This is to certify that the

thesis entitled

Using Decision-Analytic Techniques to Constrain the Parameters of and to Evaluate Models of Human Rational Decision Processes:

With Application to Diagnosis and Treatment Decisions

presented by

Thomas Henry Whalen

has been accepted towards fulfillment
of the requirements for

Ph. D. degree in Systems Science

Major professor

Date July 26, 1979

O-7639

USING DECISION-ANALYTIC TECHNIQUES TO CONSTRAIN THE PARAMETERS OF AND TO EVALUATE MODELS OF HUMAN RATIONAL DECISION PROCESSES:

WITH APPLICATION TO DIAGNOSIS AND TREATMENT DECISIONS

By
Thomas Henry Whalen

A DISSERTATION

Submitted to
Michigan State University
in partial fulfilment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Electrical Engineering and Systems Science

1979

ABSTRACT

USING DECISION-ANALYTIC TECHNIQUES
TO CONSTRAIN THE PARAMETERS OF AND TO EVALUATE
MODELS OF HUMAN RATIONAL DECISION PROCESSES:

WITH APPLICATION TO DIAGNOSIS AND TREATMENT DECISIONS

By

Thomas Henry Whalen

A new methodology for the study of human decision making processes is introduced. The method, called the "cartographic paradigm," observes a human decision maker's choices among costly sources of imperfect (probabilistic) information in order to make deductions about internal cognitive processes by which a final decision is made, by comparing the human's choices with those predicted by a formal theoretical model for various regions in the model's parameter space. The dimensions of the parameter space are psychologically meaningful variables characterizing individual differences in decision-making approach, and the comparison between the model and actual human behavior is facilitated by a map of parameter space demarcating the regions which lead to different behavior patterns in response to a given decision-making task -- hence the name "cartographic paradigm".

The paradigm, which is intended for use with a particular class of theoretical models, is demonstrated in a small pilot study using one such model, the Myopic Conservative Bayesian Decision Maker. This model was found to be

inadequate as an explanation of behavior at the levels of motivation and expertise studied, especially when data purchased later in the decision process tends to contradict the hypothesis favored by data purchased earlier.

As an additional check on the assumptions of the Myopic Conservative Bayesian Decision Maker, the pilot study also collected data on self-reported subjective probability according to a paradigm developed by Edwards (1968). Analysis of this data revealed subjective probability generally in excess of the optimal Bayesian estimate in the light of data selected and paid for by the subject; earlier studies in which the subject had no control over the amount or type of data showed the opposite effect.

Concepts such as differential processing of consistent and inconsistent data, batching of information purchases, and premature termination of information gathering are briefly discussed as candidates for inclusion in improved theoretical models. Opportunities for future research involving the paradigm include, in addition to laboratory studies, a continuing research project in structural analysis of problem complexity as related to human problem solving and an opportunity for applied research in the decision support system component of a management gaming system.

A successful program of research on human decision processes carried out according to the cartographic paradigm could benefit professional education in a number of fields by providing a multidimensional means of measuring the differences between students and expert decision makers in any given specialty. Research of this nature could also benefit designers of organizational and automated decision support systems by identifying feasible enhancements to a given decision maker's information processing capabilities, changing the parameters of the decision making process so that they lie in a region which improves expected performance without introducing elements alien to the thought processes of the human decision maker who must retain responsibility.

ACKNOWLEDGEMENTS

Thanks are due a number of persons and institutions for assistance and support during the preparation of this dissertation. First of all, I would like to thank my advisor, Dr. Sui-Wah Chan of the College of Human Medicine (Office of Medical Education Research And Development and the Dean's Office for Educational Programs) for his guidance, support and encouragement over the past two years, and Dr. Gerald Park of the Department of Electrical Engineering and Systems Science for his work as co-chairman of the guidance committee. Thanks also go to the other members of the committee: to Dr. Robert Schlueter, who provided a point of continuity to my entire graduate program by also serving on the guidance committees for my M.S. in Systems Science and my M.A. in Sociology; to Dr. John Kreer; and to Dr. Thomas Manetsch.

Informal seminars with fellow graduate students Allen Knapp, Douglas Franco and Suzanne Jennings served as important challenges to develop and organize my thinking.

Computer access was provided through the Medical Education Resources Sharing Network (CDC 6500 and Amdahl 470), the Department of Electrical Engineering and Systems Science (CDC 6500) and the Computer Institute for Social Science Research (Hewlett-Packard 2000). The text of the dissertation was prepared using the word-processing capability of the HP 2000.

Patricia Simon assisted in the entry of text and code into the computers, and conventional typing help was provided by Sharon Schwab and Marguerite Savage. Merald Clark drew figures 1 and 2 and inked in special characters and Greek letters.

The Michigan State Housing Development Authority facilitated my studies during the early part of the program by re-arrangements and reductions of my hours of work to fit my course schedule; later, support was provided by graduate assistantships within the Michigan State University College of Human Medicine.

Last, but not least, I am sincerely grateful to my wife, Evelyn, for her encouragement and understanding during the years of graduate study, and especially for reading and listening to innumerable preliminary fragments of this dissertation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

SYMBOL	DEFINITION	INTRODUCED ON PAGE
S	Set of states of the world	28
s	Number of states in S	28
σ_i	One particular state in S	28
X	Set of alternative final choices	28
c	Number of choices in X	28
x_j	One particular final choice in X	28
Y	Payoff Matrix	29
y_{ji}	Payoff for x_j when σ_i holds	29
Y_j	s-vector of possible payoffs for x_j	29
Q	Set of available tests	29
$\emptyset$	Number of available tests in Q	29
q_k	One particular test in Q	29
C_k	Cost of test q_k	29
R_k	Set of possible results of q_k	29
n_k	Number of possible results in R_k	29
r_k^ℓ	One particular result of test q_k	29
$F_k(r_k^\ell,\sigma_i)$	Representativeness of r_k^ℓ to state σ_i	29
t	Discrete time counter	31
x_{j*}	The most favored x_j at a given time	31
π_t	s-element vector of state saliences	32
π_t^i	Salience of σ_i at time t	32

m	Time when a final choice is made	32
$J_m(x_j, \pi_m)$	Expected payoff for act x_j at time m	32
$\hat{\alpha}_t$	Model's action (q or x) at time t	34
q_{k_t}	Test purchased by model at time t	34
$\hat{\pi}_t$	Model's estimate of π_t	34
$\hat{\pi}_t^i$	Model's estimate of π_t^i	34
ψ_M	Model's set of parameters	36
Ψ_M	Space defined by ψ_M	36
$\hat{B}$	List of the model's behavior	36
φ	Point in Ψ_M giving parameter values	37
$M_{(S,X,Y,Q)}(\varphi)$	Model viewed as a $\hat{B}$-valued function	37
z	A region in Ψ_M	38
ψ_1	Conservatism	42
$E_{S,Q}(\hat{\pi}_t, r_k^\ell, \psi_1)$	Conservative update function	42
ψ_2	Myopia	44
$V(k, \hat{\pi}_{t+1}, \psi_1, \psi_2)$	Value of test k at time t+1	45
q_{k^*}	The most favored q_k at a given time	47

1. INTRODUCTION

1.1 OVERVIEW

The goal of this dissertation is to develop a new methodology for studying the ways human beings use imperfect information in making economic decisions under conditions of uncertainty. Previous research has presented subjects with a fixed quantity of information and used statements of (subjective) probability as the response variable (Edwards, 1972); in contrast, the method to be presented here will use a task environment in which the stimulus consists of opportunities to purchase probabilistic information prior to a final choice with an uncertain payoff, amd the response consists of the sequence of information purchases and the final choice selected. The method is fundamentally an empirical one, but it demands a highly-developed theoretical model to serve as the hypothesis to be evaluated. The result of analyzing one such decision maker's behavior in terms of a given theoretical model may be either the classification of the decision maker within a region of the model's parameter space, or the statement that the model is inadequate to explain the observed behavior.

The purchases of costly information can be considered as a behavioral 'window' into the decision maker's cognitive processes, since they allow a clearer view of what information he considers so valuable in arriving at a decision that he is willing to pay for it. In addition, the study of choice among costly sources of probabilistic information is directly relevant to such practical problems as market research studies or medical diagnostic tests, which require careful balancing of the costs and benefits of information gathering. Improved understanding of this process can benefit professional education by analyzing the differences between students and experts; it can also benefit decision support systems by identifying natural extensions to human information processing which do not dilute the understanding needed for responsible control.

The paradigm is based on a system composed of a decision-making task, a human decision maker, a candidate theoretical model of human decision-making, and an observer who compares the behavior of the model with that of the decision maker. In Chapter 2, the decision-making task will be formally defined in terms of: (1) a set of possible states of the world; (2) a set of alternative final decisions; (3) a matrix giving the payoff for each final decision under each state of the world; and (4) a collection of opportunities to purchase costly probabilistic information about states of the world. The decision maker will be described by a set of axioms characterizing his goals, resources, and constraints. A block diagram will show the interface between the general system and any particular theoretical model. A key part of any model in this paradigm is the definition of a parameter space within which the characteristics of the individual's approach to the problem, as inferred from his behavior, may be placed. An additional set of axioms corresponds to the observer's resources and constraints, and to his goal, which is to find the region of the model's parameter space within which the representation of the decision maker's characteristics must lie if the model is to explain the decision maker's behavior (or to determine that no such region exists). Because of the central role played by a map of the abstract parameter space, the methodology will be referred to herein as the "cartographic paradigm."

The first prerequisite of the cartographic paradigm is a well-defined hypothesis about the decision-making process. This hypothesis must be expressed in a form capable of making predictions about an individual's sequence of information purchases and final choice given a particular problem. In its present form, the methodology is only applicable to deterministic models; however, stochastic models can be accommodated by straightforward generalizations using either conventional statistics or fuzzy set theory.

Since different individuals (or different populations in a stochastic model) behave differently when solving the same decision problem, the theoretical model will have one or more parameters corresponding to psychologically significant inidividual differences presumed to underly the differences in behavior. The overt actions of information purchases and final decision must be treated as discrete selections out of finite choice sets, but the parameters of the model may lie on whatever interval, ordinal, or nominal scales are theoretically appropriate.

1.2 MYOPIC CONSERVATIVE BAYESIAN DECISION MAKER

The central focus of this research is the cartographic paradigm as a general methodology for evaluating theoretical models of human information processing and decision making in a context of costly imperfect information; the Myopic Conservative Bayesian Decision Maker introduced in this section is intended primarily as a simple example of a theoretical model which satisfies the formal requirements of the paradigm. While the model is based on well-established empirical and analytic research (Raiffa,1968; Edwards, 1968,1972; Schum and Martin,1968; Gorry et al, 1973), it is not proposed as a complete substantive statement about human behavior, only as a plausible first step which incorporates two concepts, conservatism and myopia, from out of the many that must be examined in the construction of a full-scale model. A few of the most important of these omitted concepts are discussed in Chapter 4.

The overall structure of the Myopic Conservative Bayesian Decision Maker is derived from the decision analysis algorithm (Raiffa, 1968), in which Bayes' theorem and dynamic programming are used to evaluate each possible sequence of information purchases and final decision so as to maximize expected payoff net of the cost of information. Because of this, the model, like the decision analysis algorithm, assumes a well-structured problem in which all the costs and payoffs are measured on a common scale of utility and all the prior and conditional probabilities are known. This restriction imposes a limitation on the experimental task environments that can be used with this model; in fact, it excludes those ill-structured problems with no provably optimal solution for which understanding the strategies of expert and would-be expert human beings is of the most practical importance. However, if a well-structured problem is sufficiently complex relative to the time and other resources available for its solution, or if the cost of information processing is not negligible compared to the costs and payoffs of the task environment, then the exhausitve analysis called for by the decision analysis algorithm is no longer optimal; instead, the problem must be solved using the same kind of heuristic approaches that are used for ill-structured problems. Since different individuals will solve the same problem in different ways in such a case, the Myopic Conservative Bayesian Decision Maker is provided with two parameters, conservatism and myopia, in an attempt to capture some of this variability. It is conjectured that the similarities and differences among individuals measured by research in this environment can later be extended to similarities and differences in behavior on ill-structured problems, so that research findings using well-structured experiments can be relevant to the wider range of problems in which the "recognized expert" is the only standard of quasi-optimality.

The concept of conservatism is derived from the work of Ward Edwards (1968, 1972). In Edwards' experiments, subjects were presented with (costless) data from one of two initially equally-likely binomial probability distributions, and asked to estimate the a posteriori probability that the data was derived from a given distribution. When the two distributions differred strongly, subjects changed their opinion less than the data warranted, but when the distributions were nearly the same, subjects tended to overvalue the data by changing their opinions more than was warranted.

Edwards found that these results could be summarized by using a modification of the odds form of Bayes' Law in which the posterior odds equal the prior odds times some power "c" (for "conservatism") of the odds ratio for the data observed. When c=1, this is simply the optimal, Bayesian formulation; c=0 would imply ignoring the data totally and sticking to the original belief that the two distributions are equally probable, while c = infinity would mean jumping to complete certainty on the basis of any data whatever. The exponent c can thus be used as a measure of conservatism, with lower numeric values corresponding to more conservative behavior. Schum and Martin (1968) found that Edwards' model applied equally well to a problem in which six multinomial distributions took the place of the two binomial distributions used by Edwards. Thus, to accomodate this finding about human information processing, the model uses Edwards' general equation, of which Bayes' is a special case.

It should be noted that, even if the a priori subjective probabilities add up to one, as new data is used to update these subjective probabilities, the total of the resulting a posteriori probabilities will vary. Edwards (1967) gives a mathematical basis for theories of subjective probability distributions that need not add up to one; in fact, Edwards casts considerable doubt on the possibility that any consistent theory of subjective probabilities satisfying reasonable assumptions could require that the sum of the subjective probabilities of a set of exclusive exhaustive events must add up to any pre-specified constant.

While the model thus addresses an issue first raised by Edwards, the methodological technique differs sharply from his. Prior research has focused primarily on measuring the importance of information to human decision makers in terms of its after-the-fact effect on subjective probabilitites, either by asking subjects to give introspective statements of their confidence in a hypothesis or by offering choices among simple bets based on the hypotheses. The cartographic paradigm, on the other hand, draws the majority of its data from economically-motivated choices among data sources; the impact of information will be measured more by the circumstances under which a subject is willing to pay for that information than by backwards inference from subsequent behavior.

Analysis of economically-motivated purchase of information leads one immediately to the literature of information value theory, which in principle can provide the optimal choice of information purchase or final decision at all times, given either Bayes' or Edwards' model of information processing. However, as Gorry et al (1973) point out, the computational burden of this algorithm quickly becomes insupportable for problems of even moderate complexity. In order to implement their computer program for medical diagnosis and treatment selection, they evaluate tests as if it were necessary to diagnose immediately after getting the results of the test under consideration, neglecting the option of continuing to test. (Once the test has been selected and the results have been received, the process is repeated and another test may, in fact, be purchased if the value of so doing exceeds that of an immediate diagnosis; thus, the expected value of a test by Gorry's method is a lower bound for the "true" expected value given by the full decision analysis algorithm.)

The "myopia" parameter of the Myopic Conservative Bayesian Decision Maker is derived from a generalization of Gorry's variant of decision analysis. Gorry's model, which looks only one move ahead, corresponds to the present model with a myopia of 1. When myopia = 2, each possible test is evaluated as if at most one additional test could be purchased after getting the results of the test under consideration. The original decision analysis algorithm corresponds to the present model with a myopia parameter equal to infinity, or at least very large with respect to any observed sequence of actions; in this case, there is no arbitrary limit on the length of testing sequences that may be considered.

This completes the overview of the "Myopic Conservative Bayesian Decision Maker." This model is an example of the class of models to which the cartographic paradigm is applicable; any such model can be viewed abstractly as a function which, in the context of a given decision-making task, maps a point in an abstract parameter space (such as the space defined by conservatism and myopia) to a particular sequence of observable behaviors.

1.3 METHOD OF STUDY

The goal of the cartographic paradigm is to connect the theoretical insight expressed in such a formal model with a system of observations, in order to learn about the similarities and differences among decision makers. Sequences of information purchases and final choice are replaced by equivalent points or regions in an abstract space of psychological parameters hypothesized to underly behavior on the decision-making task.

For a one-to-one function from parameter space to behavior sequence, this conversion can be accomplished simply by inverting the function. However, a one-to-one model has a serious drawback; if two decision makers emit the same behavior sequence in response to one particular decision-making task, they cannot differ from each other on any decision-making task to which the model applies. A more flexible model, such as the Myopic Conservative Bayesian Decision Maker, will produce the same behavior sequence within small regions of its parameter space for a given decision-making task, but different behaviors for a sufficiently distant point in parameter space. A different task will define a different partition of the same parameter space, thus allowing for two points to be in the same region for one decision-making task but two different regions for another.

