious. This satisfactory solution is an incentive to actually programming the maximization algorithm, which remains the only problem. It should be remembered that the above solution did not occur earlier as then there was no substantive reason to exclude indifference from the models. But if it can safely be adopted the future of QFA looks much more promising.

Functional Measurement.

The substantive findings from the functional measurement analysis of the static and sequential experiments (numbers 1, 3 and 5) will be compared to previous results regardless of what dependent variable they used. How much weight to give to present results in view of their reliability will be one of the main points discussed. The Monte Carlo experiments carried out are particularly important in this context.

The data from the single subject in experiment 1 allowed tests of different models to be carried out on 2 sets of data. Both tests showed considerable support for Slovids additive model. This contradicted the results of the major studies using simple gambles, i.e. Tversky (1967a)b)) and Anderson and Shanteau (1970). These studies, which involved bids or ratings for gambles supported a multiplicative model. In experiment 1 more than 50% of the response probabilities were zero or one. This would lead one to expect, in view of what was said in chapter 3 that there would be bias in the estimates of subjective value. This undermines the result of experiment 1 with respect to the functional measurement. Since also there was only one subject and only a narrow range of simple gambles was used caution should be exercised in drawing conclusions from the support for Slovic's model. From the point of view of functional measurement this experiment should be regarded as exploratory only. Before it could be shown that a real contradiction of the rating and bidding experiments occurred data from more subjects with a wider range of gambles is needed.

This leaves experiments 3 and 5 which both used the same pair comparison design involving duplex gambles (though in other aspects the experiments were very different). The previous studies directly relevant to these are Anderson and Shanteau (1970) which used ratings, Slovic and Lichenstein (1968a)) and Andriennsen (1971) which both used ratings and bids . The results from these studies are contradictory, some supporting Slovic's additive model and others the SEU model. The present experiments can shed some light on this because their functional measurement results are not likely to be suspect. Only a moderate proportion of zero or one preference frequencies were found. For most subjects in experiment 3 there were about 33% zero or one preference frequencies and less for subjects in experiment 5. Evidence from the Monte Carlo simulations suggests that the estimation and goodness of fit procedures will be relatively reliable in such cases.

In experiment 3 it was possible to reject the hypothesis that all subjective values were equal for 9 of the 10 subjects. In experiment 5 this was possible for only 1 of the 5 subjects. The data from these "non-random" behaving subjects was examined for goodness of fit with respect to Slovic's additive model and a linear SEU model. For 8 of the subjects the linear SEU model was significantly superior to Slovic's model even if the former was itself rejected. This contradicts Slovic and Lichenstein (1968a)) and Andriennsen (1971). By comparing Slovic's model to others we have shown up its failure. A criticism of these other studies is that they failed to make such a comparison, merely being content with stating that pilot studies "indicated" this was not necessary.

The above rejection of Slovic's model does not automatically imply acceptance of the SEU integration rule. A separate test of this was made. Two more general models were compared: the general additive model and the general SEU model. Both of these models found general support in the experiments.

But, a specific test of whether the general SEU model was significantly better than the general additive model was possible. This test revealed that for 7 of the "non-random" subjects the general SEU model should be accepted and for 2 the additive model should be accepted. For the remaining "non random" subject neither model could be accepted. These results confirm the general conclusion that was made in chapter 4 regarding previous research using functional measurement. That is, simple additive models account quite well for people's responses in many cases. But, where these have been compared to the SEU model this has generally been shown to be superior.

The examination by functional measurement of additive and SEU information integration rules in static and sequential decision situations has been successful. The Monte Carlo results, and the results of the actual experiments have vindicated the use of Bock and Jones methods on small amounts of choice data from individual subjects. As long as care is taken not to apply the techniques inappropriately there is no reason why the procedures should not be used extensively in the future.

The Sequential Factors Previous Outcomes and Current Capital.

In chapter 4 it was proposed that experiments should be carried out to examine sequential factors which were not susceptible to certain criticisms of earlier studies. In chapter 6 this was done. Properties of the experiments (4 and 5) in this chapter included: i) experimental controls to elliminate artefacts due to response bias and to "unconfound" the principle sequential factors and ii) rather better statistical tests of changes in responding than had previously been used. Also, the situation was realistic, in the sense that the sequential factors were not deliberately emphasized more than they would be in a natural setting.

Generally, the results supported the finding of Lichenstein (1965) and Edwards (1954a)b)d)). That is, neither sequential factor influences people's choices in a setting where subjects information corresponded to actual probabilities and payoffs. Experiment 4 resembled that of Siguel (1969), but did not find strong sequential dependencies. This suggests that his results were artefactual.

These negative findings of chapter 6 are in fact rather useful. They mean that static models will probably be adequate in sequential situations. Furthermore, it helps us to understand behaviour in dynamic situations where previous outcomes and current capital apparently determine what people choose. It suggests that it is not the change in fortune, or the run of success or failure itself which alters peoples choices. Rather, it is due to dynamic factors which change concommitantly with sequential ones. Most likely, the fact that people's opportunities for future decisions change when their capital does is the true determinant of their changing behaviour.

Finally, there is the finding that many people appear to choose at random in independent sequential situations. In the present, rather narrow discussion of information integration models this can be parsimoniously explained by saying there was no difference in subjective value in the pairs of alternatives presented to the subjects. Or it may be that with a small amount of data it was not possible to detect differences which were actually there. If apparently random behaviour is in fact common in decision making situations, however the mechanisms underlying it must be the subject of a wider explanatory theory of decision making. It does not seem likely that if a strategy embodied in some functional relationship model was actually followed by people then they would arrive at indifference so often.

Dynamic Factors in Decision Making.

The findings of chapter 7 are clearly very limited. They showed that people took account of the factor: probability of going to favorable/unfavorable future decision states in a simple, decision dependent, dynamic situation. Further to this they showed that for only some people is it likely that the above factor would be integrated additively with other first stage information dimensions. The future of the additive model looks bleak in the context of dynamic decision making situations. However, some information integration model must be able to account well for the data as the basic Thurstone scaling model fitted very well for all subjects.

In the discussion at the beginning of chapter 7 it was suggested that any information integration strategy employed by DMs in dynamic situations must necessarily be rather complex, mirroring the complexity of the alternatives that they face. The only models similar to the ones discussed in static situations which seem plausible are ones which propose that evaluations and information integration is broken up into stages. For instance, evaluations of some single information dimensions are made, followed by a partial integration of their subjective values so that impressions are formed of the subjective values of various groups of information dimensions. Further integration then takes place to produce subjective evaluations of larger groups of information dimensions and so on until an evaluation of the whole alternative is made. Within the context of information integration models considered in this study such a stage-like model seems necessary if only because any DM has only a limited information processing capacity.