The cartographic paradigm developed for dealing with such models begins with the precise specification of the model -- a computer simulation will often be the most efficient form, although simulation is not a sine qua non of the method. In addition to the model, the task environment must also be selected. The task environment must be rich enough to elicit a significant variety of behavior, yet simple enough to be understood equally well by all the subjects and to keep the analysis of the results tractable. In a laboratory environment, the characteristics of the task may be "tuned" by simulation runs using various combinations of task and parameter values, or analytically using model equations.

Once the model and the task have been specified, the behavior of the model in response to the task must be studied. The result of this stage in the research procedure is a collection of all the behavior patterns of which the model is capable, together with the regions in parameter space corresponding to each pattern. Following this, human decision makers are presented with the same problem, and classified into the regions of parameter space corresponding to their observed behavior patterns. Any subject whose behavior does not match the behavior of the model for any parameter values is not describable by the model; while such an occurrence need not totally destroy the usefulness of the model as a description of some decision makers, it would at the least mean a theoretical incompleteness in the model, if not grounds for its rejection.

A possible practical application of this type of analysis, should it prove successful, lies in the training of student decision makers, such as medical students. If a model is discovered such that skilled physicians cluster in an identifiable region of that model's parameter space, the ways in which the behavior of a student differs could be analyzed in terms of that parameter space. This would provide a clearer language with which to give feedback to the student, as well as the possible basis for a new educational technology. Another benefit could be in the design of organizational and automated decision support systems to extend human information processing capabilities without sacrificing the comprehension needed for responsible human control.

1.4 REVIEW OF RELATED RESEARCH

This section of the dissertation will review some representative examples of each of four general methodologies for the study of human information processing and decision-making: Protocols, Subjective Probabilities, Weighting Coefficients, and Information Choice. Each study, in turn, will be discussed in terms of the kind of information provided to the decision maker, the nature of the decision to be made, the implicit or explicit assumptions about human information processing in general, and the role of individual differences. In the final part of the section, (1.4.5), the contributions of each of the four schools of research to the present study will be analyzed.

1.4.1 Protocols

The most direct way to begin to learn about how a person goes about making a decision is to ask him. The best-known example of this approach is the work of Newell and Simon (1972). In this research, subjects were presented with a complex and highly-structured task environment such as a chess position. All the data to be used in making a decision (the next move in the chess problem) were there at the start, and the subject were instructed to "think aloud" in the course of deciding what his best course of action might be. The recordings of these sessions, called "protocols," were then interpreted by the researchers in terms of a general theory of problem solving. The result of this interpretation was a computer program that simulated the behavior of one individual subject on the given problen; both the final choice and the general sequence of processing steps (corresponding to the verbal reports of thought processes) were simulated. In effect, Newell and Simon's assumptions about human information processing are mostly imbedded in the structure of their programming language (IPL-V), while separate programs are written to conform to the idiosyncracies of individual human subjects.

Kleinmuntz (1968) presents two protocol-based studies of medical diagnostic reasoning. The first study, in which a single expert psychologist thought aloud while classifying a number of MMPI profiles on a scale from maladjusted to well-adjusted, was similar in outlook to Newell and Simon' work. The second of Kleinmuntz' studies, however, adds the dimension of information-seeking. In this experiment, neurologists at varying levels of training and experience were presented with a brief description of a hypothetical patient. The subject then proceeded to ask yes-or-no questions about the patient until he reached a diagnosis. The same collection of cases was diagnosed twice, with a lapse of several weeks between sessions; the second time, subjects were asked to state their reasons for each question they asked.

Critics of protocol research have pointed out the danger that the introspections called for may alter the mental processes and so invalidate the findings; Kleinmuntz' method allowed a partial measure of this effect through comparison of the questions asked by a given subject with and without thinking aloud -- a test-retest rank order correlation coefficient of .92 was found, which lends credibility to the assumption that the thought processes with and without verbalization were comparable. Based on this, Kleinmuntz was able to state several hypotheses about the ways that neurologists in general go about selecting what question to ask next, and how students, residents, and experienced clinicians differ. These hypotheses are assessed in a general way in terms of the data, but no specific predictions of behavior like those of Newell and Simon or Kleinmuntz' MMPI study are attempted in this less-structured task environment.

1.4.2 Subjective probability

Most of the research on subjective probability in the last fifteen years has been influenced to a greater or lesser extent by the work of Ward Edwards. Edwards (1968, 1972) obtained human estimates of the probability of different "states of the world" in situations where the prior probabilities of the states are known and data with known conditional probability given each state are to be used in obtaining an estimate of the posterior probability; he then compared these human estimates with optimal estimates derived using Bayes' rule. (The concept and mathematical operationalization of "conservatism" used in the Myopic Conservative Bayesian Decision Maker derives from the results of Edwards' research in this area.) Edwards used many different experimental situations, but the simplest of them will sufice to show his paradigm. In this experimant, the subject was told that there existed two bookbags, one with p red poker chips and 100-p white poker chips, and the other with p white poker chips and 100-p red ones. One of the bags had been chosen by the flip of a coin, and a number of chips were drawn with replacement from the selected bag. The subject was then shown the results of this sample, and asked to state how likely he feels it is that the chips were drawn from the mostly red bag. Many variations and elaborations of this scenario have been used, including choice among bets or bidding for bets instead of directly stating numeric probability estimates, the use of unequal prior probabilities, et cetera. -- for details, see Slovic and Lichtenstein (1971). Most of the work by Edwards and others on subjective probability that was published in time to be reviewed by Slovic and Lichtenstein was concerned with comparing human performance with optimal Bayesian performance on this well-defined task; the consistent finding was that humans change their opinions about probabilities too little when the data are strongly diagnostic, and too much when the data are weak.

More recent studies have moved away from reliance on Bayes' rule to search for a formula or formulas which can more accurately reflect human information processing. Wallsten (1976), an outstanding example of this trend, uses four different algebraic "composition rules" as alternative models for how subjects use probabilistic information to alter their a priori subjective probabilities, and tests these models for goodness of fit using sophisticated rank-order based mathematical techniques which avoid earlier researchers' reliance on the numeric qualities of self-reports of subjective probability. The models are analyzed for goodness of fit using two different measures, on a subject-by-subject basis as well as overall; the principal model was found to fit all subjects, while one of the three alternative models was capable of fitting all but two, and the other two alternatives showed substantial lack of fit with respect to every subject.

1.4.3 Weighting Coefficients

Research aimed at finding numeric weights to represent the relative importance a given human judge attaches to various aspects of a person or thing he is evaulating date back to Wallace's 1923 paper in the Journal Of The American

Society Of Agronomy. In that study, experienced corn judges in Iowa rated 500 ears of seed corn in terms of likely yield when planted; correlation coefficients and multivariate path coefficients between the estimated yields and various significant characteristics of the corn itself (e.g. length of ear, weight of kernel) were then computed "to make out the score card which really existed in the judges' minds" (Wallace, 1923).

Slovic and Lichtenstein (1971) review a large number of studies performed in the 1950's and 1960's in which subjects made a number of judgements (usually of students or of psychiatric patients) on a numeric or dichotomous scale on the basis of a set of characterisitcs of the persons being judged. These subjective judgements were then compared with the results of a multiple regression analysis with the same characteristics as predictors, and either the judgements or an objective criterion taken as the dependent variable. The studies, taken as a group, seemed to firmly establish the hypothesis that human judgment can be well represented, at least "paramorphicly," by a simple weighted sum of the input data, and furthermore that, when human judgement did differ from the linear prediction of the judgement itself, the linear prediction was more likely to match the criterion! (A "paramorphic representation" is simply a black-box model which matches the system modeled in an input-output sense;

the concept was used to separate the empirical demonstrations of the efficacy of the regression equation from the theoretical or introspective question of whether the regression beta weights really capture what is "in the judge's mind.")

These results were the center of a long and heated "clinical versus statistical judgement" controversy, summarized by Slovic and Lichtenstein (1971). At one extreme were those who saw the linear models as providing a major advance in man's ability to improve his own consistency and accuracy of judgement and to convey the judgemental processes of many professions to the next generation of students, by using precisely determined weights for evidence that capture the optimal policy. These views were opposed by others who decried the loss of nonlinear intuitive or gestalt judgements based on human interaction with both the "hard" data processed by the equations for all cases, and the "soft" data that is unique to each particular case, especially when (as was usual in this research) it was human beings that were being evaluated.

However, more recent work by Dawes and Corrigan (1974) has cast the question in an entirely new light; they found, in five different judgement tasks, that equations with random coefficients constrained only as to sign are, on the average, at least as accurate as equations modelling the judgements of human experts or equations derived by standard multiple regression techniques to optimally predict the criterion, given the sample size generally used. Dawes and Corrigan analyzed this surprising result mathematically; in summary, they found it to be due to the extreme generality of the linear approximation technique, combined with the very low sensitivity of the regression equation to random changes in the relative weights of the coefficients when the predictor variables are intercorrelated. Thus, while finding the relative importance of various sources of evidence in an individual's decision-making may remain a desirable goal, it now appears highly questionable whether any method of fitting a linear equation to observed (or optimal) judgements can be of much help.

1.4.4 Choice of Information

The systematic study of subjects' choices of information from a defined set of alternatives begins with Glaser et al (1954). In that study, electronics trainees were presented with a description and diagram of a malfunctioning electronic device and a set of possible tests they could perform to troubleshoot. Next to each test was a paper tab that could be ripped off to reveal the result of the test printed underneath. The purpose of this was to evaluate the proficiency of each trainee on the basis of the appropriateness of the tests selected, as judged by experts in electronics. This goal of evaluating individual competence in terms of conformity to some standard has characterized the majority of research of this type.

A number of papers in the applied area of medical diagnosis have focused on the questions (medical history, physical examination, or laboratory) asked by human physicians and medical students about real or simulated patients. Rimoldi (1964) evaluated questions by measuring how frequently they were asked by members of a population of physicians and medical students, under the assumption that the questions that were asked the most were the ones with the highest utility for the diagnostic task. He also proposed using the degree to which the cumulative utility of the best observed sequence of questions differs from that of the worst observed sequence as a metric for the difficulty of the task. Sprosty (1964) compared the sequences of questions asked by students who ended up with the correct diagnosis with the sequences of questions asked by students who diagnosed incorrectly. Finally, Barnett (1972) used a sequential Bayesian analysis to determine the increase in certainty provided by the answer to each of a set of possible questions, using this value to evaluate the effectiveness of students' and clinicians' question sequences. None of these studies, however, has really attempted to model the process by which a person actually decides when to ask what question and when to stop asking and make a decision.

The work of Fried and Peterson (1969) is particularly relevant to the present research because they used explicit monetary costs for information and payoffs for correct decisions, whereas in the other studies cited above the only cost of information was the risk that the question might not conform to the standard. Fried and Peterson offered subjects only one kind of information (samples from one of two binomial distributions), and compared the amounts of information purchased under conditions of fixed stopping (where the subject must decide on the sample size before he sees any of the data) and optional stopping (where the decision whether to continue sampling is made after each binomial event). Subjects conformed well to the optimal Bayesian sample sizes under fixed stopping, but undersampled under optional stopping; the authors discuss possible interpretations of this result in general terms, but do not offer a predictive model.

1.4.5 Comparison With the Present Research

The cartographic paradigm introduced in this dissertation is related to the protocol studies (section 1.4.1) in the use of a formal, algorithmic model of human behavior to predict the steps a person goes through in making a decision; it differs, however, in avoiding introspection in favor of information purchases as the means of detecting these intermediate stages. Another difference is that the cartographic paradigm implicitly calls for more highly-structured theories of human information-processing, as compared with research such as that of Newell and Simon (1972) where most of the general theory of human information processing is embedded in the syntax and semantics of the language used to write the simulation programs. (For discussion of this issue, see Scandura, 1977).

The present research resembles the research in subjective probability (section 1.4.2) in its concern with the effects of new information on a person's prior opinions. In addition, the particular theoretical model used in the pilot experiment, the "Myopic Conservative Bayesian Decision Maker," uses experimental stimuli and a system state variable ($\hat{\pi}$) that adhere very closely to precedents established in subjective probability research. In future research using different theoretical models, this latter similarity of probabilistic stimulus and quasi-probabilistic state variables will not necessarily continue to hold, although the concern for the effect of information on opinion will remain a central concern of the overall paradigm.

The research on weighting coefficients (section 1.4.3) resembles the new paradigm in the effort to find a common scale of measurement for the impact of qualitatively different kinds of information. In prior research, this scale has been a statistical one such as regression coefficients, while the method being presented in this dissertation uses an economic measure to assess the relative impact of different kinds of information.

The data collection procedures of the cartographic paradigm are a direct development of those of the "choice of information" research discussed above (section 1.4.4). Several important elements of difference do exist, however. Whereas earlier research generally relied on presumed analogies to practical problems and role-playing by the subjects to establish the costs of information and benefits of correct decisions, the present paradigm quantifies these costs and benefits with cash payments to the subject. Earlier research used a single dimension of efficiency or appropriateness of search to compare the behavior of different individuals; the present paradigm recognizes the multidimensional variations in ill-structured problem solving, and is specifically designed to use and evaluate multiparameter models which embody hypotheses about the relations between different kinds of problem-solving heuristics. Past research of the "choice of information" variety has contrasted neophyte and expert decision makers in terms of the surface characterisitcs of the sequences of questions they ask about one or a few particular problems; the present paradigm aims at discovering stable underlying characteristics of individuals as revealed by problem-solving behavior, and identifying the clusters of characteristics associated with expert performance.

2. SYSTEM DEFINITION

2.1 OVERVIEW

The purpose of this chapter is to define the theoretical and experimental environment within which research is to be conducted, both in the pilot study reported on in this dissertation and in later larger-scale substantive research which will be carried out using the cartographic paradigm. This environment is analyzed in terms of a system composed of the following elements: a decision-making task with costly, imperfect data (discussed in section 2.2); a human decision-maker who performs the task and experiences the associated costs and payoffs (section 2.3); a candidate theoretical model of human decision-making which attempts to explain the similarities and differences between the behavior of different persons in like tasks (section 2.4); and an observer who compares the behavior of the model with that of the human decision maker (section 2.5). Section 2.6 gives the formal statement of the particular candidate theoretical model used in the pilot study (the Myopic Conservative Bayesian Decision Maker), together with the consequences for the model-specific aspects of other system elements.

2.2 DECISION-MAKING TASK

The following set of formal elements define a class of decision-making tasks. This class of tasks is of interest in part because of the practical importance of related problems such as cost-effective medical diagnostic testing, but also because of the opportunity to use a little-studied behavior, information purchases, to make and test inferences about aspects of human information processing that apply in other situations as well. In this research, mathematical techniques developed in the normative theory of decision-making with costly imperfect data, "decision analysis," will be generalized in an attempt to model human behavior; while the performance of skilled human decision makers does not equal that of the normative model on simple problems subject to exhaustive analysis, human beings seem to maintain an acceptable performance level on complex and ill-structured problems for which the normative model fails to reach any decision within feasible time constraints.

Any individual decision-making task in this class of tasks is identified by the 4-tuple (S,X,Y,Q), where:

(T1) S is a set of s possible states of the world σ_i, exactly one of which is assumed to hold.

$S = \{\sigma_1, \sigma_2, \sigma_3, \ldots, \sigma_i, \ldots, \sigma_s\}$

(T2) X is a set of c alternative final choices x_j, of which the decision maker must choose exactly one.

$X = \{x_1, x_2, \ldots, x_j, \ldots, x_c\}$

(T3) Y is a c-by-s payoff matrix whose ji'th element y_{ji} is the payoff that the decision maker would receive after selecting final choice x_j when the state of the world is actually σ_i. The j'th row of Y, written Y_j, is thus an s-element row vector which gives all the possible payoffs for final choice x_j (depending upon which state of the world holds).

(T4) Q is a set of $\emptyset$ costly tests available to the decision maker,

$Q = \{q_1, q_2, \ldots, q_k, \ldots, q_\emptyset\}$

Any test q_k consists in turn of:

(a) A cost C_k that the decision maker must pay to learn the result of test q_k;

(b) A set R_k of n_k possible results r_k^ℓ of test q_k

$R_k = \{r_k^1, r_k^2, \ldots, r_k^\ell, \ldots, r_k^{n_k}\}$;

(c) A function $F_k(r_k^\ell, \sigma_i)$ with range [0,1] which defines the relative degree to which the result r_k^ℓ to test q_k is representative of state σ_i.

Without loss of generality, the function is normalized so that $\sum_{\ell=1}^{n_k} F_k(r_k^\ell, \sigma_i) = 1$ for all k and i.

(In Bayesian and other statistical models, $F_k(r_k^\ell, \sigma_i) = P(q_k \Rightarrow r_k^\ell | \sigma_i)$, the conditional probability of result r_k^ℓ to test q_k given state σ_i; but see e.g. Tversky and Kanheman (1974) for an alternative formulation of representativeness in human information processing.)

2.3 DECISION MAKER

The decision maker is an individual human being about whom the following axioms are assumed to hold:

(D0) The decision maker encounters the task at time $t=0$, and performs actions at times $t = 1, 2, \ldots, m$, where the m'th action is the selection of a final choice x_{j*} and each of the other actions is the selection of a test.

(D1) At all times $t = 0, 1, 2, \ldots, m$, the decision maker retains complete knowledge of:

(a) the set of possible states of the world S (but not which one holds)

(b) the set of alternative final choices X

(c) the payoff matrix Y

(d) the set of possible tests Q

(e) the results of all tests purchased at any time prior to t.

(Note that the decision maker need not attend to all of this information; what information the decision maker actually uses in making his decision is a matter for particular theoretical models. Axiom D1 can be satisfied by written records or other similar aids.)

(D2) Just prior to the decision maker's action at time t, the decision maker's opinion about the relevance of each of the s possible states in S may be represented by a salience vector

$$\pi_t \stackrel{d}{=} [\pi_t^1, \pi_t^2, \ldots, \pi_t^i, \ldots, \pi_t^s]^T; \quad \pi_t^i \in [0,1]$$

(Salience may be interpreted as subjective probability in models such as the Myopic Conservative Bayesian Decision Maker which define subjective probability.)

(D3) $\pi_1 = [\frac{1}{s}, \frac{1}{s}, \ldots, \frac{1}{s}]^T.$

Initially, saliences are defined to be equal and may, without loss of generality, be scaled to add up to 1.

(D4) After purchasing a test q_k at time t for cost C_k and learning that the result is r_k^ℓ, the decision maker revises his opinion to one represented by π_{t+1}, and selects his next action.

(D5) The decision maker's final choice $x_{j*} \in X$ at time $t = m$ is the choice which maximizes the subjective expected payoff function $J_m(x_j, \pi_m) = \sum_{j=1}^{s} \pi_m^i y_{ji}.$

This definition of the decision maker is intended as a general framework only; necessary details such as the interaction between costs and payoffs or the stopping rule which determines m are left to particular theoretical models such as the Myopic Conservative Bayesian Decision Maker.

2.4 THEORETICAL MODEL

The block diagram in Figure 1 shows the input-output requirements for a theoretical model to be used in this research process.

The lines labeled S, X, Y, and Q indicate that the model's behavior is dependent on the specification of the task known to both the decision maker and the observer: S is the set of possible states, X is the set of alternative final choices, Y is the payoff matrix, and Q is the set of available tests.

$\hat{\alpha}_{t+1}$ is the action selected by the model for time t+1, either purchasing test $q_{k_{t+1}}$ or selecting final choice x_{j*}. If a final choice is selected, $t+1 = \hat{m}$ and the model terminates. Otherwise, the "next" test $q_{k_{t+1}}$ becomes the "last" test q_{k_t} in the next time period. (This is the function of the box labeled "DELAY").

r_k^{ℓ} is the result of the last test q_{k_t}, defined for $t = 1, 2, \ldots, m-1$. It represents the feedback of information from the environment that the decision maker receives in return for incurring the cost of diagnostic tests.

$\hat{\pi}_t$ is an s-element vector which serves as an estimate of the vector π_t representing the decision maker's opinion at time t. (See assumption D2).

$\hat{\pi}_1 = [\frac{1}{s}, \frac{1}{s}, \ldots, \frac{1}{s}]^T.$

$\hat{\pi}_t \stackrel{d}{=} [\pi_t^1, \pi_t^2, \ldots, \pi_t^s]^T$

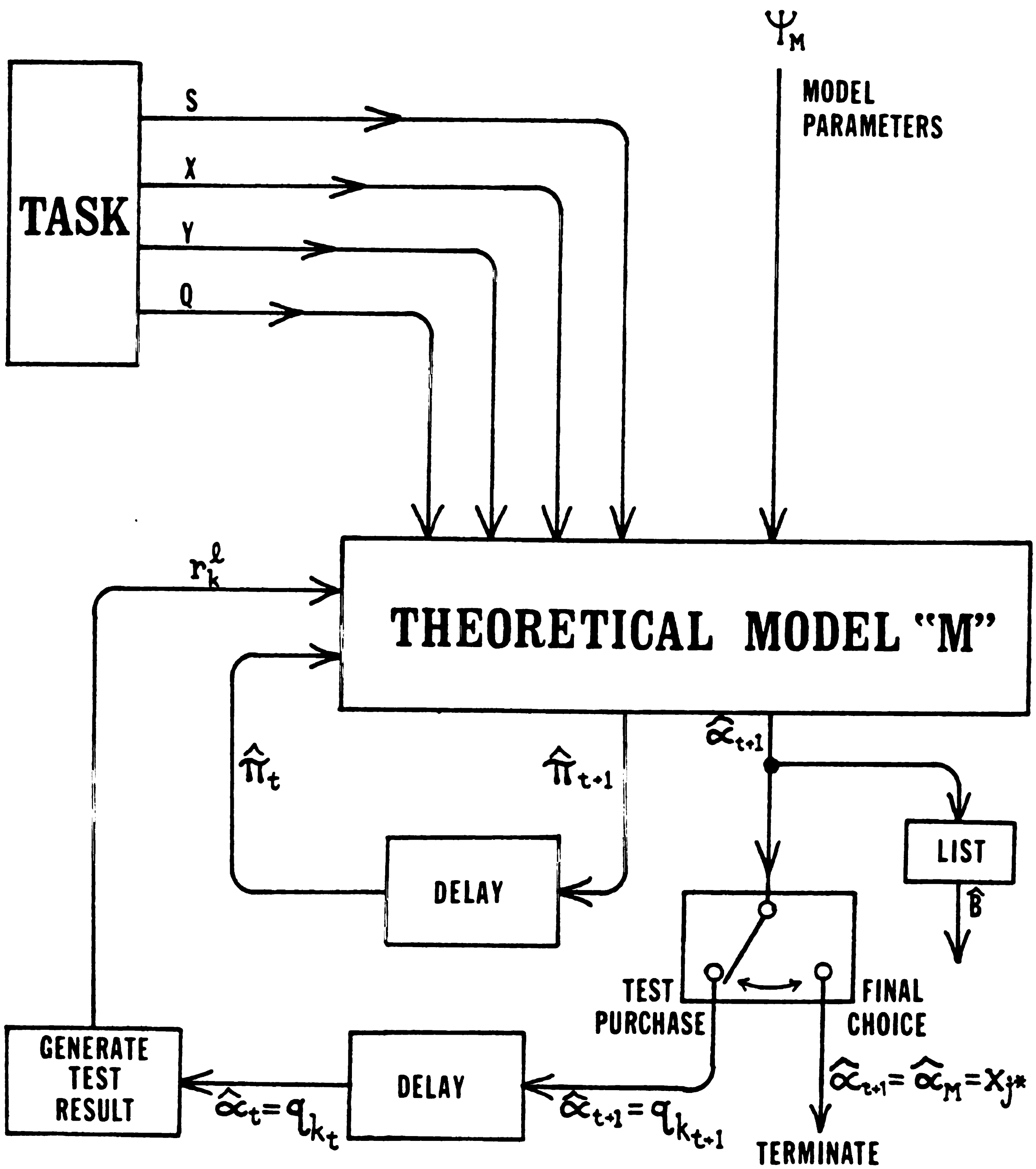


FIGURE 1: BLOCK DIAGRAM

$\hat{\pi}_{t+1}$ is the estimate of the vector representative of the decision maker's new opinion at time t+1, after taking into account the information r_k^{ℓ}. $\hat{\pi}_{t+1}$ becomes $\hat{\pi}_t$ for the next time period, as indicated by the DELAY box.

Ψ_M is the set of parameters by which the model M represents individual differences that affect behavior on the class of decision-making tasks under consideration. The number of parameters and their interpretation will vary from one theoretical model to another, but each parameter should correspond to a psychologically meaningful concept, such as the concepts of conservatism and myopia in the two-parameter Myopic Conservative Bayesian Decision Maker. For a given theoretical model, the values of that model's parameters will vary form one decision maker to another.

The parameter space $\boldsymbol{\Psi}_M$ for a particular theoretical model M is defined as the abstract space whose dimensions correspond to the elements of Ψ_M, the set of parameters of the model M.

$\hat{B}$, the principal output of the model, is the accumulated, ordered list of inferred actions $\hat{\alpha}_t$; that is, the model's sequence of information purchases and its final choice. Since the model inputs $\hat{\pi}_0$ and r_k^{ℓ} are determined by the constant features of the task environment interacting with the model itself, the behavior of a given theoretical model M on a given task (S,X,Y,Q) depends only on the values of the elements of Ψ_M. This can be expressed by representing the model M by a family of generalized functions, indexed by task characteristics, from Ψ_M to the space of possible behavior lists: $M(S,X,Y,Q)(\varphi) = \hat{B}$, where $\varphi \in \Psi_M$. (That is, φ is the specification of all the elements of Ψ_M, characterizing a particular real or hypothetical individual decision maker.

2.5 OBSERVER

The observer is an educator, cognitive scientist, or other person interested in objectively studying the behavior of the decision maker. The role of the observer is given by the following axioms:

(O1) The observer has the same knowledge of the task (S,X,Y,Q) as that assumed for the decision maker in axiom D1.

(O2) The observer knows what test q_{k_t} the decision maker purchases at each time t = 1, 2, ..., m-1 and the final choice x_{j*} the decision maker selects at time t=m. By analogy with the list $\hat{B}$ of the model's behavior, the ordered list of the decision maker's observable behavior, known to the observer, will be referred to as B.

(O3) The observer does not know the decision maker's opinion nor its true representation π_t, for any time t > 1.

(O4) The observer conjectures that model M (from the class of theoretical models defined above) is an adequate representation of the decision maker's behavior; the observer's goal is to find the region $z \subset \Psi_M$ such that if $\varphi \in z$, then $M_{(S,X,Y,Q)}(\varphi) \overset{d}{=} \hat{B} = B$.

If the observer succeeds in finding a region z as in (O4), he will map its boundaries within Ψ_M and state, for a given decision maker, that if the model adequately represents that decision maker's behavior process on the given task, then the values of that model's parameters characterizing the decision maker lie somewhere in region z. If no such region is found, the observer must conclude that the theoretical model M is not adequate to represent that decision maker's behavior on that task.

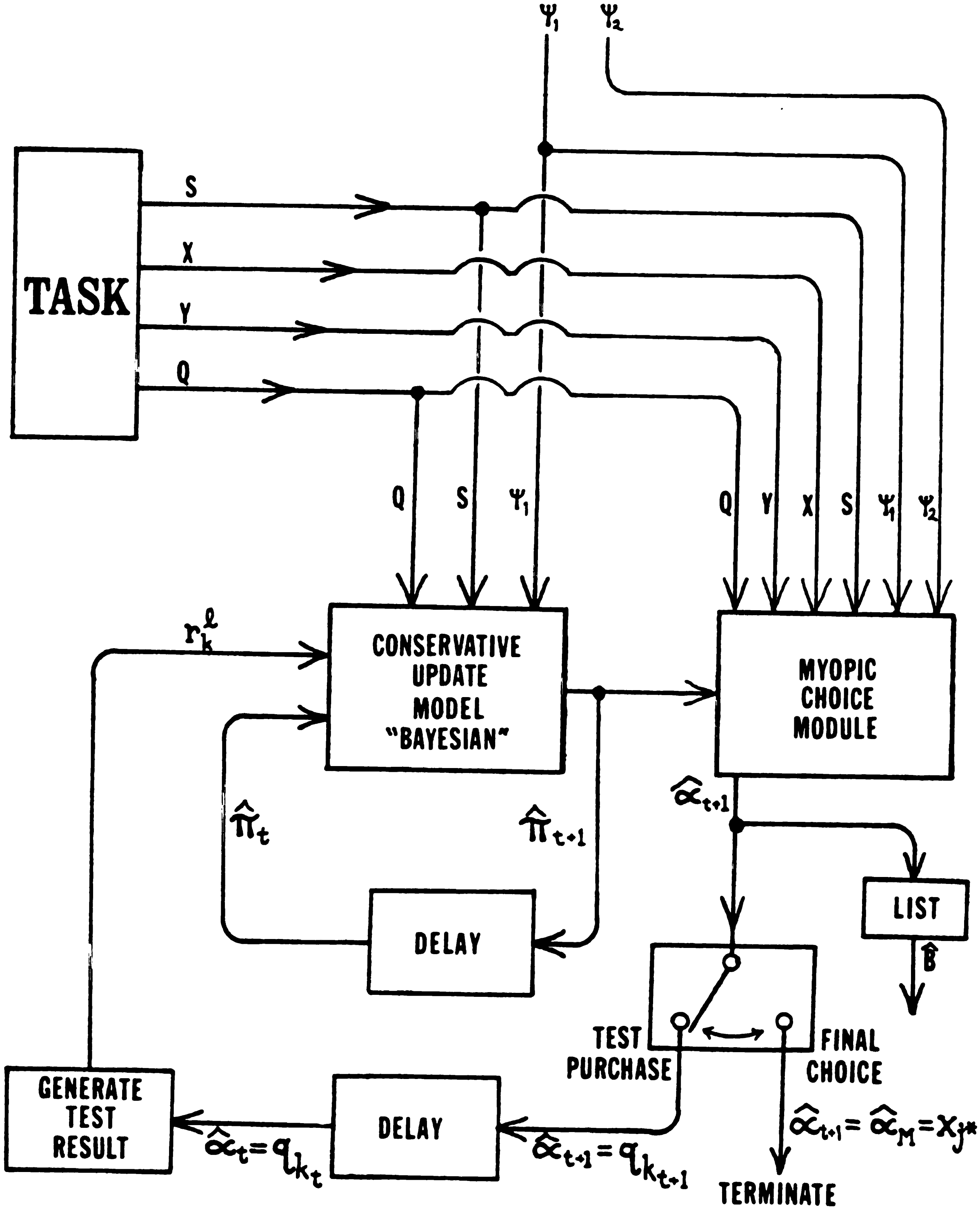


FIGURE 2: MYOPIC CONSERVATIVE BAYESIAN DECISION MAKER

2.6 MYOPIC CONSERVATIVE BAYESIAN DECISION MAKER

The purpose of this section is to give the mathematical specification of the Myopic Conservative Bayesian Decision Maker, the theoretical model introduced in Section 1.2; this model will be used in the pilot experiment descibed in Chapter 3. The model can be conveniently divided into a "conservative" update module which changes the value of the state vector $\hat{\pi}_t$ representing the decision maker's salience vector (here interpreted as subjective probability) in the light of new information, and a "myopic" selection module which determines the next action to be taken (information purchase or final choice).

2.6.1 Conservative Update Module

The update module of the Myopic Conservative Bayesian Decision Maker is based on the findings of Schum and Martin (1968), who showed that Edwards' concept of conservatism in human information processing applies to subjective probability distributions with multiple alternatives, not just the two-alternative case investigated by Edwards. What these investigators found was that human beings' stated judgements of a posteriori probabilities, given equal a priori probabilities and data with known conditional probabilities, can be modeled consistently by a variant of Bayes' rule in which the impact of evidence upon opinion about relative odds is raised to a power (called c by Edwards and by Schum and Martin, but Ψ_1 in this model). Bayes' rule is the special case where $\Psi_1 = 1.0$, but the research cited has shown that Ψ_1 in general is not equal to 1.0 in human information processing. Thus, the value of the exponent Ψ_1 is taken as the measure of "conservatism," with lower values corresponding to more conservative behavior -- that is, behavior in which the impact of new information on previously-held opinion is low.

The update module uses the following algorithm to determine $\hat{\pi}_{t+1}$ from $\hat{\pi}_t$ and $\hat{\alpha}_t$:

If the latest act $\hat{\alpha}_t$ was the selection of a final choice x_{j*}, the model terminates. ($t = \hat{m}$ in this case.)

Otherwise, the estimated probability vector $\hat{\pi}_{t+1}$ is defined by the following equations:

(1) $\hat{\pi}_1 \stackrel{d}{=} [\frac{1}{s}, \frac{1}{s}, \ldots, \frac{1}{s}]^T$.

(2) For any $t \in \{1, 2, \ldots, m-1\}$, let k be the index of the test q_{k_t} purchased at time t, and let ℓ be the index of its result, r_k^ℓ.

Define a vector-valued "Conservative Bayesian function"

E by: $E_{S,Q}\left(\hat{\pi}_t, r_k^\ell, \Psi_1\right) \stackrel{d}{=} \left[\frac{w_{t+1}^1}{1+w_{t+1}^1}, \frac{w_{t+1}^2}{1+w_{t+1}^2}, \ldots, \frac{w_{t+1}^s}{1+w_{t+1}^s}\right]^T$, where $$w^i_{t+1} \stackrel{d}{=} \left(\frac{\hat{\pi}^i_t}{1-\hat{\pi}^i_t}\right)\left(\frac{(F_k(r^\ell_k,\sigma_i))\ (1-\pi^i_t)}{\sum_{h=1}^{s} F_k(r^\ell_k,\sigma_h)\hat{\pi}^h_t - F_k(r^\ell_k,\sigma_i)\hat{\pi}^i_t}\right)$$

is the updated conservative odds for state of the world σ_i, which reduces to the updated Bayesian odds when Ψ_1 is equal to 1.0.