The n-replicate pair comparison design is clearly not appropriate for a comprehensive examination of this whole process, because of the number of information dimensions involved. These increase exponentially with the number of time stages in the dynamic situation. More efficient experimental paradigms are requires, involving such data as ratings, bids, statements of preference etc. Of course, the n-replicate design has its place in confirming results from the more efficient data found to broadly support some model or other. This would have to be at one of the intermediate stages of the grand evaluative process described in the previous paragraph. Such studies to verify the conclusions about decision making from "non-decision making" studies of decision-making behaviour are obviously necessary.

The above is meant to defend experiment 6 against the charge that it is too narrow in range. As implied above, the next stage in studying information integration models in dynamic situations is to try to obtain some more comprehensive, though less direct evidence about the models. For instance, by such approaches as discussed by Goldberg (1968) in his review of empirical research on clinical and other complex judgements. The studies discussed by Goldberg differ from studies of dynamic decision making in the kind, though not the complexity of the alternatives which must be evaluated. They show that it is possible to carry out empirical research to find how people evaluate complex objects. An objection levelled against them has been that the information integration models proposed may only be paramorphic descriptions of the judgemental process. This charge becomes particularly plausible when the number of information dimensions of the alternatives judged by subjects becomes very large. However, it is a charge that would be levelled at the whole of the present study since no alternative classes of models were considered at any stage. For this reason it is discussed in the next section in the context of the whole study.

Information Integration Models: Homomorphic or Paramorphic?

It is appropriate in this last section to reconsider the philosophy underlying the experimental investigation which has been reported. Above all the view has been taken that verbal theories hide assumptions which one should be aware of and implications which should be tested. The answer has been to try to state any models used in mathematical terms so that their assumptions become explicit and their consequences derivable and testable. Theoretical formalism has been accompanied by the use of formal experimental methods. What were seen as important variables were controlled by either holding them constant or varying them independently and systematically to bring out their effect on the dependent variable, the choice. The virtues of formalization are well known and its validity in scientific endeavour will not be discussed. But, formal models are necessarily specific and can only capture a microcosm of reality. They need to be set in a broader, less formal conceptual framework and they are only valid to the extent that this superstructure is.

Let us consider the framework within which the formal models of the present study are set. The appropriateness of this framework determines the major strengths and weaknesses of the study. The class of information integration models of behaviour derive historically from an economic theory of man. They are the result of the gradual generalization by psychologists of classical Utility Theory (see Edwards 1954c, 1961). The generalization was found necessary as the years went by to account for the discrepancies between the way actual and ideal decision makers behaved. An ideal decision maker has been described many times. He is rational, fully informed and intent on maximizing expected utility. The success of classical Utility Theory in the context of economic decision making is due to the fact that a concensus can be reached about what is rational, what information is required before choosing and what the objective of the choice is.

What remains of Utility Theory in the information integration models of this study is the idea that people integrate the available information within each alternative and select the "best". The realism of even this aspect of utility models is questionable though. It is natural that economic man should gather his information, work out the expected payoff of each alternative and select that with the highest. The real decision maker however only has the limited resources of his own cognitive processes. He must use them to arrive at a decision within the constraints of the situation. For the individual in such a position processing the information within each alternative in turn may not be the most natural, or the most efficient means of deciding. Perhaps such considerations should have been to the fore from the earliest psychological studies of decision making. "Man the Thinker" would have been a better focus than "Man the Profit Maker."

This root of the information integration models has directed the aims of psychological studies which focus on them. The major aim of Utility Theory is to prescribe what decisions business men should take. The major aim of the information integration models is to describe what decisions people actually do take. In the transfer from the economic to the psychological investigation it is natural that the aims should transfer from prescribing to describing behaviour. However, a description of behaviour is not an explanation of it. The purpose of the early studies of information integration models was not to explain behaviour. To explain behaviour it is necessary to find out why people make the choices they do. This is done by describing the cognitive processes underlying peoples' actions not by describing what they do.

The present study of decision making has been discussed as applying the explanatory mode of investigation. That is, information integration models have been presented as as being homomorphic with a class of cognitive processes underlying the behaviour. This they are, and thus it is legitimate to consider them as explanatory models. Traditionally though they have been descriptive of behaviour. Whether they actually correspond to cognitive processes has not been an issue. They have been evaluated in terms of how well they describe behaviour. If they are only paramorphic models of the cognitive processes but make good predictions of behaviour then it has been felt that explanations are not important.

The explanatory mode of investigation has gained ground in the study of decision making since Slovic and Lichenstein's important paper (1968a). Recently a few quantitative models of behaviour based on alternative cognitive processes have been proposed. It is important to know under what conditions they would be superior to our models. Many clinical judgment studies have addressed themselves to this question (see Goldberg (1968)). A few have been concerned with more simple situations and three of these Einhorn (1970), (1971), Tversky (1969) will be discussed.

Einhorn suggested two models which might plausibly be employed by decision makers in some situations: the conjunctive model, where an alternative is only considered attractive if it is reasonably attractive on all its dimensions and the disjunctive model, where any alternative which is attractive on at least one dimension is considered attractive. Although Einhorn is not the first to have considered these models he appears to be the first to have attempted to evaluate them experimentally. Some evidence that for the same situation, different peoples' evaluations are approximated better by different models was found in an experiment where four judges rank ordered a set of job applicants. Three things were known about the applicants, who had to be ranked with respect to acceptability for graduate school.

Evidence in a later study, Einhorn (1971) also pointed to the conclusion that for some tasks and some people the conjunctive model fitted better than the best information integration model. This was more marked in a task where subjects rank ordered jobs with respect to attractiveness than in a task where subjects rank ordered a set of applicants for graduate school.

Tversky (1969), as already discussed, studied the lexicographic semi-order model and the class of difference models to which it belongs. Suppose DM is considering a pair of multi-altribute alternatives. The difference models suppose that he considers the difference between the values on each dimension and then integrates the subjective values of these differences in some way. A special case would be the additive difference model, where functions of the subjective differences are integrated additively to give an overall subjective difference with respect to subjective value, whose direction indicates the preferred alternative. Tversky showed that the special case where the functions of the subjective differences are linear is equivalent to the additive information integration model. Experimentally as discussed earlier he showed that transitivity does not hold in certain situations. This he regarded as supportive of the additive difference model, which does not require transitivity for multi-attribute alternatives. He suggests that information integration models may approximate actual processes better in situations where alternatives appear sequentially and difference models are better when alternatives appear simultaneously. Such ideas, although intuitively reasonable, obviously require backing up with experiments. The need to compare the information integration models with such alternative models as the above cannot be over-stresses, as what little evidence there is in relatively simple situations seems to indicate that they are very plausible alternatives.

From recent studies such as the above it looks likely that when information integration models are compared to other information processing models they will be shown to be worse in many cases. Viewed as explanatory models of decision making they are likely to be superceded. We have found, however that they describe behaviour reasonably well even if they do not explain it. They will make reasonable predictions at the gross level even if they fall down under close scrutiny as, for instance Tversky (1969) found. As descriptive models providing a paramorphism of peoples behaviour which enables good predictions to be made they are likely to have a long and useful life.