(3) $\hat{\pi}_{t+1} \stackrel{d}{=} E_{S,Q}(\hat{\pi}_t,\ r^\ell_k,\ \Psi_1)$ for $t \in \{1, 2, \ldots, m-1\}$.

(See Appendix 1 for the derivation of this formula.)

2.6.2 Myopic Selection Module

The role of the myopic selection module of the Myopic Conservative Bayesian Decision Maker is to connect the internal opinion states, generated by the conservative update module, to the observable actions (test purchases and selection of final alternative) which are the focus of the research paradigm. The module, derived from a variant of decision analysis used by Gorry et al (1973) for computerized medical diagnosis and treatment recommendations, measures myopia by the parameter Ψ_2; the selection module is "myopic" in the sense that, at each time t, each diagnostic test in the set Q is evaluated as if it were necessary to cease testing and choose a final treatment from the set X at or before time $t+1+\Psi_2$. (Time, in the model, is measured by the number of actions that have occurred, where an action is a test purchase or the selection of a final choice). Thus, after learning the results of the t'th test, the model considers at most Ψ_2 additional tests. (Gorry et al consider only the case $\Psi_2 = 1$.)

The selection module determines the act $\hat{\alpha}_{t+1}$ using the updated estimated subjective probability vector $\hat{\pi}_{t+1}$ by the following myopic decision analysis algorithm, with integer-valued myopia parameter Ψ_2:

(A) Compute the subjective expected value to the simulated decision maker of selecting final alternative x_j at time t+1, for each $j \in \{1, 2, \ldots, c\}$. This value is found by the sum of the payoffs conditional on each state of the world weighted by the subjective probabilities of the states: $\sum_{i=1}^{s} y_{ji} \hat{\pi}_{t+1}^{i} \stackrel{d}{=} Y_j \hat{\pi}_{t+1}$.

(B) Determine the best final alternative x_{j*} such that $Y_{j*}\hat{\pi}_{t+1} \geq Y_j\hat{\pi}_{t+1}$ for all $j \in \{1, 2, \ldots, c\}$.

(C) Compute the net value of the simulated decision maker's expectation assuming that, rather than selecting a final alternative, the decision maker purchases test q_k at time t+1, for each $k \in \{1, 2, \ldots, \emptyset\}$. This value, denoted $V(k, \hat{\pi}_{t+1}, \Psi_1, \Psi_2)$, is found by the following recursive set of equations:

(1) $V(k, \hat{\pi}_{t+1}, \Psi_1, 0) = 0$

(2) $V(k, \hat{\pi}_{t+1}, \Psi_1, \Psi_2) =$

$$\sum_{\ell=1}^{n_k} U(\Psi_1, \Psi_2, E_{S,Q}(\hat{\pi}_{t+1}, r_k^{\ell}, \Psi_1)) \sum_{i=1}^{s} \hat{\pi}_{t+1} F_k(r_k^{\ell}, \sigma_i) - C_k \quad \text{if } \Psi_2 > 0$$

(3) $U(\Psi_1, \Psi_2, E_{S,Q}(\hat{\pi}_{t+1}, r_k^{\ell}, \Psi_1)) =$

$$\max \left[\max_{k'} V(k', E_{S,Q}(\hat{\pi}_{t+1}, r_k^{\ell}, \Psi_1), \Psi_1, \Psi_2 - 1), \right.$$

$$\left. \max_{j} Y_j E_{S,Q}(\hat{\pi}_{t+1}, r_k^{\ell}, \Psi_1) \right]$$

In Equations 2 and 3, the function U is the value of the simulated decision maker's expectation if his opinion is $E_{S,Q}(\hat{\pi}_{t+1}, r_k^{\ell}, \Psi_1)$ and he considers Ψ_2-1 additional tests. $E_{S,Q}(\hat{\pi}_{t+1}, r_k^{\ell}, \Psi_1)$, henceforth written E, is the opinion that the decision maker would hold at time t+2 if $\hat{\alpha}_{t+1}$ were the purchase of test q_k, and this test gave result r_k^{ℓ}. $U(\Psi_1, \Psi_2, E)$ is either the value of the best treatment given opinion E, $\max_j Y_j E$, or the value of the best test at time t+2 given opinion E and myopia Ψ_2-1 (since the test would be the second of the Ψ_2 tests in the longest possible plan), $\max_{k'} V(k', E, \Psi_1, \Psi_2-1)$.

The subjective probability that the purchase of test q_k at time t+1 will in fact yield the result r_k^{ℓ} is found by multiplying the conditional probability of result r_k^{ℓ} to test q_k qiven that the true state of the world is σ_i, $F_k(r_k^{\ell}, \sigma_i)$, by the subjective probability (at time t+1) of state σ_i, $\hat{\pi}_{t+1}^{i}$, for each possible state σ_i, and summing over i.

The total expected value of purchasing test q_k at time t+1 is found in equation 2 by multiplying the value of the decision maker's expectation at time t+2 if test q_k at time t+1 gives result r_k^{ℓ}, $U(\Psi_1, \Psi_2, E_{S,Q}(\hat{\pi}_{t+1}, r_k^{\ell}, \Psi_1))$, by the subjective probability of result r_k^{ℓ} for each ℓ, summing over ℓ, and subtracting C_k, the cost of the test itself.

Note the recursive definition of $V(k, \hat{\pi}_{t+1}, \Psi_1, \Psi_2)$ in terms of $U(\Psi_1, \Psi_2, E_{S,Q}(\hat{\pi}_{t+1}, r_k^{\ell}, \Psi_1))$ and of the latter in terms of $V(k', E_{S,Q}(\hat{\pi}_{t+1}, r_k^{\ell}, \Psi_1), \Psi_1, \Psi_2-1)$. The value of a test at time t+1 depends on the values and probabilities of the possible states of knowledge at time t+2 that could result from the test, and the values of these possiblle states of knowledge depend on the values of the various possible actions at time t+2, evaluated at a reduced myopia. The recursion is terminated after Ψ_2 levels by invoking Equation 1.

(D) Determine the best test q_{k*} such that

$$V(k^*,\hat{\pi}_{t+1},\Psi_1,\Psi_2) \geq V(k,\hat{\pi}_{t+1},\Psi_1,\Psi_2) \quad \text{for all } k \in \{1, 2, \ldots, \emptyset\}.$$

(E) Determine $\hat{\alpha}_{t+1}$ as follows:

$$\hat{\alpha}_{t+1} = \text{Select final choice } x_{j*} \text{ and put } \hat{m} = t+1 \quad \text{if } Y_{j*}\hat{\pi}_{t+1} > V(k^*,\pi_{t+1},\Psi_1,\Psi_2),$$

$$\hat{\alpha}_{t+1} = \text{Purchase test } q_{k*} \quad \text{if } Y_{j*}\hat{\pi}_{t+1} \leq V(k^*,\hat{\pi}_{t+1},\Psi_1,\Psi_2).$$

When $\Psi_1 = 1$ and $\Psi_2 = \infty$, this model is the standard decision analysis algorithm (Raiffa, 1968). When $\Psi_1 = 1$ and $\Psi_2 = 1$, the model is the same as that reported on by Gorry et al (1973).

3. PILOT STUDY

The primary goal of the pilot research study of the Myopic Conservative Bayesian Decision Maker model is to demonstrate and gain experience with the use of the paradigm presented above in evaluating theoretical models of human decision making. A secondary goal is to evaluate the Myopic Conservative Bayesian Decision Maker model itself as an explanatory theoretical model for observed patterns of behavior. The study was carried out in three phases: (1) Creating the task environment, (2) Mapping the parameter space, and (3) Observing human decision makers.

3.1 CREATING THE TASK ENVIRONMENT

The following one-player simulation game was used as the basic framework for the task environment:

The decision maker takes the role of a veterinarian in an agricultural cooperative which is raising a large number of a new variety of poultry. The co-op is currently experiencing losses due to an unknown disease among the birds; the identity of the disease has been narrowed down to four possibilities, each of which is considered equally likely.

Because the poultry is a new variety, the decision maker has no relevant personal experience; however, he has access to a number of diagnostic tests with published findings (conditional probabilities) resulting from the application of the respective tests to birds of the variety in question known to have each of the four considered diseases. Each test has a cost which is several percent of the maximum expected profit from the venture. (All the inexpensive tests may be considered to have already been used in narrowing the search to four diseases). Each disease has a specific treatment which results in a complete cure and a large profit when applied correctly, but to a total loss when applied in the presence of a different disease. In addition, there is a fifth, "broad spectrum" treatment which results in partial control of any of the diseases, and thus in an intermediate profit.

The decision maker will receive a bonus proportional to the profit made on the venture, net of the cost of any diagnostic tests used.

The principal equipment of the simulation game consists of (1) introductory material giving a background explanation plus the structure of costs and payoffs, and (2) a set of cards, each of which summarizes a particular test (in terms of its cost, possible outcomes, and conditional probabilities) or treatment (in terms of its payoff given each possible disease state). These materials, with the actual values used in the experiment, are exhibited in Appendix 3.

In terms of the formal model of a decision-making task, S is the set of four diseases, X is the set of five treatments, Y is the profit arising from each disease -- treatment pair, and Q is the set of five cards giving information about each test (q_k), with its conditional probabilities $(F_k(r_k^\ell,\sigma_i))$.

This framework was selected because it is parallel to the important problem of cost-effective use of diagnostic procedures in human medicine, and yet is simple enough to be tractable in a small-scale pilot study of human decision processes. In particular, the agricultural setting was chosen in order to substitute simple monetary costs and benefits for the highly ambiguous value tradeoffs necessary in the treatment of human illness.

Once this general framework was established, the next step was to specify the values of the probabilities, costs, and payoffs involved. The goal of this phase of the study was to create a decision-making task that was sensitive to variations in the parameters of conservatism and myopia, and yet was simple enough that the human decision makers could safely be assumed to comprehend it. In addition, the task could not be so over-controlled that the only behavior possible was behavior predicted by the model, since such a situation would prevent any real test of the model's accuracy.

The last of these concerns was addressed first by creating a first approximation to the task with a very high degree of symmetry, so that only a few numbers must be learned by the subjects. The Myopic Conservative Bayesian

Decision Maker was then implemented as a computer simulation program written in PASCAL on MSU's CDC 6500 computer (Appendix 2), and simulation experiments were performed using the simple task and various values of conservatism and myopia. The variation in the model's behavior was found to be too small, so changes were made to the costs, probabilities and payoffs and the simulation was repeated, until a task environment that appeared to meet all these requirements was found. (This iterative simulation method of creating a task environment was chosen in preference to an analytic design of an optimal environment for the exercise of the Myopic Conservatibe Bayesian Decision Maker's equations in part to preserve maximum generality of the paradigm across models, since the model is treated as a "black box"; the process whereby the task environment becomes tuned to the model depends only on the model's behavior, not on its structure.) The final values of the costs, payoffs and probabilities are shown in Table 1.

Once the numeric values were established, the next step was to express them in a form suitable for presentation to human subjects. The manner of presentation which was developed is presented in Appendix 3. The two most important difficulties in creating a task environment for human subjects, comprehension and motivation, will be discussed below.

TABLE 1: COSTS, PAYOFFS AND PROBABILITIES
(TASK ENVIRONMENT FOR THE PILOT STUDY)

TEST A: COST = 100

	IF ACTUAL DISEASE IS:			
PROBABILITY OF::	VIRUS A	VIRUS B	BACTERIUM C	BACTERIUM D
POSITIVE RESULT:	.90	.20	.20	.20
NEGATIVE RESULT:	.10	.80	.80	.80

TEST B: COST = 100

	IF ACTUAL DISEASE IS:			
PROBABILITY OF::	VIRUS A	VIRUS B	BACTERIUM C	BACTERIUM D
POSITIVE RESULT:	.20	.90	.20	.20
NEGATIVE RESULT:	.80	.10	.80	.80

TEST C: COST = 100

	IF ACTUAL DISEASE IS:			
PROBABILITY OF::	VIRUS A	VIRUS B	BACTERIUM C	BACTERIUM D
POSITIVE RESULT:	.20	.20	.90	.20
NEGATIVE RESULT:	.80	.80	.10	.80

TEST D: COST = 100

	IF ACTUAL DISEASE IS:			
PROBABILITY OF::	VIRUS A	VIRUS B	BACTERIUM C	BACTERIUM D
POSITIVE RESULT:	.20	.20	.20	.90
NEGATIVE RESULT:	.80	.80	.80	.10

TEST V: COST = 100

	IF ACTUAL DISEASE IS:			
PROBABILITY OF::	VIRUS A	VIRUS B	BACTERIUM C	BACTERIUM D
POSITIVE RESULT:	.80	.80	.20	.20
NEGATIVE RESULT:	.20	.20	.80	.80

PAYOFFS FOR DISEASE/TREATMENT PAIRS:

	IF ACTUAL DISEASE IS:			
	VIRUS A	VIRUS B	BACTERIUM C	BACTERIUM D
TREATMENT A	1000	0	0	0
TREATMENT B	0	1000	0	0
TREATMENT C	0	0	1000	0
TREATMENT D	0	0	0	1000
TREATMENT X	300	300	300	300

Comprehension: The substantive focus of the cartographic paradigm is the study of how different people approach the resolution of the same decision-making task. The task has to be complex enough that the optimal solution path is not obvious, or all the subjects would respond in the same way; the demand effect of the task would have overcome their individual differences. On the other hand, if the task itself is so complex that the subjects do not fully understand it, they will only be able to respond to their own partial models of the task, and there is no reason to expect these partial models to be the same from subject to subject; in this sense, we are no longer observing different people's response to the same task. Thus, the task must be understood fully and equally by all subjects, while the solution must be so complex that they vary in their approaches to it. These concerns were in part addressed by using a highly symmetrical set of numbers and experimenting with the simulation program to make sure behavior is sensitive to differences in conservatism and myopia. The subject instructions and information cards in Appendix 3 are intended as a clear and redundant presentation of the task; in particular, the information cards provide a continuing reminder to reduce the danger of forgetting and misremembering.

Motivation: Just as classical decision analysis computes the objective value of information in terms of costs, payoffs, and Bayesian probabilities, the Myopic Conservative Bayesian Decision Maker judges the subjective value of information in terms of utilities and subjective probabilities. Thus, the accuracy of an experimental study of the model depends heavily on the accuracy with which the utilities of payoffs and disutilities of costs are determined. In the preliminary experiments, a linear approximation to the utility of money is used. Since most theoretical and empirical utility curves are most nonlinear near zero (see Baumol, 1977), the variable payoff to the subject is restricted to the range from -$1.00 to +$1.00, and combined with a fixed payment of $5.00. In other words, the subject's total payoff is in the range $4.00 to $6.00, in order to reduce the distortion introduced by the linear approximation to the utility of money. In Experiment 4 and subsequent research using the paradigms, more sophisticated models of utility, such as that of VonNeumann and Morgenstern, will be substituted to increase the precision of the measurements.

3.2 MAPPING THE PARAMETER SPACE

In order to address the observer's goal of developing a system for classifying overt patterns of behavior in terms of the parameters of the theoretical model, the next step in the pilot study was the preparation of a map of the parameter space defined by conservatism and myopia. The map was prepared by running the simulation program with the task environment selected for experimentation (including the "rigged" algorithm for generating test results) and roughly fifty different conservatism/myopia pairs. These pairs were restricted to myopia less than five, since the amount of calculation required with myopia of four was so great that a longer planning horizon, at the full breadth of search assumed by the model, can be considered highly unlikely in unassisted human beings at the levels of motivation used. Conservatism was also initially restricted to be no more than 1.25, as no conservatism values approaching this limit have been reported in the literature. Within these bounds, mapping was done separately for each of the four levels of myopia considered. For a given value of myopia, widely separated conservatism values were first evaluated; then, for pairs of conservatism values leading to different sequences of overt behavior, intermediate values were examined to find the values at which behavior shifts from one pattern to the other. If a third pattern existed at conservatism values intermediate between the first two, the interpolation procedure automatically uncovered this fact, then found transition points delimiting all such behavior regions.

FIGURE 3: MAP OF PARAMETER SPACE

BEHAVIOR PATTERNS

(Note: on all tests except the second in each sequence, the most likely result given disease A occurs. On the second test in each sequence, the least likely result given disease A occurs.)