Appendix 1.

The A Priori Choice of α, the Parameter of Probabilistic Preference Model 1 for Binary Choice Data.

The basic experimental situation to which the probabilistic preference model is relevant is the n-replicate, pair comparison one, described in chapter 2. Let the set of alternatives be A with a, b etc. $\in$ A. If there are m elements in A then the subject's n x m(m-1)/2 choices are assumed to be independent, bernoulli trials. The binary preference probability for some pair (a,b) is denoted p(a,b). When this pair is presented to the subject suppose he chooses "a" k times. The set of k/n values for the pairs presented are the basic data from the experiment.

Under the bernoulli model, H_0 the maximum likelihood estimate (MLE) of p(a,b) is k/n. The likelihood function of this observation given H_0 is

$$L(k/n|H_0) = \frac{n!}{k!(n-k)!} \cdot p(a,b)^k \cdot (1-p(a,b))^{n-k}$$

Under probabilistic preference model 1, p(a,b) can only be in a certain range, determined by the model parameter, α. The range is different for each of the possible hypotheses $a < b$, $a \sim b$ and $b < a$. Suppose these hypotheses are denoted H_1, H_2 and H_3 respectively. The maximum likelihood of an observed preference frequency, k/n given one of these hypotheses can be denoted by $ML(k/n|H_i)$ where i = 1, 2, 3. These functions are special cases of the likelihood function when the MLE of p(a,b) under the appropriate hypothesis is substituted. The MLE's depend on α and were given in chapter 2.

The major consequence of choosing α is that it affects the values of $ML(k/n|H_i)$. Thus it directly affects the results of likelihood ratio tests of information integration models based on probabilistic preference model 1. The information relevant to the choice of α is the plot of $ML(k/n|H_i)$ against α for all values of k, n and i. Alpha should be chosen based on this information using certain criteria. The first criterion is that for any particular pair (a,b) each of the hypotheses, $H_i(i = 1,2,3)$ should be plausible whatever k/n is observed. This is so that it is difficult to reject an information integration model on the basis of observations from a single pair. Rejection should be due to observing incompatible pattern of choices which suggest violations of the QFA conditions. The second criterion is that an observed k/n should be useful in differentiating among the hypotheses, H_i. If this were not the case a statistical test of an information integration model would not be powerful. These criteria are not complementary since the first requires an α close to 1 and the second requires an α close to 0.5.

Now, values of α which are suitable may be different for different n. Even if n is fixed, however it is difficult to assess all the plots of $ML(k/n|H_i)$ as functions of α which are relevant to the choise of α . For this reason a criterion function was used the evaluation of which results in some compromise between the above criteria.

The criterion function was based on the functions

$$\frac{1}{\sum_{i=1,2,3} ML(k/n|H_i)}$$

which decreases as the likelihood of any of the hypotheses increases and

$$\frac{\min_{i=1,2,3}\left\{(ML(k/n|H_i)\right\}}{ML(k/n|H_2)}$$

which decreases as the difference between the likelihoods of the hypotheses increases.

For a particular observed preference frequency, k/n the value of α which minimizes the product of these functions should compromise reasonably between the two criteria of "good" α. Suppose the product of the functions is denoted by $C(k/n)$. The "best" α for a particular n is that which is best on average when all possible k are taken into account. Thus, it was decided to choose that α for which

$$\sum_{k=0,..n} C(k/n) \quad \text{is minimum.}$$

The values arrived at for different n are to a great extent arbitary. They are shown in figure A1.1 and referred to by the euphemism "optimal α". It is felt that whether the values of optimal α appear to give a sensible compromise between the two requirements is all that is wanted at present. To enable the suitability of the choice of α for $n = 7$, $\alpha = 0.80$ and $n = 15$, $\alpha = 0.73$ to be assessed plots of $\log(ML(k/n|H_i))$ as functions of k are shown in figures A1.2 and A1.3. Inspection of these figures indicates that the compromise aimed for between the two criteria are met.

FIGURE A1.1

Graph of optimal α against n.

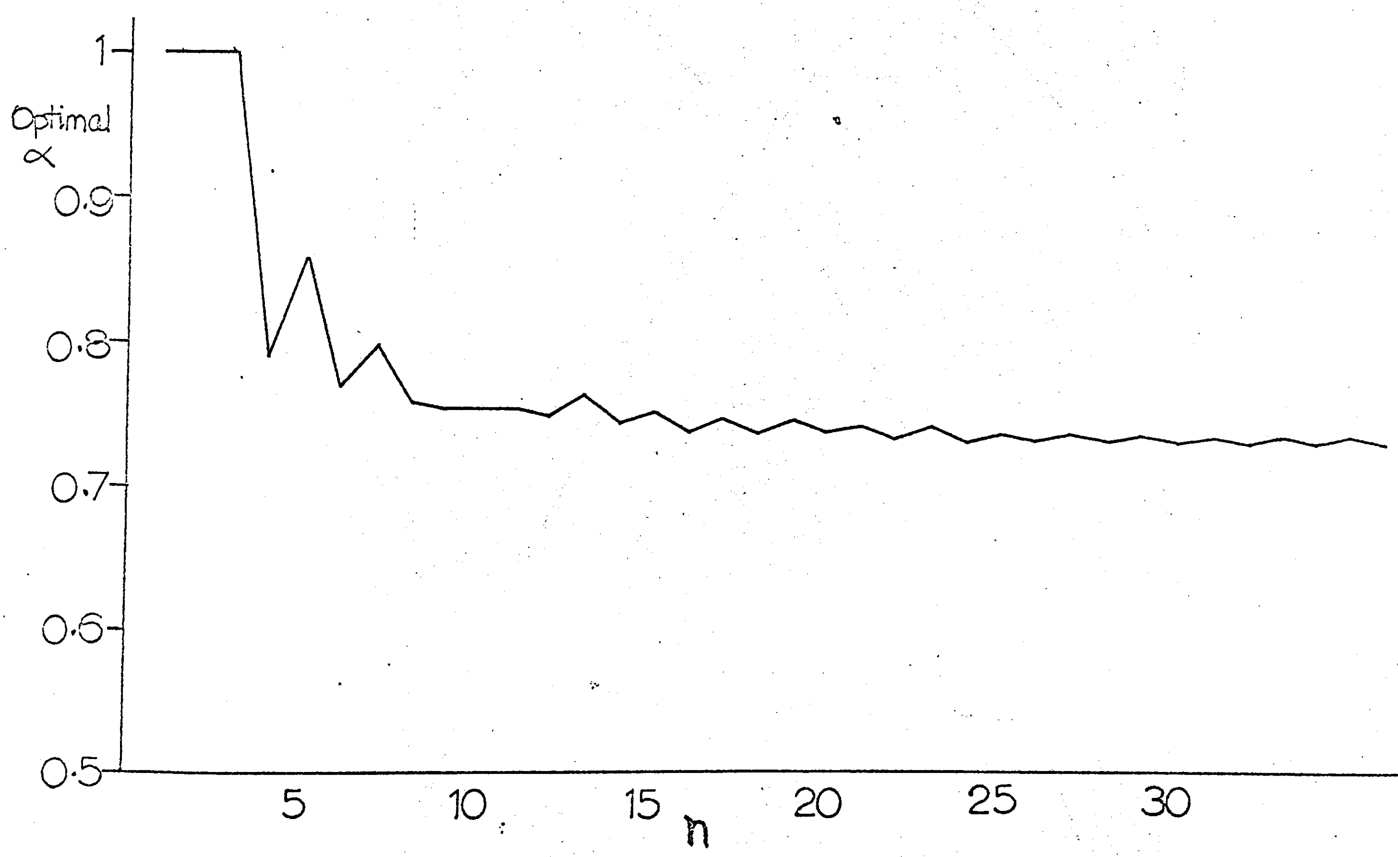


FIGURE A1.2

Graph of log ML(k/7/Hi) for i = 1, 2, 3, against k, Under Probabilistic Preference Model with $\alpha = .8$.

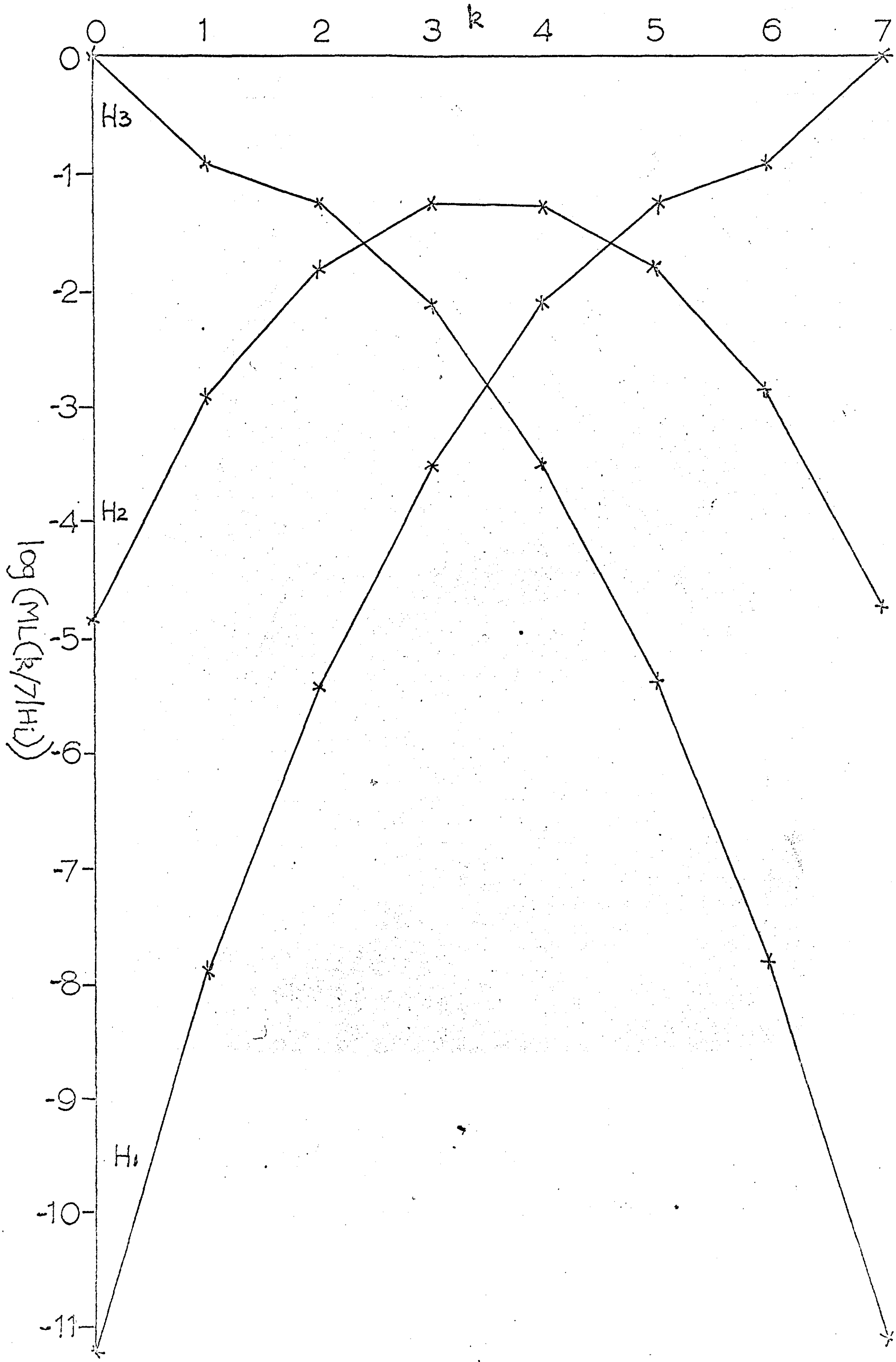
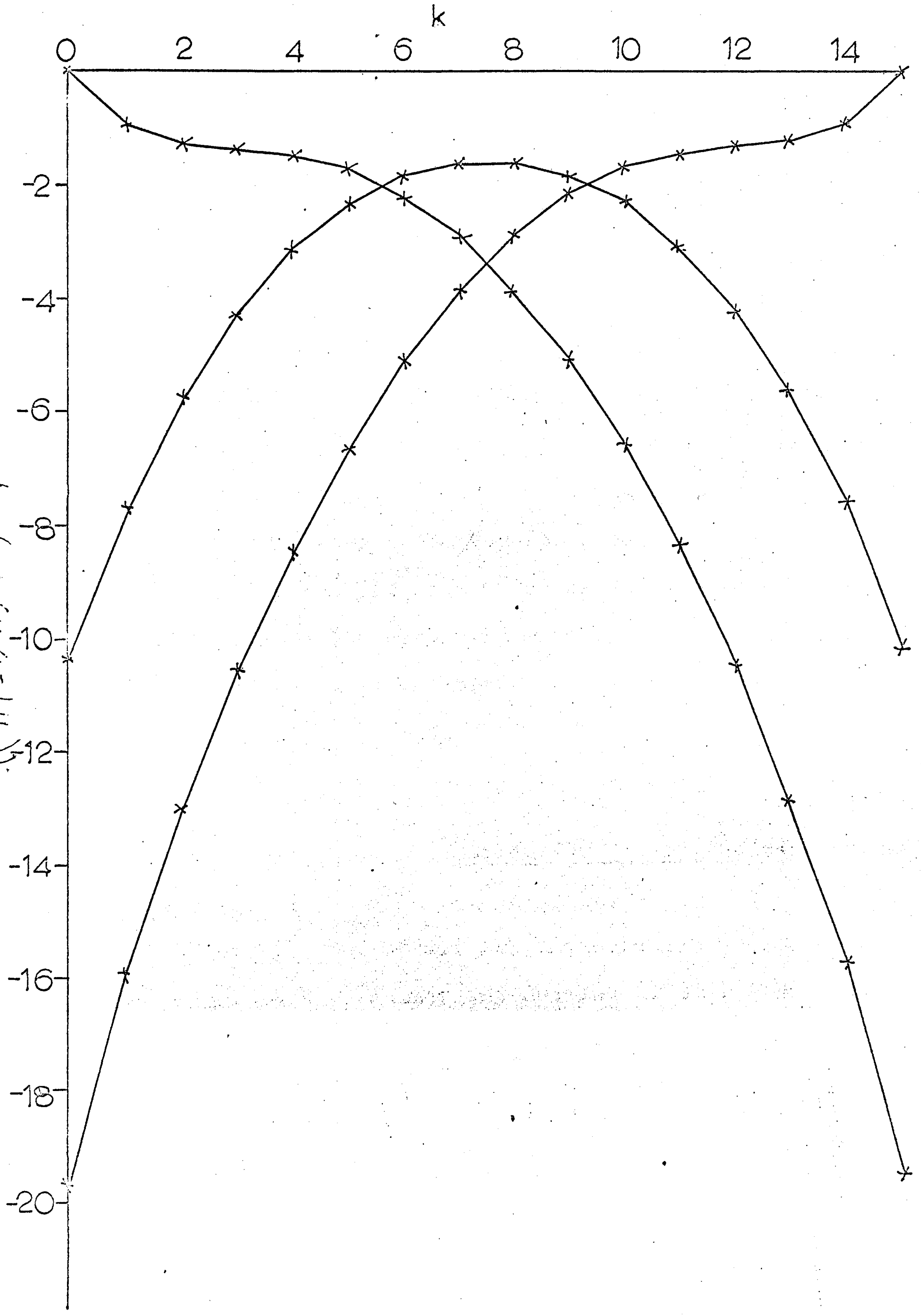