O = treatment E

P = test 1, treatment A

Q = test 5, test 1, treatment B

R = test 5, test 1, test 2, test 3, test 4, test 2, test 1, treatment A

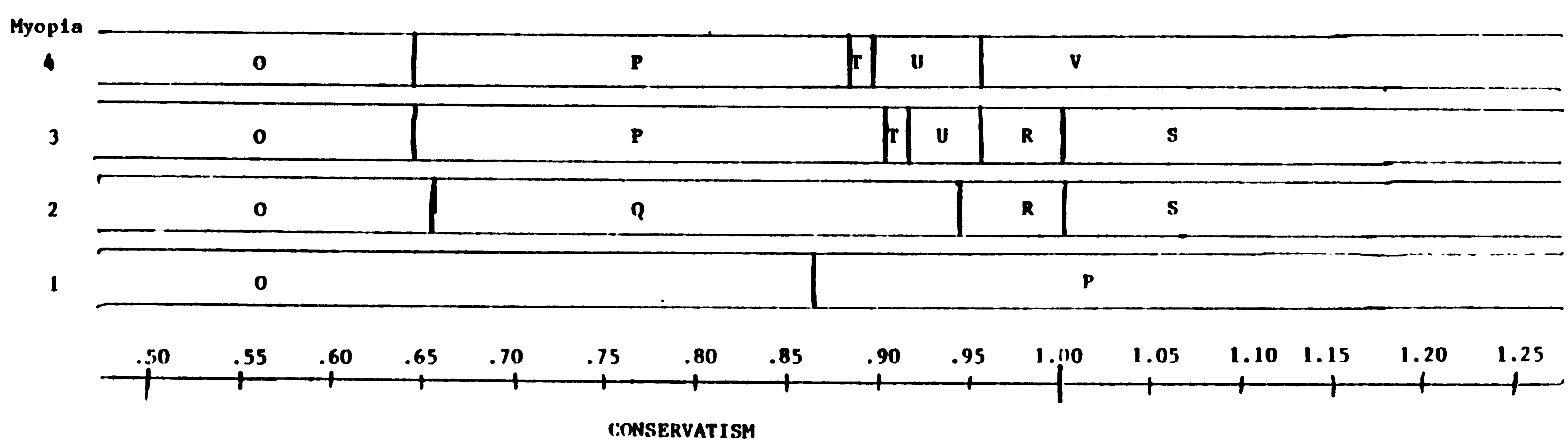
S = test 5, test 1, test 2, test 3, test 4, test 1, treatment A

T = test 1, test 1, test 2, test 3, test 4, treatment A

U = test 1, test 1, test 2, test 3, test 4, test 1, treatment A

V = test 1, test 1, test 3, test 2, test 4, test 1, treatment A

MAP OF PARAMETER SPACE



The resulting map of parameter space, shown in Figure 3, shows that, as conservatism and myopia vary freely within the range $\Psi_1 < 1.25$, $\Psi_2 < 5$, the model produces exactly eight different behavior patterns in response to the given task environment. Any values of the parameters within a particular labeled region produce overt behavior that is indistinguishable, and is given by the behavior list corresponding to the label. At the boundaries of this region, however, a small change in parameter values produces a sudden shift in behavior.

In addition to providing a typology of individual differences based on behavior sequences, the results of the simulation runs are also of interest as a formal study of the mathematical consequences of the assumptions of the Myopic Conservative Bayesian Decision Maker model given various parameter settings. In the remainder of this section, the eight behavior patterns found during the simulation will be examined in turn, to gain insight into the behavior of the Myopic Conservative Bayesian Decision Maker viewed as a formal system.

Pattern 0: Treatment E. The explanation of this behavior pattern is very simple: The decision maker finds none of the tests to be worth their price, and immediately applies the broad-spectrum treatment. This low opinion of the value of testing is associated with low values of conservatism, which mean that information has little effect on opinion. To a lesser extent, myopia determines how high conservatism must be before any information is purchased, since a high myopia means that information which is not worth much by itself may gain value as part of a long-range information-seeking plan.

Pattern P: Test 1, treatment A. The region in parameter space underlying this response to the given decision-making task is in two parts. When myopia is equal to 1, the model evaluates each test as if a final decision had to be made on the basis of just the result of that test plus the current opinion. If the potential impact of information is great enough (measured by conservatism greater than about 0.86), then one of the specific tests (called test 1 without loss of generality; the model is biased towards low-indexed tests when subjective expected utilities are equal) is selected. This test gives the result that favors state A. (See the discussion of test results and selection of "true" state in Appendix 3). State A now has an increased subjective probability while states B, C, and D have equal, decreased subjective probabilities. If only one more test can be thought of before the final decision, a different test is valueless because it cannot provide sufficient reason to choose a different treatment, and a repetition of the same test, while it could conceivably provide sufficient grounds to switch to the broad-spectrum treatment, is not likely enough to do so to offset its cost. Thus, no more information is purchased, and treatment A is selected.

When myopia equals 2, the strategy of test 5 on the first test and test 1 on the second is dominant, and Pattern P does not appear for any value of conservatism. However, at myopia of 3 or 4 there are values of conservatism for which longer combinations of tests have higher subjective expected utility. For myopia of 3 and conservatism between .64 and .91, or for myopia of 4 and conservatism between .64 and .89, longer and more complicated chains of logic come to essentially the same conclusions as discussed above, and pattern P is emitted.

Pattern Q: Test 5, test 1, treatment B. The first thing one might notice about this pattern is that it is the only one that leads to an incorrect application of a specific treatment and thus to zero gross profit. Although this is primarily an artifact of the particular way the test results were set up unknown to the model or to the subjects (results favoring the actual state were given on all tests but the second in a sequence), most published methods of analyzing decision-making processes would begin by putting pattern Q in a class by itself as a "failure" and lump all the others as "successes" prior to analyzing the different kinds of successful pattern. Using the parameter-space method, however, it can be clearly seen (see Figure 3) that Pattern Q lies next to Patterns O and R in the Conservatism dimension and Patterns O, P, T, and U in the Myopia Dimension. The region of parameter space where myopia = 2 and conservatism is between .94 and .95 is particularly interesting, since a decision maker so characterized is nearly indifferent between the "successful" pattern R and the "unsuccessful" pattern Q, because the subjective expected utility of the third test in the sequence is nearly equal to its cost.

Pattern R: Test 5, test 1, test 2, test 3, test 4, test 2, test 1, treatment A; and Pattern S: Test 5, test 1, test 2, test 3, test 4, test 1, treatment A. These two patterns are discussed together because the difference between them is the result of an unanticipated consequence of the conservative Bayesian information-processing algorithm which, at least in the context of this task environment, mimics the important phenomenon of primacy/recency effects. The two patterns make identical choices for the first five tests, and get results favoring disease A or B on the first (test 5), followed by negative results on each of the four specific tests. At this point, if conservatism = 1 (as in Bayes' law), the subjective probabilities of diseases A and B are equal (and greater than those of diseases C or D), so the decision maker is indifferent between test 1 and test 2.

However, when conservatism is less than one, the negative result on test 1 on the second iteration has a greater effect on opinion than the negative result on test 2 on the third iteration; thus, the subjective probability of disease B is higher, giving test 2 a higher subjective probability of a positive result on iteration 6. Since a positive result would result in sufficient grounds to come to a final treatment decision, this means a higher subjective expected utility for test 2 on iteration 6 and so test 2 is selected. (Since in fact a negative result ensues, test 1 is selected on iteration 7, a positive result is observed, and treatment A is prescribed on iteration 8.) On the other hand, if conservatism is greater than 1, the more recent negative result on test 2 (iteration 3) has a greater effect on subjective probability than the less recent negative result on test 1 (iteration 2), so disease A is more subjectively likely and test 1 is selected on iteration G. These two patterns also illustrate another feature of conservatism, the non-monotonicity of the relation between the impact of information and the amount of information purchased -- pattern R, with conservatism less than 1, contains seven information purchases, while pattern S, with conservatism greater than 1, contains only six.

Pattern T: Test 1, test 1, test 2, test 3, test 4, treatment A. The most striking thing about pattern T is the small size of the region of parameter space for which it occurs. The reason for this is that, when conservatism is equal to 1 (Bayes' law), the negative results on each of the four specific tests on iterations 2 through 5 bring the Bayesian subjective probabilities exactly back to the values they had after the first test. Since this set of subjective probabilities resulted in the highest subjective expected value being associated with purchasing test 1 (to try to make more certain of the indicated diagnosis of disease A) the first time, it does so again, as in Pattern V when conservatism = 1. On the other hand, when conservatism is low enough, Pattern P occurs because, given one positive result, further information will not have enough impact for its subjective expected utility to exceed its cost, and a final choice is made at iteration 2. However, the effects of primacy combined with conservatism at the critical values of 0.91 when myopia = 3 and 0.89 when myopia = 4 have made the situation after 5 tests just sufficiently different (in violation of Bayes' law) from the situation after 1 test that no sixth test is purchased and a final prescription of treatment A is made instead.

Pattern U: Test 1, test 1, test 2, test 3, test 4, test 1, treatment A. This pattern differs from pattern T only in the fact that it occupies a range of conservatism closer to the Bayesian value of 1.0, which means that the negative results on iterations 2 through 5 come close to canceling each other out. This returns the decision maker almost to the state of belief he was in after the first test, so he again purchases a test 1. (Note that at conservatism values even closer to 1.0, when myopia = 3 the increasing value of information in general causes the decision maker to prefer patterns R or S which begin with the broader but weaker test 5. When myopia = 4, pattern V is chosen for higher values of conservatism, as discussed below).

Pattern V: Test 1, test 1, test 3, test 2, test 4, test 1, treatment A. This pattern differs from Pattern U only in that test 3 and test 2 are reversed in sequence. The reason for this is related to the transition from Pattern U to Pattern R when myopia = 3, in that both depend on an increase in the value of test 5. In Pattern V, however, the change takes place in a more subtle way due to the greater ability to think ahead when myopia = 4. Because of the way test results came out in this experiment, test 5 never appears, but it enters into the simulated decision-maker's planning as follows: After test 1 on iteration 1 returned a positive result and test 1 on iteration 2 returned a negative result, test 3 is preferred over test 2 because, if test 3 gave positive results on iteration 3, then test 3 would be repeated on iteration 4. If test 3 on iteration 4 then gave neagtive results, the first four iterations would, on the whole, count against states A and C. Therefore, Test 5 would be purchased on iteration 5; it would then favor state B if the result were positive, and state D if it were negative. The specific test for the favored state would then be purchased on iteration 6; if it were positive, this would be sufficient grounds to treat for that disease. While this scenario is unlikely, test 3 is just as good as test 2 on all other possibilities and superior on this one, since test 2 does not allow taking advantage of test 5. Test 4, of course, is just as good as test 3, so the slight bias for the lower-indexed test prevails.

3.3 OBSERVING HUMAN DECISION MAKERS

Four experiments with human subjects were performed to demonstrate the use of the research paradigm.

Experiment 1 was a preliminary test of the ability of the Myopic Conservative Bayesian Decision Maker to model the behavior of a very wide range of individuals (readers of Michigan State University's student newspaper) on first exposure to the decision-making task presented above (including one erroneous datum, on iteration 2), and by implication problems of similar complexity. As the results described below will show, the experiment provided strong evidence that the model is not adequate for this task.

Experiments 2 and 3 were then performed varying the subject population, the task, and the model itself; the goal of these experiments was to analyze the ways in which simulated and actual human behavior differred in order to provide a foundation for the next generation of theoretical models and experimental task environments to be exercised and combined according to the paradigm presented in this dissertation.

Experiment 4 was then designed using the results of the preliminary experiments. The two principal goals of this experiment were to provide a test of the ability of the original Myopic Conservative Bayesian Decisiom Maker to model the variations of behavior within a more homogeneous group of subjects who have had a chance to practice with the original decision-making task, and to compare the bounds set upon conservatism by the model's interpretation of overt economic behavior with the point estimate of conservatism provided by Edwards' introspective method.

In this report, results will be presented in a "canonical form" due to the purely arbitrary nature of some of the subjects' decisions. "Test 5" always means Test V, "Test 1" is the first specific-virus test (Test A or Test B) purchased by the subject while "Test 2" is the other specific-virus test, and "Test 3" is the first specific-bacterium test (Test C or Test D) purchased by the subject while "Test 4" is the remaining specific-bacterium test. Treatments and diseases are recoded correspondingly; the treatment specific to the first virus tested for is referred to as "Treatment A", and similarly for the remaining treatments.

TABLE 2

RESULTS OF EXPERIMENT 1

Subject

S1: Treatment E

S2: Test 5, Test 1, Test 2, Test 1, Test 1, Treatment B

S3: Test 5, Test 5, Test 5, Test 1, Test 2, Test 5, Test 1, Test 2, Treatment A

S4: Test 5, Test 1, Test 2, Test 3, Test 4, Treatment E

S5: Test 5, Test 1, Test 5, Test 1, Test 1, Test 2, Test 2, Test 1, Treatment A

S6: Test 5, Test 1, Test 2, Test 3, Test 4, Treatment E

S7: Test 5, Test 1, Test 5, Test 1, Test 1, Treatment A

S8: Test 5, Test 1, Test 2, Test 3, Test 4, Treatment E

S9: Test 5, Treatment A

S10: Test 5, Treatment A

3.3.1 Preliminary Experiments

Experiment 1 consisted of presenting ten subjects individually with the task environment described above. Subjects were recruited by a classified advertisment in the Michigan State News, which read "Earn $4.00 or more in decision-making experiment" followed by instructions for contacting the experimentor. No selection beyond schedule compatibility was exercised; most subjects were undergraduates, but graduate students and university clerical staff were also included. Typically, subjects had no coursework in statistics, although one had three terms and one was currently taking a course in management which included a treatment of Bayesian statistics.

For the nine subjects in Experiment 1 who chose to purchase diagnostic tests, results were provided according to the scheme presented in Appendix 3; in particular, it should be noted that the second test a subject purchased produced the less likely outcome given the actual disease, while all other tests produced the more likely outcome. As a result, seven subjects received data which obviously contained at least one inconsistent element. (In addition to the subject who purchased no tests, two other subjects each purchased only one test.)

The results of Experiment 1 are exhibited in Table 2. Of the ten subjects, only subject S1 showed a behavior pattern that matched any of the eight patterns of which the

Myopic Conservative Bayesian Decision Maker is capable. (Subject S1 matched Pattern 0). Even considering the small sample size, these results are a strong indication that the variety of ways in which the behavior of naive, unselected persons encountering this problem for the first time differs from the behavior prescribed by the decision analysis algorithm cannot be accounted for by any combination of conservatism and myopia.

Examination of the results of Experiment 1 led to the conjecture that much of the behavior of the naive subjects was essentially random, due to a lack of experience with and knowledge of tasks of this nature. Experiment 2 was performed to explore the ability of the Myopic Conservative Bayesian Decision Maker to model the behavior of more sophisticated persons. The five subjects in Experiment 2 were four graduate students in engineering and a laboratory technician employed by the Michigan Department of Agriculture. The task which was presented to these individuals was the same as the task in Experiment 1, except that the fixed payment was reduced from $5.00 to $1.00 by eliminating the $4.00 "sure-thing" payment.

TABLE 3

RESULTS OF EXPERIMENT 2

Subject

X1: Test 5, Test 1, Treatment B

X2: Test 5, Test 1, Test 2, Test 5, Test 2, Test 1, Treatment A

X3: Test 5, Test 1, Test 2, Test 3, Test 4, Treatment E

X4: Test 5, Test 1, Test 2, Test 3, Test 4, Test 1, Test 1, Treatment A

X5: Test 5, Test 1, Test 2, Test 5, Test 1, Treatment A

The results of Experiment 2 are shown in Table 3; as in Experiment 1, only one subject matched any of the behavior patterns generated by the model. (Subject X1 matched Pattern Q). Thus, it is clear that a significant range of behavior remains unaccounted for even when the population of decision makers is restricted to persons with training and experience in complex, numerically-based decision making.

Of the fifteen subjects in Experiments 1 and 2, neither of the two whose behavior matched that of the Myopic Conservative Bayesian Decision Maker purchased enough tests to perceive conflicting data; subject S1 in Experiment 1 purchased no tests and chose the broad-spectrum treatment, while subject X1 in Experiment 2 purchased Test 5, got a positive result, purchased Test 1, got a negative result, and prescribed Treatment B, which was the only specific treatment not contraindicated by any data so far.

The other four subjects in Experiment 2, plus four of the subjects in Experiment 1 (S2, S4, S6 and S8), reached states of knowledge equivalent to that reached by subject X1. However, these subjects (like patterns R and S generated by the model) proceeded to test for the indicated disease (Test 2) instead of treating immediately. When these eight human decision makers received a negative result on Test 2, which contradicted the most likely expectation given the data thus far, their subsequent behavior diverged from that of the model.

These findings on the first two experiments led to the conjecture that, while the Myopic Conservative Bayesian Decision Maker was inadequate to predict human behavior given data tending to disconfirm subjects' likely initial hypotheses, it might still be useful as a partial model of behavior given consistent data, supplemented by other parameters or submodels for the more complex case of inconsistent data.

To examine this conjecture, a third experiment was performed. In Experiment 3, the instructions and payments to the subject were the same as in Experiment 2; however, the method for determining test results and the "actual" disease was different. As in the first two experiments, the first test always yielded a positive result, and the actual disease was constrained to be consistent with this result. On the second and all subsequent tests, the result was constrained to be consistent with all results given so far. If the first test were a specific test, this constraint would be sufficient to determine all subsequent results. If (as turned out to be the case in every instance) the first test were Test 5, then the first specific virus (A or B) tested for would yield a negative result, the disease would be the other virus, and all subsequent test results would be determined by the requirement of consistency. In this way, those subjects (a majority) whose first two actions were Test V and a specific-virus test (Test A or Test B) would get the same test results for these initial tests whichever experimental group they belonged to, but different test results thereafter.