FIGURE A1.3

Graph of log (ML(k/15|Hi)) against k,
Under Probabilistic Preference Model 1 with $\alpha = \cdot 73$.



Appendix 2.

Maximizing the Likelihood of Data During the QFA of an Information Integration Model when Preference is Defined Probabilistically.

In chapter 2 likelihood ratio tests for the QFA of information integration models were proposed. It was indicated that the most tractable cases were for i) statements of preference data assuming the simple error model and ii) binary choice data assuming probabilistic model 1. All uses of the LR test require that the likelihood of the data is maximized under the hypothesis of interest. This is the difficult problem in our cases, and a solution to it is proposed based on a re-phrasing of the problem to a pseudo-boolean programming one. In this appendix two examples of an appropriate re-phrasing are given for case ii) above within classes of preference patterns which satisfy certain testable consequences of SEU. The alternatives are a) simple gambles and b) three parameter duplex gambles. In the final section a simple amendment to deal with both QFA tests for case i) above will be given.

The Likelihood Maximization Problem for Binary Choices Between Simple Gambles.

The solutions of all the maximization problems rely heavily on the work of Hammer and Rudeanu (1968). The first step, in all cases is to list the testable consequences of the model under examination which are of interest. When the model under examination is the SEU model, the testable consequences of interest are H5C2, H5C3 and H5C3'. Reasons for this are given in chapters 2 and 5. The problem then becomes one of maximizing the likelihood of a set of binary choices between simple gambles given that the conditions H5C2, H5C3 and H5C3' and probabilistic preference model 1 all hold. The way this problem is transformed into one to which Hammer and Rudeanu's techniques can be applied will now be described.

The properties H5C2, H5C3 and H5C3' must be converted into equations and inequalities in binary variables. Suppose the set of gambles are denoted P x S with $(p_i, s_j) \in$ P x S, $i = 1, \ldots p, j = 1, \ldots s$. Define the variables

$$x_{ijkl} = \begin{cases} 1 & \text{if } (p_k, s_l) < (p_i, s_j) \\ 0 & \text{otherwise} \end{cases}$$

and $$z_{ijkl} = \begin{cases} 1 & \text{if } (p_i, s_j) \sim (p_k, s_l) \\ 0 & \text{otherwise} \end{cases}$$

It is stated without proof (which is just a matter of careful checking) that the properties above are equivalent to the system of inequalities/equalities below.

H5C2.

i) $x_{ijkl} + x_{klij} + z_{ijkl} = 1$ (connectedness)

and $z_{ijkl} - z_{klij} = 0$ (symmetry of indifference)

both for all $ps(ps - 1)/2$ unordered pairs,

$(p_i, s_j), (p_k, s_l) \in$ P x S

ii) $x_{ijkl} + x_{klmp} - x_{ijmp} \leq 1$ (transitivity of strict preference)

for all ordered triples $(p_i, s_j), (p_k, s_l), (p_m, s_p) \in$ P x S

$$\left.\begin{aligned} \text{iii)}\quad & z_{ijkl} + z_{klmp} - z_{ijmp} \leqslant 1 \\ & z_{ijmp} + z_{mpkl} - z_{ijkl} \leqslant 1 \\ & z_{klij} + z_{ijmp} - z_{klmp} \leqslant 1 \end{aligned}\right\}\text{(transitivity of indifference)}$$

for all unordered triples $(p_i, s_j), (p_k, s_l)(p_m s_p) \in P \times S$

H5C3'

iv) $x_{ijkl} + x_{kpmj} - x_{ipml} \leqslant 1$

(cancellation of strict preference within win or lose gamble sets)

for all i, j, k, l, m, p within a win or lost set such that not [(i = k and k = m) or (j = l and l = p) or (i = k and j = l) or (k = m and p = j) or (i = m and p = l)]

H5C3

v) $x_{ijkl} + z_{kpmj} - z_{ipml} \leqslant 1$ (cancellation of indifference)

for all $1 \leqslant i \leqslant k \leqslant m \leqslant p$ and $1 \leqslant j \leqslant l \leqslant p \leqslant s$

such that not [(i = k and k = m) or (j = l and l = p) or (i = k and j = l) or (k = m and p = j) or (i = m and p = l)]

Any solution of this system corresponds to a pattern of preference - indifference which satisfies the conditions H5C2, H5C3 and H5C3'. The log-likelihood function of a set of data given a preference pattern and assuming probabilistic model I can be put in terms of what Hammer & Rudeanu call a pseudo-boolean function of the x_{ijkl} and z_{ijkl} variables. Let the log likelihood of the data given the model be denoted by z. This function z is to be maximized subject to the restrictions above. It is the following function of the x_{ijkl}'s and z_{ijkl}'s:

$$z = \sum_{\substack{\text{all unordered} \\ \text{pairs } (p_i, s_j), \\ (p_k, s_l) \in P \times S}} {}_pC_{ijkl}\, x_{ijkl} + {}_I C_{ijkl}\, z_{ijkl} + {}_{\bar{p}} C_{ijkl}\, x_{klij}$$

where the coefficients are

$${}_p C_{ijkl} = \text{the log likelihood when } (p_k, s_l) < (p_i, s_j)$$

$${}_I C_{ijkl} = \text{the log likelihood when } (p_i, s_j) \sim (p_k, s_l)$$

$${}_{\bar{p}} C_{ijkl} = \text{the log likelihood when } (p_i, s_j) < (p_k, s_l)$$

which are constants, functions of the data and the parameter of probabilistic model I, α. The problem to find the maximum of the log likelihood (and therefore the likelihood) subject to the conditions being considered is thus transformed to what Hammer & Rudeanu call a linear pseudo-boolean programming problem. Since they give an algorithm which will solve these types of problems (see reference) it is in principle possible to find the maximum. The same method could obviously be used to test the SEU in semi-orders model and also to test both models in the three parameter gamble case. Also, the algorithms are suitable for data obtained from partial pair comparison experiments. This is important as it is quite impossible to obtain complete data for even moderate m and n. In the incomplete case, the coefficients of the variables for pairs not observed are simply set to zero in the linear function to be maximized. Let us now consider the problem of maximization under SEU and preference model I for three parameter gambles.

The Likelihood Maximization Problem for Binary Choices between Three Parameter Duplex Gambles.

The important conditions for the QFA of SEU in this case are, as discussed in chapters 2 and 5 the conditions H9C2, H9C3 and H9C3'. As an illustration, the reformulation of the maximization problem subject to the constraints H9C2, H9C3' and H9C3 will be set out, again assuming probabilistic model 1.

When the "chance of losing" parameter is held constant let the set of three parameter duplex gambles be denoted by P x W x L with elements (p_i, w_j, l_k) where i = 1, ...p, j = 1,...w and k = 1, ...l. Denote the variables by

$$x_{ijklmp} = \begin{cases} 1 \text{ if } (p_l, w_m, l_p) \prec (p_i, w_j, l_k) \\ 0 \text{ otherwise} \end{cases}$$

$$z_{ijklmp} = \begin{cases} 1 \text{ if } (p_i, w_j, l_k) \sim (p_l, w_m, l_p) \\ 0 \text{ otherwise} \end{cases}$$

The log likelihood of the data is expressed in terms of these binary variables as:

$$Z = \sum_{\substack{\text{all unordered} \\ \text{pairs } (p_i, w_j, l_k), \\ (p_l, w_m, l_p)}} p\, C_{ijklmp}\, X_{ijklmp} + I\, C_{ijklmp}\, Z_{ijklmp} + \bar{p}\, C_{ijklmp}\, Z_{lmpijk}$$

where the coefficients are analagous to those of the previous case.

The system of constraints corresponding to the consequences of SEU are as follows:

H9C2:

i) $X_{ijklmp} + Z_{ijklmp} + X_{lmpijk} = 1$ (connectedness)

ii) $Z_{ijklmp} - Z_{lmpijk} = 0$ (symmetry of indifference)

for all unordered pairs (p_i, w_j, l_k), (p_l, w_m, l_p)

iii) $X_{ijklmp} + X_{lmnpqv} - X_{ijkpqv} \leqslant 1$

(transitivity of strict preference)

for all unordered triples (p_i, w_j, l_k), (p_l, w_m, l_n), (p_p, w_q, l_v)

iv) $Z_{ijklmn} + Z_{lmnpqr} - Z_{ijkpqr} \quad 1$

$Z_{ijkpqr} + Z_{pqrlmn} - Z_{ijklmn} \quad 1$

$Z_{lmnijk} + Z_{ijkpqr} - Z_{lmnpqr} \quad 1$

Transitivity of indifference.

for all unordered triples

(p_i, w_j, l_k), (p_l, w_m, l_n), (p_p, w_q, l_r)

H9C3.

v) $X_{ijklmp} + X_{lmsqrk} - X_{ijsqrp} \leqslant 1$

(dual-distributive cancellation of strict preference)

for all i, j, k, l, m, p, q, r, s,

such that not

[(i = l and l = q and j = m and m = r) or (k = p and p = s)
or (i = l and j = m and k = p) or (l = q and m = r and s = k)
or (i = q and j = r and s = p)]

vi) $Z_{ijklmp} + Z_{lmsqrk} - Z_{ijsqrp} \leqslant 1$

(dual-distributive cancellation of indifference for all $1 \leqslant i \leqslant l \leqslant q \leqslant p$

$1 \leqslant j \leqslant m \leqslant r \leqslant w$

$1 \leqslant k \leqslant n \leqslant s \leqslant l$ such that the same condition as in v) holds.

H9C3'

vii) $Z_{ijpklp} + Z_{kqpmjp} - Z_{iqpmlp} \leqslant 1$

(double cancellation of indifference)

for all $1 \leqslant i \leqslant k \leqslant m \leqslant p$

$1 \leqslant j \leqslant l \leqslant q \leqslant w$ and all p

such that not

{(i = k and k = m) or (j = l and l = q) or (k = m and q = j) or (i = k and j = l) or (i = m and q = l)}

within signed classes.

viii) $X_{ijpklp} + X_{kqpmjp} - X_{iqpmlp} \leqslant 1$

(double cancellation of strict preference)

for all i, k, m, j, l, q and all p such that the condition of vii) holds.

These solutions seem rather laborious but the object has been to show that in principle at least the problems can be solved. Only reasonable computation facilities need to be available to carry out the procedure. Now, let us turn to the maximization problem when statements of preference are the data.

The Maximization Problem for Statements of Preference, Assuming the Simple Error Model.

In chapter 2 it was indicated that the likelihood of the data given a class of preference patterns would be at a maximum for any pattern which differed from the observed pattern by minimum k preferences. This suggests a simple amendment to the algorithms proposed in the last two sections. In the case of simple gambles the constants of the objective function are changed as follows:

$$p\,C_{ijkl} = \begin{cases} 1 \text{ when } (p_k, s_l) < (p_i, s_j) \\ 0 \text{ otherwise} \end{cases}$$

$$I\,C_{ijkl} = \begin{cases} 1 \text{ when } (p_k, s_l) \sim (p_i, s_j) \\ 0 \text{ otherwise} \end{cases}$$

$$\bar{p}\,C_{ijkl} = \begin{cases} 1 \text{ when } (p_i, s_j) < (p_k, s_l) \\ 0 \text{ otherwise} \end{cases}$$

This sets the objective function, z equal to the number of latent preferences the same as the observed ones. The similar substitution for the case of three parameter duplex gambles can be made, and when the pseudo-boolean programming problems are solved and z is maximized, the minimum k pattern of preferences is found.

Bibliography

ADAMS, E.W. (1965) Elements of a theory of inexact measurement. Phil. of Sci., 32, 205-229.

ADAMS, E.W. & FAGOT, R.F. (1959) A model of riskless choice. Behavl. Sci., 4, 1-10.

ADAMS, E.W., FAGOT, R.F. & ROBINSON, R.E. (1965) On the empirical status of axioms in theories of fundamental measurement. Oregon U.P.