The subjects for Experiment 3 were five students drawn from a graduate course in systems science, "Advanced Systems Methodology and Simulation."

The results of Experiment 3 are displayed in Table 4. Subject Z4's behavior matches that of the Myopic Conservative Bayesian Decision Maker with myopia=2 and conservatism between .65 and .95 (Pattern Q); the test results received are the same as would be received in Experiments 1 or 2. The other four subjects in Experiment 3 received sequences of test results different from those used in preparing the map of parameter space used in the first two experiments. However, since they all began with Test V, only the region of parameter space which produces behavior beginning with Test 5 needed to be re-mapped. Furthermore, the region with myopia = 2 and conservatism between .65 and .95 produces the same behavior pattern in either condition; although the absence of error in Experiment 3 changes the final payoff, the (real or simulated) decision maker does not know this until the behavior pattern is complete. Thus, only the region of parameter space underlying patterns R and S in the original experiment needed to be re-run for Experiment 3.

TABLE 4

RESULTS OF EXPERIMENT 3

Subject

Z1: Test 5, Test 1, Test 2, Treatment B

Z2: Test 5, Test 1, Test 2, Test 2, Treatment B

Z3: Test 5, Test 5, Test 1, Test 1, Treatment B

Z4: Test 5, Test 1, Treatment B

Z5: Test 5, Test 1, Test 2, Test 2 Treatment B

(Note that in Experiment 3 the actual disease is the second one tested for, and all tests produce the most likely result given this disease).

Computer simulation runs indicated that, in this entire region (myopia = 2 and conservatism > .94 or myopis = 3 and conservatism > .95), the only behavior pattern that emerges under the conditions of Experiment 3 is: Test 5, Test 1, Test 2, Treatment B. This pattern matches the behavior of subject Z1.

Two of the subjects in Experiment 3 whose behavior was not matched by the Myopic Conservative Bayesian Decision Maker, subjects Z2 and Z5, used a series of tests which differs from the pattern produced by subject Z1 (and by the model when myopia = 2 and conservatism > .94 or myopia = 3 and conservatism >.95) only in the purchase by subjects Z2 and Z5 of an extra repetition of Test 2 before selecting Treatment B. While no variation of conservatism and myopia alone could account for this behavior, a modified version of the model was able to match the pattern. In the modified version, a third parameter was introduced by varying the utility of a correct diagnosis. For simplicity, conservatism was held constant at 1.0. When myopia equaled three and the utility of a correct diagnosis and specific treatment was raised from its original value of 1,000 to 6,000, the revised model produced behavior matching subjects Z2 and Z5; this pattern was stable for utilities from 6,000 at least to 10,000 for this value of conservatism, while utilities from 1,000 to 5,000 led to behavior matching subject Z1 and the original model (utility = 1,000, proportional to cash profit).

The behavior of the remaining subject, Z3, cannot be accounted for by either the original model or the extended version; while the second purchase of Test 5 might or might not be indicated for some 3-tuple of conservatism, myopia and utility, the second purchase of Test 1 on iteration 4 occurs at a time when the evidence indicates that, for any positive value of conservatism, Virus B is more likely than Virus A. As a result, no matter what positive values are chosen for conservatism, myopia and utility, the Myopic Conservative Bayesian Decision Maker could never prefer Test 1 to Test 2 on the fourth iteration of the sequence of tests purchased by subject Z3. In other words, whatever heuristics this subject used (such as "do all tests twice in succession to improve reliability"), these heuristics are located outside the realm of Bayesian strategy regardless of any modifications of conservatism, myopia or utility.

3.3.2 Refinements to the Task Environment

The results of the three preliminary experiments raised several questions about the conduct of the experiment and about the Myopic Conservative Bayesian Decision Maker. The prevalence of behavior patterns difficult to justify under any principle of quasi-optimality implies a lack of comprehension among some of the subjects. The analysis of different levels of utility carried out using the data from Experiment 3 inidcates that motivation may have been inadequately controlled. On the other hand, the deviation between the behavior of the subjects and that of the model may indicate that the conservative Bayesian information processing pattern, found by Edwards and his co-workers when subjects were given costless, experimenter-selected data, does not apply to the present situation where subjects select and pay for information.

Experiment 4 was designed to shed further light on each of these three issues. The subjects were twenty students from an undergraduate course in statistical methods in psychology; the experiment was conducted near the end of the term, so the subjects had recent and continuing exposure to elementary statistical concepts. To assure comprehension, the experimental session consisted of ten repetitions of a replication of Fried and Peterson's (1969) experimental task which is similar in concept to but simpler than the present one, followed by ten repetitions of the game used in Experiments 1, 2, and 3. The first five of these repetitions of the experimental game were "just for practice," to allow the subject to gain familiarity with the consequences of various ways of responding without affecting his outcome, while the last five were played for lottery chances (see below). All ten repetitions of the Fried and Peterson replication, all five of the "practice games," and the first three out of the five games for lottery chances were played honestly according to the stated probabilities, but the last two games were rigged, unknown to the subject, so that test and treatment outcomes in the next-to-last game matched those in Experiments 1 and 2, and outcomes in the last game matched those in Experiment 3. (A programming error resulted in one subject, #13, receiving erroneous results for the last two tests on the last game; however, this subject had already departed from the model's behvaior before the first erroneous result.) Data were analyzed for the last two games only, to maximize subjects' comprehension of and familiarity with the experimental task environment.

In order to obtain better control of motivation, the payment schedule was also changed. Subjects received a flat cash payment of $4.00 for participation in the experiment, plus chances in two lotteries, each with a prize of $20.00. Each subject's chances in Lottery I were determined by his performance on the ten repetitions of the Fried and Peterson replication, and his chances in Lottery II were determined by his performance on the last five repetitions of the diagnosis and treatment game. The reason for the lotteries is that, while the utility of money varies in a nonlinear manner which differs from individual to individual and from time to time, Von Neumann and Morgenstern (1947; see also Raiffa, 1968) have shown that, given certain reasonable assumptions, the utility to a given individual of a chance in a given lottery, measured relative to the utilities of winning and of losing the lottery, is a linear function of the probability of winning. Thus, in the experiment, a subject received one chance in Lottery II for each $100 of net profit in the last five games, which means that, under the Von Neumann-Morgenstern assumtions, the subject's utility for "money" in the simulation game (net profit of the imaginary turkey farm) should be linear.

Questions concerning the applicability of the conservative Bayesian model itself to an information selection and purchase situation are indirectly addressed by the abovementioned improvements in experimental control. A more direct comparison between information processing in the present task environment and in the literature of conservatism (Edwards 1968, 1972; Schum and Martin, 1968; Slovic and Lichtenstein, 1971) was made by asking the subject to mark a line calibrated from "completely impossible" to "absolutely certain" to show his subjective probability of the disease he considers most likely. This was done once per game, after a treatment had been selected but before treatment results were given to the subject. In this manner, a direct measure of introspected subjective probability was obtained by a replication of Edwards' general methodology; the principal difference between this subjecive probability and those obtained by Edwards, Schum and Martin and others is that, in the present experiment, the subjects themselves have selected and paid for the data, wheras in the Edwards-type experiment the experimentor determined the amount and type of the data to be given to the subject, whose only task was to react to these data.

The experiment was conducted using two BASIC programs, one for the Fried and Perteson replication and one for the diagnosis and treatment game. The programs ran on Michigan State University's HP2000 ACCESS system, and the subjects interacted with the programs via a TeleType model 33 terminal.

TABLE 5
Comparison Between Behavior Patterns of Subjects in Experiment 4 and the Myopic Conservative Bayesian Decision Maker

		Behavior on Game 10					
		O	P	Q	W	Nomatch	Total
	O:	1	1	0	0	0	2
Behavior	P:	0	0	0	0	2	2
On Game	Q:	0	0	2	3	1	6
9	W:	0	0	0	0	0	0
	Nomatch:	0	1	0	2	4	7
	Total:	1	2	2	5	7	17

(Model behavior patterns R, S, T, U and V did not occur among the human subjects.)

TABLE 6
Matches, Near-Matches and Non-Matches of Subjects With the Model Given Consistent anf Inconsistent Data

	Consistent Data	Inconsistent Data	Total
Matches	20	0	20
Near-Matches	N/A	4	4
Non-Matches	7	3	10
Total	27	7	34

3.3.3 Analysis of Economic Behavior

Table 5 shows the degree to which the behavior of 17 subjects in Experiment 4 conformed to that of the Myopic Conservatibe Bayesian Decision Maker. The remaining three subjects were excluded from analysis, one due to noncompliance and the other two due to obvious lack of comprehension of the task.

Each row of Table 5 corresponds to one class of behavior on Game 9, and each column corresponds to one class of behavior on Game 10. Behavior classes O, P, and Q are as defined in Figure 3: O means prescribing the broad-spectrum treatment with no testing; P means purchasing a specific test, receiving a positive result, and prescribing the corresponding specific treatment; and Q means purchansing test V, receiving a positive result, purchasing a specific-virus test A or B, receiving a negative result, and prescribing the specific treatment for the virus not tested for. Because of the way that results are determined, pattern W can only occur on Game 10; it is the behavior sequence found in the third preliminary experiment and in the associated simulation runs with consistent data. The sequence of pattern W is as follows: Purchase test V, receive a positive result, purchase a specific-virus test, receive a negative result, purchase the specific test for the remaining virus, receive a positive result, and prescribe the specific treatment for the latter virus. The remaining class, "Nomatch," refers to any pattern of observed human behavior which no choice of parameters for the Myopic Conservative Bayesian Decision Maker can produce.

One of the most striking things about these results is the fact that no subject who received inconsistent data matched any of the ways the model deals with iunconsistent data (patterns R, S, T, U and V). Table 6 gives the number of matches, near-matches and non-matches between subjects and the model. A near-match is defined as a behavior pattern which diverges from a pattern generated by the model only after inconsistent data is received. For example, subject eleven on Game 9 begins with test 5, gets a positive result, purchases test 1, gets a negative result, and purchases test 2. Up to this point, he matches patterns R and S of the Myopic Conservative Bayesian Decision Maker. However, the result of test 2 under the conditions of Game 9 is negative, which implies to the subject that one of the three tests is an error. The model continues with test 3 in both patterns R and S, but subject eleven departs from the model by purchasing a repetition of test 5.

Note that, while any subject who purchased at least two tests on Game 9 received a test result which was in fact an "error," that is, the result less likely gives the actual disease, only six such subjects purchased a collection of tests whose results were internally inconsistent. The seventh case of inconsistent data occurred on Game 10 as a result of a programming error.

In general terms, the results displayed in Table 6 indicate that, while the assumptions of the Myopic Conservative Bayesian Decision Maker may apply to some human decision makers at the levels of training and motivation used in the experiment when the test results are consistent, no evidence in either Experiment 4 or the three preliminary experiments supports the application of the model in the case of inconsistent test results.

To be of practical or theoretical use, however, a model of this type must not only be able to model an individual's behavior on isolated instances of decision-making; it must also capture the stable underlying parameters of the information-processing strategy the individual is using. In the case of the Myopic Conservative Bayesian Decision Maker, these parameters are assumed to be conservatism and myopia. While examination of the subjects' choices in the games and impressions of their comments imply that major learning is still ocurring after 8 or 9 games and thus fully stabilized strategies are not to be expected, it is nevertheless instructive to examine the degree of consistency from Game 9 to Game 10 shown in the present experiment.

In Table 5 it can be seen that one of the two subjects who used behavior pattern 0 on Game 9 continued using pattern 0 on Game 10, neither of the two subjects who used behavior pattern P on Game 9 continued using pattern P on Game 10, and two of the six persons who used behavior pattern Q on Game 9 continued using pattern Q on Game 10. Examination of the raw data revealed that both of the two subjects who failed to match the model on Game 9 but matched Pattern W on Game 10 were "near matches" on Game 9 who diverged from pattern W only after receiving inconsistent data; thus, these two subjects can be considered "stable" in the same sense as the three subjects whose overt behavior was identical on the two games.

Of the 17 subjects in the anaysis, 7 were not even near-matches on at least one of the two games, and thus the question of stability is undefined. 5 subjects used behavior patterns corresponding to different regions of parameter space on the two games; thus, while their behavior on each individual game fit the model, no single {conservatism, myopia} pair could account for both their behavior on Game 9 and their behavior on Game 10. For the remaining 5 "stable" subjects, however, their behavior on Game 9 (or at least their behavior up to the receipt of inconsistent test results) comes from the same region of the model's parameter space as their behavior on Game 10.

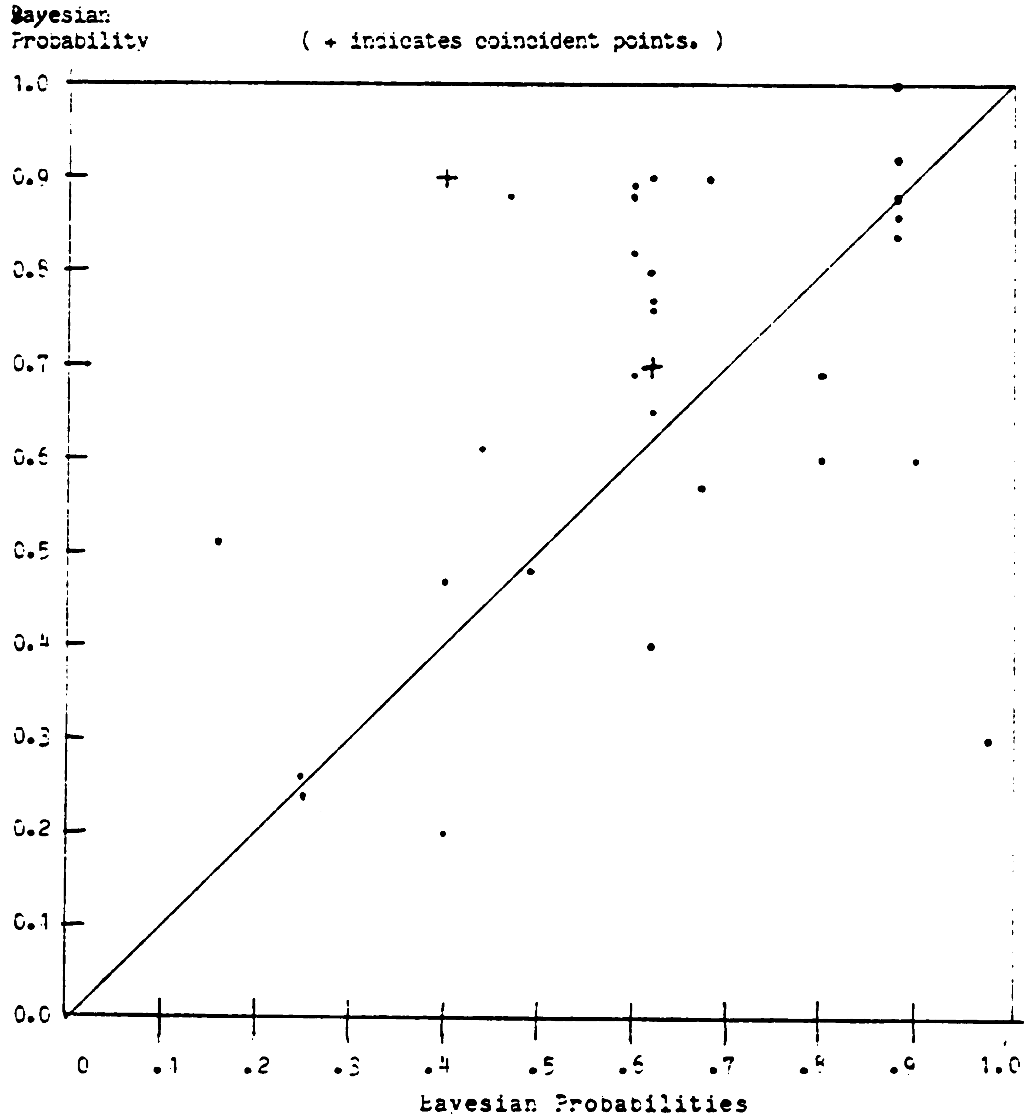


FIGURE 4:
COMPARISON OF SUBJECTIVE AND BAYESIAN PROBABILITIES

3.3.4 Analysis of Subjective Probabilities

The subjective probability of the subjectively most likely disease was elicited in each game by asking the subject to mark a line calibrated from "totally impossible" to "totally certain" after the subject had selected a treatment but before he learned the results of that treatment. In Figure 4, this probability is compared with the objective Bayesian probability of the same disease given the test results received in the course of that game. Plus signs (+) indicate two occurrences of the same subjective probability -- Bayesian probability pair.

The most important thing that is shown by the graph is that less than a third of the points lie below the diagonal line which represents subjective probability equal to Bayesian probability. This contradicts the finding of Schum and Martin (1968) that subjects given probabilistic data regarding the likelihood of several posible states tended to give subjective probabilities for the most likely state that were lower than the Bayesian probability of that state; the present result also contrasts with the findings of the many reports of research in the two-alternative case reviewed by Slovic and Lichtenstein (1971).