ADAMS, E.W., FAGOT, R.F. & ROBINSON, R.E. (1970) On the empirical status of axioms in theories of fundamental measurement. J. Math. Psychol., 7, 379-409.

ANDERSON, N.H. (1970) Functional measurement and psychophysical judgement. Psychol. Rev., 77, 153-170.

ANDERSON, N.H. & ALEXANDER, G.R. (1971) Choice test of the averaging hypothesis for information integration. Cog.Psychol., 2, 313-324.

ANDERSON, N.H. & SHANTEAU, J.C. (1970) Information integration in risky decision making. J. exp. Psychol., 84, 441-451.

ANDERSON, T.W. & GOODMAN, L. (1957) Statistical inferences about Markov chains. Ann. math. Statis., 28, 89-110.

ANDRIESSEN, J. (1971) Comments on a new risk-taking model. Acta psychol., 35, 173-187.

BECKER, G.M., DEGROOT, M.H., & MARSCHAK, J. (1963a) Stochastic models of choice behaviour. Behavl. Sci., 8, 41-45.

BECKER, G.M., DEGROOT, M.H. & MARSCHAK, J. (1963b) An experimental study of Stochastic models for wagers. Behavl. Sci., 8, 199-202.

BECKER, G.M., DEGROOT, M.H. & MARSCHAK, J. (1963c) Probabilities of choice among very similar objectives. Behavl. Sci., 8, 306-311.

BECKER, G.M., DEGROOT, M.H. & MARSCHAK, J. (1964) Measuring utility by a single-response sequential method. Behavl. Sci., 9, 226-232.

BECKER, G.M. & MCCLINTOCK, G.G. (1967) Value : behavioural decision theory. Ann. Rev. Psychol., 18, 239-286.

BERKSON, J. (1955) Maximum likelihood and minimum chi-square estimates of the logistic function. J. Am. statis. Ass., 50, 130-162.

BOCK, R.D. & JONES, L.V. (1968) The measurement and prediction of judgement and choice. Holden-Day.

BOX, G. & MULLER, M. (1959) A note on the generation of normal deviates. Ann. math. Statist., 28, 610-611.

BRADLEY, R.A. & TERRY, M.E. (1962) Rank analysis of incomplete block designs. I. The method of paired comparisons. Biometrika, 39, 324-345.

BURKE, C.J. & ZINNES, J.L. (1965) A paired comparison of pair comparisons. J. math. Psychol., 2, 53-76.

COHEN, J., BOYLE, L.E., & SHUBSACHS, A. (1969) Rouge et noir : influence of previous play on choice of binary event outcome and size of stake in a gambling situation. Acta psychol., 31, 340-352.

COOMBS, C.H., BEZEMBINDER, T.G., GOODE, F.M. (1967) Testing expectation theories in decision making without measuring utility or subjective probability. J. math. Psychol., 4, 72-103.

COOMBS, C.H. & HUANG, L.C. (1970) Polynominal psychophysics of risk. J. math. Psychol., 7, 317-338.

COOMBS, C.H. & KOMORITA, S.S. (1958) Measuring utility of money through decisions. Am. J. Psychol., 71, 383-389.

COOMBS, C.H. & PRUITT, D.G. (1960) Components of risk in decision making : probability and variance preferences. J. exp. Psychol., 60, 265-277.

COX, D.R. (1958) The regression analysis of binary sequences. Jl. R. statis. Soc. series B., 20, 215-242.

COX, D.R. (1970) Analysis of binary data. Methuen.

DALE, H.C.A. (1962) On choosing between bets. Q.J. exp. Psychol., 14, 49-51.

DAVID, H.A. (1963) The method of paired comparisons. Griffin.

DAVIDSON, D., MCKINSEY, J. & SUPPES, P. (1955) An outline of a formal theory of value. Phil. of Sci., 22, 140-160.

DAVIDSON, D., SUPPES, P. & SEIGEL, S. (1957) Decision making : an experimental approach. Stanford U.P.

DAVIS, D.M. (1958) The transitivity of preferences. Behavl. Sci., 3, 26-33.

DE CANI, J.S. (1969) Maximum likelihood paired comparison ranking by linear programming. Biometrika, 56, 537-545.

DURAP, H. (1957) Recherche de plans d'experience temporels à transitions exhaustives. Bull. du C.E.R.P., 16, 21-39.

EDWARDS, W. (1953) Probability-preferences in gambling. Am. J. Psychol., 66, 349-364.

EDWARDS, W. (1954a) The reliability of probability preferences. Am.J.Psychol., 67, 68-95.

EDWARDS, W. (1954b) Probability preferences among bets with differing expected values. Am. J. Psychol., 67, 56-57.

EDWARDS, W. (1954c) The theory of decision making. Psychol. Bull., 51, 370-417.

EDWARDS, W. (1954d) Variance preferences in gambling. Am. J. Psychol., 67, 441-452.

EDWARDS, W. (1961) Behavioural decision theory. Ann. Rev. Psychol. 12, 473-498.

EDWARDS, W. (1962a) Dynamic decision theory and probabilistic information processing. Human Factors, 4, 59-73.

EDWARDS, W. (1962b) Subjective probabilities inferred from decisions. Psychol. Rev., 69, 109-135.

EDWARDS, W. (1969) Decision processes research : an intemperate review. Proc. XIX International Psychology Congress.

EDWARDS, W., LICHTENSTEIN, S. & SLOVIC, S.P. (1965) Boredom-induced changes in preferences among bets. Am. J. Psychol., 78, 208-217.

EINHORN, H.J. (1970) The use of nonlinear, non compensatory models in decision making. Psychol. Bull., 73, 221-223.

EINHORN, H.J. (1971) Use of nonlinear, non compensatory models as a function of task and amount of information. Org. Behavl. Hum. Perf., 6, 1-27.

FAGOT, R.F. (1956) A theory of decision making. Unpublished Phd thesis. Sanford University.

FAGOT, R.F. (1959) A model for ordered metric scaling by comparison of intervals. Psychometrika, 24, 157-162.

FISHBURN, P.C. (1964) Decision and value theory. Wiley.

FISHBURN, P.C. (1969) Intransitive indifference with unequal indifference levels. J. math. Psychol., 6, 144-149.

GIOIA, A.A. (1967) A note on Roskies axioms for multiplicative representation. J. math. Psychol., 4, 151-153.

GOLDBERG, L.R. (1968) Simple models or simple processes? Am. Psychol. 23, 483-496.

GREENBERG, M. & WEINER, B. (1966) Effects of reinforcement history upon risk taking behaviour. J. exp. Psychol., 71, 587-592.

HACHAUER, B. (1970) Decision making in roulette. Acta Psychol., 34, 357-366.

HAMMER, P.L. & RUDEANU, S. (1968) Boolean methods in operations research. Springer-Verlag.

HURST, P.M. & SEIGEL, S. (1956) Predictions of decisions from a higher ordered metric scale of utility. J. exp. Psychol., 52, 138-144.

IRWIN, F.W. (1958) An analysis of the concepts of discrimination and preference. Am. J. Psychol., 71, 152-163.

KRANTZ, D. (1967) A survey of measurement theory. Lectures in Applied Mathematics, 12, 314-350.

KRANTZ, D. & TVERSKY, A. (1971) Conjoint measurement analysis of composition rules in psychology. Psychol. Rev., 78, 151-169.

KRANTZ, D., LUCE, R., SUPPES, P. AND TVERSKY, A. (1971) Foundations of Measurement, volume 1. Wiley.

LICHTENSTEIN, S.C. (1965) Bases for preferences among three-outcome bets. J. exp. Psychol., 59, 162-169.

LICHTENSTEIN, S.C. & SLOVIC, P. (1971) Reversals of Preference between bids and choices in gambling decisions. J. exp. Psychol., 89, 46-55.

LICHTENSTEIN, S.C., SLOVIC, P. & ZINK, D. (1969) Effect of instruction in expected value on optimality of gambling decisions. J. exp. Psychol., 79, no. 2, pt 1.

LINDMAN, H.R. (1971) Inconsistent preferences among gambles. J. exp. Psychol., 89, 390-397.

LUCE, R.D. (1956) Semiorders and a theory of utility discrimination. Econometrica, 24, 178-191.

LUCE, R.D., (1959) Individual choice behaviour. Wiley.

LUCE, R.D. (1966) Two extensions of conjoint measurement. J. math. Psychol., 3, 348-370.

LUCE, R.D. & SUPPES, P. (1965) Preference, utility and subjective probability. In R.D.Luce et al., (Eds.) Handbook of mathematical psychology, volume II. Wiley.

LUCE, R.D. & TUKEY, J.W. (1964) Simultaneous conjoint measurement : a new type of fundamental measurement. J. math. Psychol., 1, 1-27.

MAY, K.O. (1954) Transitivity, utility and aggregation in preference patterns. Econometrika, 22, 1-13.

MOOD, A. & GRAYBILL, F. (1960) Introduction to the theory of statistics. McGraw-Hill.

MORRISON, H.W. (1963) Testable conditions for triads of paired comparison choices. Psychometrika, 28, 369-390.

PAYNE, J.W. & BRAUNSTEIN, M.L. (1971) Preferences among gambles with equal underlying distributions. J. exp. Psychol., 87, 13-18.

PEARSON, E.S. & HARTLEY, H.O. (1958) Biometrika tables for statistitians, volume 1. C.U.P.

PFANZAGL, J. in co-operation with BAUMANN, V. & HUBER, H. (1968) Theory of measurement. Wiley.

RANYARD, R.H. (1971) Conjoint measurement hypotheses for three dimensions. Meeting of the Mathematical and Statistical section of the B.P.S. University of Birmingham.

RAPOPORT, A. (1966a) A study of human control in a stochastic multistage decision task. Behavl. Sci, 11, 18-32.

RAPOPORT, A. (1966b) A study of a multistage decision making task with an unknown duration. Human Factors, 8, 54-61.

RAPOPORT, A. (1967a) Dynamic programming models for multistage decision tasks. J. math. Psychol., 4, 48-71.

RAPOPORT, A. (1967b) Variables affecting decisions in a multistage inventory task. Behavl. Sci., 12, 194-204.

RAPOPORT, A. (1968) Choice behaviour in a Markovian decision task. J. math. Psychol., 5, 163-181.

RAPOPORT, A. & JONES, L.V. (1970) Gambling behaviour in two-outcome, multistage betting games. J. math. Psychol., 7, 163-187.

RAPOPORT, A., JONES, L.V., & KAHAN, J.P. (1970) Gambling behaviour in multiple choice betting games. J. math. Psychol., 7, 12-36.

ROSKIES, R. (1965) A measurement axiomatization for an essentially multiplicative representation of two factors. J. math. Psychol., 2, 266-276.

SANDERS, D.L. (1961) The intensity of newspaper content preferences. Unpublished doctoral dissertation. Stanford University.

SCOTT, D. (1964) Measurement structures and linear inequalities. J. math. Psychol., 1, 233-247.

SCOTT, D. & SUPPES, P. (1958) Foundational aspects of theories of measurement. J. Symbol. Log., 23, 113-128.

SEIGEL, S. (1956) A method of obtaining an ordered metric scale. Psychometrika, 21, 207-216.

SHANTEAU, J. & ANDERSON, N.H. (1969) Tests of a conflict model for preference judgement. J. math. Psychol., 6, 312-325.

SIGUEL, E.N. (1969) Sequential decision making under risk. Unpublished Phd. thesis. University of Michigan.

SJOBERG, L. (1966) Rotation to a hypothesis of constant factor loadings. Br. J. math. statist. Psychol., 19, 87-95.

SJOBERG, L. (1968) Studies of the rated favorableness of offers to gamble. Scand. J. Psychol., 9, 257-273.

SLOVIC, P. (1969a) Differential effects of real vs. hypothetical payoffs upon choices among gambles. J. exp. Psychol., 80, 434-437.

SLOVIC, P. (1969b) Manipulating the attractiveness of a gamble without changing its expected value. J. exp. Psychol., 79, 139-145.

SLOVIC, P. (1969c) Analysing the expert judge. A descriptive study of a stockbroker's decision processes. J. appl. Psychol., 53, 255-263.

SLOVIC, P. & LICHTENSTEIN, S.C. (1968a) The relative importance of probabilities and payoffs in risk taking. J. exp. Psychol., 78, no. 3 part 2.

SLOVIC, P. & LICHTENSTEIN, S.C. (1968b) The importance of variance preferences in gambling decisions. J. exp. Psychol., 78, 646-654.

SUPPES, P. & WINET, M. (1955) An axiomatization of utility based on the notion of utility differences. Management Science, 1, 259-270.

SUPPES, P. & ZINNES, J.L. (1963) Basic measurement theory. In R.D.Luce et.al (Eds.), Handbook of mathematical psychology, volume 1. Wiley.

THURSTONE, L.L. (1959) The measurement of values. Chicago U.P.

TOCHER, K.D. (1963) The art of simulation. English U.P.

TVERSKY, A. (1967a) Additivity, utility and subjective probability. J. math. Psychol., 4, 175-201.

TVERSKY, A. (1967b) Utility theory and additivity analysis of risky choices. 75, 27-36.

TVERSKY, A. (1969) Intransitivity of preferences. Psychol. Rev. 76, 31-48.

URBAN, J. (1908) The application of statistical methods to the problems of psychophysics. Psychol. Clinic. Press.

VAN DER MEER, H.C. (1963) Decision making: the influence of probability preference and expected value on gambling strategy. Acta Psychol., 21, 231-253.

WALLSTEIN, T.S. (1971) SEU theory and subjects' probability estimates : use of measurement free techniques. J. exp. Psychol., 88, 31-40.