The next question to be considered is whether the subjective probability the subject expressed by marking the line accurately measures the actual salience of the corresponding state for decision-making, given the assumptions of the theoretical model being evaluated. This question cannot be investigated for the 14 cases for which the Myopic Conservative Bayesian Decision Maker failed; also, the question is trivial for the three cases in which subjects prescribed the broad-spectrum treatment without purchasing any tests.

The remaining 17 cases are divided among the three behavior patterns P, Q, and W. Each point in parameter space corresponds to a value for the subjective probability after the results of the tests purchased by the model have beeh received and processed one at a time using the model's Conservative Update Module. The range of parameter values for which the pattern occurs defines a range of subjective probabilities the pattern is capable of generating; numeric searching within that range can provide a point estimate for the conservatism which, given the actual test results received, would produce a particular subjective probability within that range.

The analytical methods used by Schum and Martin and by others to estimate conservatism rely on the fact that data are presented all at once as a single compound event, which simplifies the mathematics considerably, compared with the complex order and interaction effects which occur when data are presented one at a time and must be processed before the next data-generating event can occur. In the Myopic Conservative Bayesian Decision Maker, one-datum-at-a-time processing tends to result in subjective probabilities which deviate from the Bayesian in the same direction as and to an equal or greater degree than procesing data as a single compond event. In the Bayesian special case where conservatism equals 1, the two methods yield the same result.

In the task environment of the present study, Pattern P can result in any subjective probability of .55 or greater; all four of the subjective probabilities reported in games in which Pattern P occurred were within this range. Pattern Q can result in a subjective probability between .45 and 0.59; for conservatism of 2.64 or more, well beyond the range originally mapped, pattern Q can also occur and produce subjective probabilities greater than .96, but none of the eight games in which subjects' behavior matched Pattern Q resulted in reported subjective probabilities in either of these ranges. Finally, Pattern W can lead to subjective probabilities between .60 and .96, and the reported subjective probabilities in four out of the five games matching Pattern W fell within this range. Overall, seven out of seventeen games had a reported subjective probability consistent with the parameter region of the Myopic Conservative Bayesian Decision Maker which gives rise to a behavior pattern matching the subject's behavior. These results imply that, whether or not the Myopic Conservative Bayesian Decision Maker is an adequate model of that part of a subject's cognitive processes which leads to his choices of tests and treatments, it does not appear promising in its present form as a model of the process which generates his introspective report of subjective probability.

4. PRACTICAL AND SCIENTIFIC IMPLICATIONS

The primary purpose of this chapter is to evaluate the potential of the cartographic research paradigm presented in this dissertation as one means for improving our knowledge of the similarities and differences between human decision makers, especially between experts and neophytes. A secondary purpose is to evaluate the particular theoretical model used as an example in this study, the Myopic Conservative Bayesain Decision Maker.

Section 4.1 will analyze the results of the pilot study as an evaluation of the Myopic Conservative Bayesian Decision Maker and as an example of the use of the cartographic paradigm. Section 4.2 presents a sampling of alternative theoretical models suitable for use with the paradigm in future research projects. Section 4.3 outlines some promising avenues for future research using the cartographic paradigm or extensions of it together with theoretical models based on those presented in section 4.2. Finally, sections 4.4 and 4.5 discuss the potential benefits to professional education and to decision support systems design that would arise from a long-term research effort built on the foundations established in this dissertation.

4.1 IMPLICATIONS OF THE PILOT STUDY

4.1.1 Substantive Implications: Evaluation of the Model

The empirical findings of the pilot study reveal three important shortcomings of the Myopic Conservative Bayesian Decision Maker at the low-to-moderate levels of experience and motivation examined. These shortcomings are: a total lack of fit between model and behavior when inconsistent test results are received; a low rate of fit between model and behavior when consistent test results are received; and self-reports of subjective probability consistently greater than predicted by the theory underlying the model.

As discussed in the previous chapter, the subjects who received inconsistent test results can be divided into two classes, depending on whether their behavior diverged from the model's before or after enough tests had been purchased to make the inconsistency apparent (i.e. a positive and a negative on repetitions of the same test or a positive on Tese V and negatives on both specific-virus tests). The implications for the Myopic Conservative Bayesian Decision Maker of the subjects who diverged before receiving inconsistent results must logically be analyzed together with those who diverged without inconsistent results, as neither group's divergence can be attributed to any effect specific to inconsistent test results. The remainder, who diverged only after receiving an inconsistent set of test results, consists of eight subjects in the first two preliminary experiments, and four subjects in Game 9 of Experiment 4. While their behaviors show a variety of patterns, the data strongly suggest that when a person receives information which contradicts his expectations, this information is processed in a very different way from other information. For example, some subjects appear to be reacting in these experiments by "backtracking," returning to an earlier stage by discounting one or more tests to reduce the cognitive inconsistency. This issue will be further discussed in Section 4.2 as it affects the choice of components for new candidate theoretical models to replace the Myopic Conservative Bayesian Decision Maker in future research.

Looking only at behavior sequences without inconsistent test results, the data indicate some learning effect. In the preliminary experiments, eight subjects departed from the model without having received inconsistent test results and only four subjects made choices that the model could match. The more experienced subjects in Experiment 4 departed from the model ten times before receiving inconsistent test results and matched the model twenty times. In other words, after the cases of divergence following inconsistent test results discussed in the previous paragraph are removed, two thirds of the subjects with no prior experience on the game diverged from the model while only one third of the games played by the experienced subjects in Experiment 4 diverged. These results do not preclude the existence of higher levels of experience at which virtually all games would match one or another of the behavior patterns the model can generate under the important restriction of no inconsistent test results; however, the Myopic Conservative Bayesian Decision Maker does seem clearly inadequate for its original purpose of analyzing the differences between expert and neophyte decision makers.

The reason for this failure may be partially elucidated by examining the data on subjective probabilities. The Conservative Update Module of the Myopic Conservative Bayesian Decision Maker is based on the assumption that the same kind of information processing is used in evaluating sources of future information and in evaluating information already received. Furthermore, it assumes that information processing in the game used in this research is of the same kind as was elicited in the research paradigm developed by Edwards (1968, 1972) and used by Schum and Martin and others, in which the subjects have no choice in the amount and nature of the data they receive. As the results in the previous chapter show, the self-reported subjective probabilities are not in general within the ranges predicted by the model. More significantly, the central finding of previous work on subjective probability upon which the model is based, conservatism, is strongly contradicted by the present data. The only element of the present experimental task which distinguishes it from previous studies in this area (Slovic and Lichtenstein, 1971) is the fact that subjects select and pay for data rather than having it selected and given to them by the experimentor. An important question for future research, as discused in Section 4.3.1 below, is the separation of the effects of paying for data from the effects of the fact that the data sources were selected by the subject.

4.1.2 Methodological Implications: Evaluation of the Paradigm

The empirical research reported on in this dissertation is an example of the use of the cartographic paradigm to evaluate one particular candidate theoretical model, the Myopic Conservative Bayesian Decision Maker. This two-parameter model was derived from well-established prior research, namely Edwards' and Schum and Martin's findings on conservatism and Gorry's myopic variant of decision analysis. The model was formalized in a simulation program: the map of parameter space from which the "cartographic paradigm" takes its name was then prepared by a series of simulation runs which found the critical parameter values which mark the boundaries between regions leading to different behavior patterns in the task environment used in the study. The behavior patterns of human subjects confronted with the same task environment was then examined to determine which region in parameter space, if any, corresponded to each human behavior pattern.

This research, which followed the original design of the study according to the cartographic paradigm, revealed several important facts about the model: (1) the model is inadequate to account for behavior subsequent to the receipt of an internally inconsistent set of test results; (2) in the absence of inconsistent test results, the model performs substantially better with experienced than with inexperienced decision makers, but it still fails to match a third of the experienced decision makers' patterns even when inconsistent test results are excluded; and (3) 16 out of the 24 casees of behavior that matched the model, and thus could be categorized as to conservatism and myopia, correspond to regions in parameter space with myopia = 2 and conservatism 0.65 or greater (although 6 of these 16 cases showed Pattern 4, which also occurs for myopia =3 when conservatism is greater than .96.)

These findings are sufficient to show the paradigm, as opposed to the model, to be a success; the paradigm has exposed the flaws in an otherwise plausible model, and has identified the parameter region which warrants closest study if the model, or another model using a parameter space including conservatism and myopia among its dimensions, is used in future research. Experiment 4 also included an additional measure, self-reported subjective probability. This measure was not called for by the cartographic paradigm, nor does it have any part in the analysis based on that paradigm. However, self-reported subjective probability served as a useful supplement to the study in examining the reasons for the model's failure; the lack of conservatism in the data may well be the most important substantive finding of the pilot study. In future studies, it should be possible to incorporate more sophisticated direct measures of subjective probability into the economic behavior considered by the cartographic paradigm. Edwards (1967) gives an example of direct inferences of subjective probability from a laboratory exercise of economic choice.

4.2 ALTERNATIVE THEORETICAL MODELS

4.2.1 Update Models

If a theoretical model seeks to predict or explain the process whereby a person selects costly information upon which to base a decision, the model must consider the person's state of belief about the state of the world, and how his state of belief changes as new information about the world is received. This state of belief is formalized in Chapter 2 as the salience vector π in axiom (D2); the generalized concept is operationalized in the Myopic Conservative Bayesian Decision Maker as a subjective probability vector $\hat{\pi}$, which is updated by the Conservative Update Module.

DOCS, the Doctor Simulation System (Chan, 1974) satisfies most of the requirements placed upon a theoretical model by the cartographic paradigm; it is a simpler model than the Myopic Conservative Bayesian Decision Maker, although it is embedded in a much richer task environment. In DOCS, the salience of a disease is equal to the number of a patient's known symptoms which are also symptoms of that disease, divided by the total number of the patient's known symptoms. Provision is made for weighting the importance of symptoms, but this is not used in the current version. The generality of the weighting mechanism may make DOCS a useful testbed for alternative update models in future research, as discussed is Section 4.3.2 below.

Most other non-Bayesian formal models of opinion change, such as that of Wallsten (1976) or the regression model reviewed by Slovic and Lichtenstein (1971), address only the relative attractiveness of two hypotheses. It is frequently asserted (e.g. Wallsten page 196) that the two-alternative case can be readily generalized to several alternatives; however, even the highly simplified task environment used in the present research demonstrates that this is not necessarily the case when the availability of tests such as Test V organize the possible states of the world into hierarchies and/or subsets. Thus, published two-alternative models require analytic extension and empirical testing in the multiple-alternative case before they can be used as a component for a model of choice of information. Nevertheless, this is an important avenue for future research; models like Wallsten's, developed from the start as models of actual human behavior, promise greater accuracy than such models as the Myopic Conservative Bayesian Decision Maker which are derived from a formula such as Bayes' rule, which is optimal only when all of its restrictive assumptions are met and information processing is essentially unlimited.

Another drawback to the use of Bayesian formulas, conservative or otherwise, to model human behavior when information is received sequentially has to do with information which contradicts the currently-favored hypothesis. Bayes' rule and its variants handle confirming and disconfirming information in the same way, algebraically combining the odds ratio or likelihood ratio of the new data with the prior odds or probability. The results of the pilot experiment as well as much published research (e.g. Abelson et al, 1968) indicate that humans, whether naive or expert, can react in quite different ways to confirming versus disconfirming data. One promising line of research is backtracking, where a mutually-inconsistent set of test results is discounted or disregarded to return the subject's opinions to an earlier, consistent state. The literature of artificial intelligence (e.g. Sussmann, 1977) and cognitive consistency theory (Abelson et al, 1968) will be useful sources for future models incorporating this phenomenon.

4.2.2 Choice Models

The choice module of any theoreticl model within the cartographic paradigm is what converts the opinion state maintained by the update module into the observable choices of information sources and final alternative which form the observer's data. The choice module must handle three kinds of decision: whether to continue seeking information or make a final choice; what kind of information to purchase next; and what final alternative to select.

Essentially all the published work in ths area that uses explicitly priced information is based on the decision analysis algorithm; Raiffa (1968) and other authors advocate the full decision analysis algorithm as a normative model, Gorry et al (1973) use a myopic version of the algorithm for a project in computer-assisted medical decision-making, and Fried and Peterson (1967) compare human behavior with the optimum in the "optional stopping" task environment in which the only real decision is when to cease purchasing repeated samples of the same binary test and make a final choice between two alternatives.

When information is not explicitly priced, the decision to cease testing and select a final alternative may be made by a threshhold salience for the most salient state, or by defining some test results or combinations therof as leading to "certainty;" in the medical literature, this is referred to as a "pathognomic" set of signs and symptoms.

In most research on choices of information, it is assumed that the final choice of alternative is relatively straightforward once the set of "decision premises" has been determined: "Given a complete set of value and factual premises, there is only one decision consistent with rationality" (Simon, 1957). Thus, the key to the final choice of alternative lies almost enitrely in the choices of information sources.

One characteristic of the decision analysis algorithm is that the choice of each test is made based on the outcomes of every previous test; this feature is accentuated when myopia is introduced, because the n'th choice in the sequence is not even hypothetically analyzed at its maximum depth of search (myopia) until after the results of the (n-1)'th test have been received. However, a promising heuristic for reducing the decision-making burden is to schedule tests in batches, rescheduling only if the batch is exhausted without yielding sufficient premises for a final choice of alternative. This heuristic is used in DOCS as well as in other published normative and descriptive models in medical decision making, especially with regard to laboratory tests (Pau, 1974). The need for research on the use of this heuristic in future studies involving the cartographic paradigm is underscored by the tendency of some subjects on the preliminary experiments to make requests for information such as "Test A and Test B" -- as this heuristic was not the focus of the pilot study, these subjects were simply told "one at a time, please," but a future task environment could easily be constructed which would allow batching. (The structure of Experiment 4 automatically precluded such requests.)

Another aspect of human behavior, which Fried and Peterson demonstrated in a simple task environment and which may well have played a role in the behavior of the subjects in the present experiment, is premature termination. Neglecting fixed-budget effects not objectively present in the task environment, optimal strategies call for decisions based only on the present state of knowledge and the future costs, payoffs and probabilities. However, humans tend to resist "throwing good money after bad;" thus, a series of tests which objectively cancel each other out and should return the subject to his starting point tend instead to make him cease testing, even if this means applying the broad-spectrum treatment. Any complete theoretical model of human behavior must take into account this human trait, which is probably useful more often than not outside the laboratory where budget restraints do exist and the value of the information sources themselves (i.e. the conditional probabilities) are only imperfectly known.

4.3 LONGTERM RESEARCH PROGRAM

The principal goal of this dissertation has been to develop and to demonstrate the use of a new methodology for the study of human decision-making processes, the "cartographic paradigm." This section deals with three areas of future research which will build directly upon this work: laboratory experiments to clarify some issues raised by the pilot study and to develop new theoretical models; a large ongoing research program in problem-solving in an ill-structured clinical environment in the MSU College of Human Medicine; and a program of applied research in management information and decision in association with the management gaming program of the Georgia State University College of Business Administration.

4.3.1 Laboratory Experiments

The pilot study raises some important issues in the area of subjective probability, which can be studied using some relatively simple laboratory exercises. As discussed in Chapter 3, subjects in Experiment 4 indicated (by marking a calibrated line) higher levels of confidence than were justified by Bayes' rule. In a large number of earlier studies reported by Edwards and others, confidence levels below the Bayesian were found. The only unique elements which differentiate the present experimental task from those of earlier studies are the fact that the data presented to the subject were of the subject's own choosing (that is, the subject decided which test to call for; obviously, he did not choose whether the test was to have a positive or negative result), and that the subject was required to pay for the data. Clearly, research is needed in which classic experiments in the literature of conservatism are replicated varying these two factors of choice and payment.

The finding may also have some implications for the theory of organizational behavior; it is well known that an individual who participates in a decision-making process, especially in a small group setting, tends to be more committed to the ensuing decision. (Vroom, 1964). Without denying the obvious sociological elements in this facilitation effect, the findings in the present study imply that a person who has invested time and effort in determining some of the premises upon which the group decision was based may have higher confidence in the correcntess of the decision by virtue of that fact. This conjecture may be tested by designing and carrying out a series of experiments contrasting

(a) group decisions based on group-determined premises;

(b) group decisions based on individually-determined premises;

(c) group decisions based on externally-determined premises;

(d) individual decisions based on group-determined premises;

(e) individual decisions based on individually-determined premises; and

(f) individual decisions based on externally-determined premises.

Another avenue for laboratory research building upon this dissertation is the development of new theoretical models based on the concepts outlined in Section 4.2. None of these concepts are as fully developed as the conservative Bayesian update formula and the myopic decision analysis algorithm, which is why the latter concepts were used in the pilot study; thus, the first step is to explore the extensions needed to published models for use in a richer task environment with explicit costs and benefits. For example, research is currently underway in collaboration with Dr. Chan involving a replication of Fried and Peterson's (1969) optimal stopping experiment augmented with self-report of subjective probabilities (Whalen and Chan, 1979). Future research along these lines will include extending Wallsten's (1976) algebraic models to handle sequential receipt of information, and modifying DOCS (Chan, 1974) to operate with less user intervention. This research is expected to lead to a new generation of theoretical models suitable for use in the two large research projects described in Sections 4.3.2 and 4.3.3 below.

4.3.2 Problem-Solving Research in a Medical Environment

The cartographic paradigm will be one of many elements of the continuing program of research into structural and behavioral aspects of ill-structured problem solving being carried out within the Michigan State University College of Human Medicine (Chan, 1974a,b; 1978; Chan and Whalen, 1979; Whalen and Chan, 1979). In the next phase of this research, practicing physicians and medical students will be the subjects for experiments using a very complex task environment of medical problem-solving.

The data to be collected in these experiments consist of "decision traces," which correspond roughly to the "behavior patterns" described in this dissertation, augmented with two kinds of self-reported confidence level. The decision traces wil be recorded from the interactions between individual physicians or medical students and a computerized parient case record system, with a special process tracer acting as front end to collect the data.

One of the key phases in the analysis of this data is a cluster analysis of decision traces, to find a collection of distinct canonical strategies which characterize fundamental ways of approaching a problem. Before such a cluster analysis can be carried out, however, a metric for the distance between pairs of decision traces is needed. This metric will be defined primarily in terms of a "deep metric" over an underlying parameter space, which will have dimensions comparable to conservatism and myopia plus additional dimensions to be determined by an interdisciplinary team of information scientists and medical educators. Computer-assisted role playing will then be used, in a manner analogous to the simulation runs in this dissertation, to find the decision traces which arise from the important "landmarks" in the parameter space. Finally a "surface metric" will be used to locate observed decision traces relative to these landmarks by identifying trivial order effects or other unimportant differences. The reason for the reliance on the "deep metric" and minimization of the role of the "surface metric" lies in the occasions when a slight shift in strategy, or an arbitrary choice among alternatives to which a single strategy is indifferent, leads to a different choice early in the decision trace which in turn leads to sharply divergent surface behavior. In such a case, the two traces would seem far from each other according to the surface metric, but they would be close to respective members of a pair of landmarks known, via the deep metric, to be close to each other.

4.3.3 Applied Research Using Management Gaming

The management gaming programs of Georgia State University (Nichols and Schott, 1973; Thompson, 1978) have a strong component of decision support systems which students may optionally choose to use to aid them in making decisions in a simulated competitive business environment. At present, no record is kept of the use of these decision support systems, but preliminary conversations with the persons involved in maintaining them indicate that it is feasible to modify the programs to collect some very useful data about the information sources and statistical transformations used by students in management courses at the graduate and undergraduate levels. Knowledge obtained from comparing these data with data on the choices and performance of the students in the games themselves can then lead to changes in the game structure in the direction of greater realism, which would be of benefit both to research in decision-making and to the education of the students using the games.

4.4 IMPLICATIONS FOR PROFESSIONAL EDUCATION

One of the more difficult aspects of professional education in many fields has to do with helping the student acquire competent judgement in the use of costly, imperfect sources of data. Data sources may have a clear monetary cost, such as diagnostic tests in medicine, destructive testing in engineering and market survey research in business, and there may be risks involved in obtaining the data, such as in exploratory surgery or test-piloting experimental aircraft. A more insidious cost of information is the time and effort required to digest the information once it is on hand, which often outweighs the benefits even if the information is obtained at zero marginal cost.

In principle, the decision analysis algorithm is the optimal solution to the problem of costly imperfect data; however, in many if not most practical cases the cost of applying the algorithm itself, including the data acquisition and interpretation cost of obtaining the high-quality numeric data required by the algorithm, exceeds the potential benefits if the algorithm is feasible at all. Professional decision makers, on the other hand, successfully solve such problems on a near-routine basis. The difficulty in training their successors or expanding their number arises because the ability to solve problems does not imply the ability to explain how one solves them in a manner that can readily be put into practice by a student.

Learning by example and by supervised practice can never be superceded entirely, nor should they in a responsible profession. However, an improvement in the language with which we talk about this judgement process can greatly improve the students' preparedness to benefit from example and practice. The potential of a program of research carried out under the cartographic paradigm introduced in this dissertation is to develop and validate theoretical models and localize expert decision processes to regions within the psychologically meaningful parameter spaces of those models, using the actual behavior of expert decision makers in a standardized task environment instead of their less-reliable introspections about why they chose that behavior. Students' strategies could then be localized elesewhere in the parameter space, and the ways that the students should change if they want to become more like the experts can be explained in terms of the parametric dimensions of the space.

4.5 IMPLICATIONS FOR DECISION SUPPORT SYSTEMS

The kind of information about human decision-making processes that would be obtained from a program of research using the cartographic paradigm would be of benefit to designers of organizational and automated decision support systems in two ways. First, any methodology which improves our ability to predict what information a decision maker will find useful can be used to design systems that make that information more readily available -- and equally important, separate the useful information from the mass of data which is not useful in the particular decision process used by the decision maker.

The second potential contribution is more specifically linked to the cartographic paradigm per se. If the region of parameter space in which a decision maker habitually operates is near, but outside, a region which leads to appreciably better reults on the average, then it may be possible to supply artificial aids which are tailored to the decision maker's own characteristics and designed to help him attain higher performance without interfering with his judgement and responsibility. An example of this philosophy can be found in Edwards' work on the Probabilistic Information Processor (PIP), summarized in (Edwards, 1968). In PIP, humans interpreted incoming data by stating the conditional probability of each event that occurred (rather than each possible event, an immensely larger undertaking) given each of several hypotheses. A computer then used these conditional probabilities to update the probabilities of the respective hypotheses according to Bayes' law, avoiding the conservatism shown by a control group of subjects who updated the probabilities themselves. However, the experimental task environment Edwards used points up the need for extreme caution whenever human judgement is supplemented or supplanted by a formal algorithm. In the experiment, conducted in the nineteen-sixties, subjects interpreted events in an imaginary world crisis set in 1975 with respect to the likelihood of conventional or nuclear war, and Edwards defined improved performance in most cases as a higher (less conservative) inferred probability of imminent war after a few events pointing in that direction. The advantage of an aid such as PIP over one which is constructed without reference to human characteristics is that PIP's results can be easily critiqued because the process is close enough to human thinking to be understood; thus, one can say "I think we NEED to be more conservative in this situation." On the other hand, when a decision maker's staff or computer makes a recommendation on the basis of a process foreign to the decision maker's own thought processes, the decision maker can only "take it or leave it."

5. SUMMARY AND CONCLUSION

The cartographic paradigm for research in human decision processes, introduced and demonstrated in this dissertation, draws upon earlier research in the areas of problem-solving protocols, subjective probabilities, weighting coefficients, and choice of information. These intellectual debts are acknowledged in Chapter 1, and the present approach is contrasted with each of its antecedents.

Chapter 2 contains a formal definition of an experimental or observational system composed of a decision-making task with costly imperfect data, a decision maker who seeks to maximize his subjective expected utility, a candidate theoretical model of the decision-making process, and an observer who seeks to explain the decision maker's overt behavior (information purchases and final choice of alternative) in terms of the theoretical model. A theoretical model within this system consists of a parameter space each of whose dimensions is a psychologically meaningful characteristic of a human decision maker, plus an algorithm for determining the sequence of test purchases and final choice of alternative that would arise when a decision maker characterized by a particular point in parameter space confronts a particular decision-making task. The requirements for a theoretical model are further clarified in Chapter 2 by the presentation of a simple example, the two-parameter Myopic

Conservative Bayesian Decision Maker.

Chapter 3 reports on a pilot study which demonstrates the use of the cartographic paradigm; in the pilot study, the Myopic Conservative Bayesian Decision Maker is evaluated in the context of a simulation game in veterinary medicine which was played by four groups of subjects under different conditions of expertise, experience, and motivation. The first phase of the pilot study was to specify the details of the decision-making task in such a way that a sufficient range of behavior patterns resulted from reasonable variations in the model's parameters of conservatism and myopia, and to design the initial set of experimental materials with which to convey this task to the subjects.

The second phase consisted of roughly fifty simulation runs using a PASCAL implementation of the Myopic Conservative Bayesian Decision Maker to produce the map of parameter space from which the cartographic paradigm takes its name. This map was constructed in four segments, for myopia equal to 1, 2, 3, and 4; within each segment, the critical values of conservatism at which the pattern of overt behavior changed were found. At such a critical value, the simulated decision maker is indifferent between two substantively different strategies, while all values of conservatism between two critical values at a given level of myopia lead to identical information choices and final choice of alternative in the particular task environment used, and thus define a region within the parameter space.

The third phase of the pilot study was to carry out three preliminary experiments in which the task environment was further refined and selection criteria for appropriate subjects were developed. The final phase of the pilot study was Experiment 4, in which twenty subjects played the simulation game ten times each; the last two games each subject played had standardized (rigged) test results and were used for analysis under the cartographic paradigm and also in terms of a self-report of subjective probability similar to that used by Edwards (1968).

The results of these experiments show that the Myopic Conservative Bayesian Decision Maker is not adequate as a model of human decision processes at the levels of motivation and experience studied, especially when any of the tests purchased by the decision maker give results contradicting the hypothesis supported by the previous tests. Analysis of the 24 games in which subjects did use behavior patterns explainable by the model did show clustering within a particular region of parameter space, however; this information may prove useful in conjunction with any future model which includes conservatism and myopia among its parameters.

The supplementary analysis based on self-reported subjective probability produced an important unanticipated result. Edwards (1968, 1972) and other researchers found human estimates of a posteriori probability to be consistently less than the optimal Bayesian estimate in a wide variety of circumstances in which subjects are given probabilistic information; the present study indicates the opposite effect when subjects choose and pay for the information they receive.

Chapter 4 begins with a discussion of the substantive and methodological implications of the pilot study itself. Following this, some promising concepts around which new theoretical models can be built are presented, and some specific opportunities for future research involving the cartographic paradigm are examined. The last two sections of Chapter 4 are concerned with the potential benefits of a successful program of research under the cartographic paradigm to professional education and to decision support systems respectively.

APPENDICES

APPENDIX 1: DERIVATION OF $E_{S,Q}(\hat{\pi}_t, r_k^{\ell}, \Psi_1)$

APPENDIX: Derivation of $E_{S,Q}(\hat{\pi}_t, r_k^\ell, \Psi_i)$

Given a set of exclusive, exhaustive hypotheses σ_i, $i = 1, 2, \ldots, s$ with prior probability $P(\sigma_i)$, and given datum D with conditional probability $P(D|\sigma_i)$ for each i.

By Bayes' theorem, $P(\sigma_i|D) = P(\sigma_i)P(D|\sigma_i) \Big/ \sum_{m=1}^{s} P(\sigma_m)P(D|\sigma_m)$.

In odds form, the equivalent formula is

$$\frac{P(\sigma_i|D)}{1 - P(\sigma_i|D)} = \frac{P(\sigma_i)P(D|\sigma_i) \Big/ \sum_{m=1}^{s} P(\sigma_m)P(D|\sigma_m)}{\left(\sum_{m=1}^{s} P(\sigma_m)P(D|\sigma_m) - P(\sigma_i)P(D|\sigma_i)\right) \Big/ \sum_{m=1}^{s} P(\sigma_m)P(D|\sigma_m)}$$

$$= \frac{P(\sigma_i)P(D|\sigma_i)}{\sum_{m=1}^{s} P(\sigma_m)P(D|\sigma_m) - P(\sigma_i)P(D|\sigma_i)}$$

$$= \frac{P(\sigma_i)}{1 - P(\sigma_i)} \left[\frac{P(D|\sigma_i)\left(1 - P(\sigma_i)\right)}{\sum_{m=1}^{s} P(\sigma_m)P(D|\sigma_m) - P(\sigma_i)P(D|\sigma_i)}\right],$$

where the factor in square brackets is the "likelihood ratio" for D; that is, the probability of D given σ_i divided by the probability of D given the complement of σ_i.

Schum and Martin (1968), extending the work of Edwards (1968), found that human judgements of probabilities can be well modelled by raising the factor in square brackets to a power that (in general) differs from 1.0. Substituting $\hat{\pi}_{t+1}^{i}$ for $P(\sigma_i|D)$, $\hat{\pi}_t^{i}$ for $P(\sigma_i)$, $F_k(r_k^\ell, \sigma_i)$ for $P(D|\sigma_i)$, $F_k(r_k^\ell, \sigma_m)$ for $P(D|\sigma_m)$, and raising the factor in square brackets to the power of Ψ_i leads to the given formula for the conservative Bayesian update function $E_{S,Q}(\hat{\pi}, r_k^\ell, \Psi_i)$

APPENDIX 2: Pascal Program MYOPIC

```
PROGRAM MYOPIC (INPUT, OUTPUT);

CONST     S = 4     (* NUMBER OF STATES OF THE WORLD *);
        ALT = 5     (* NUMBER OF ALTERNATIVE FINAL CHOICES *);
        PHI = 5     (* NUMBER OF POSSIBLE TESTS *);
         NK = 2     (* NUMBER OF RESULTS FOR TEST K *);

TYPE STATE = 1..S        (* STATE OF THE WORLD *);
    CHOICE = 1..ALT      (* CHOICE OF FINAL ALTERNATIVE *);
      TEST = 1..PHI      (* TEST *);
    RESULT = 1..NK       (* RESULT OF TEST *);
   OPINION = ARRAY[1..S] OF REAL (*  SUBJECTIVE PROBABILITY VECTOR *);
   DISTRIB = ARRAY[1..S] OF REAL  (* VECTOR-VALUED ELEMENT OF F OR Y *);

VAR          I : STATE;
             J : CHOICE;
             K : TEST;
             L : RESULT;     (* I,J,K,L = INDEXES *)

           PIT : OPINION     (* "PRIOR" OPINION AT TIME T *);
     PITPLUS1 : OPINION      (* "POSTERIOR" OPINION AT TIME T+1 *);

             Y : ARRAY[1..ALT] OF DISTRIB (* PAYOFF MATRIX *);
             C : ARRAY[1..PHI] OF REAL (* COST OF TEST K *);
             F : ARRAY[1..PHI,1..NK] OF DISTRIB
                             (* CONDITIONAL PROBABILITIES *);


        KSTAR : TEST         (* INDEX OF "BEST" TEST *);
        JSTAR : CHOICE       (* INDEX OF "BEST" CHOICE *);
         BEST : REAL         (* VALUE OF CURRENT BEST ACTION *);
         TEMP : REAL         (* TEMPORARY HOLDING VARIABLE *);

   ITERATION : INTEGER       (* CONTROL VARIABLE FOR MAIN LOOP *);

         PSI1 : REAL         (* CONSERVATISM - BAYES = 1.0 *);
         PSI2 : INTEGER      (* MYOPIA - GORRY = 1, RAIFFA = INFINITE *);

FUNCTION U (PSI2:INTEGER; PITPLUS2:OPINION):REAL; FORWARD;

FUNCTION PRODUCT (VAR ROW:DISTRIB; PI:OPINION):REAL;

                             (* SCALAR PRODUCT OF TWO VECTORS *)

VAR I : STATE (* INDEX OVER ELEMENTS *);
    X : REAL (* DUMMY FOR PRODUCT *);

BEGIN  (* PRODUCT *)

X := 0.0;
FOR I := 1 TO S DO  X := X + ROW[I] * PI[I] ;

PRODUCT := X;

END (* PRODUCT *);

PROCEDURE E (VAR PITPLUS2:OPINION; PITPLUS1:OPINION; K:TEST; L:RESULT;
     PSI1:REAL);

VAR  I : STATE; (* INDEX TO CONSIDER EACH TEST *)
  TEMP : REAL;

FUNCTION OMEGA (VAR I:STATE) : REAL;

     VAR PRIOR : REAL (* A PRIORI ODDS *);
         RATIO : REAL (* ODDS RATIO FOR L AND I *);
           NUM : REAL (* NUMERATOR OF ODDS RATIO *);
         DENOM : REAL (* DENOMINATOR OF ODDS RATIO *);
   BEGIN  (* FUNCTION OMEGA *)
     PRIOR := PITPLUS1[I] / (1-PITPLUS1[I]);
     NUM := F[K,L][I] * (1-PITPLUS1[I]);
     DENOM := PRODUCT(F[K,L],PITPLUS1) - F[K,L][I]*PITPLUS1[I];
     RATIO := NUM / DENOM;
     IF PSI1 = 1.0 THEN OMEGA := PRIOR*RATIO
          ELSE OMEGA := EXP(LN(PRIOR)+PSI1*LN(RATIO));
END (* FUNCTION OMEGA *);

BEGIN  (* PROCEDURE E *)
FOR I := 1 TO S DO (* CONVERT ODDS TO PROBABILITY *)
     BEGIN
```

```
        TEMP := OMEGA(I);
        PITPLUS2[I] := TEMP / (1.0+TEMP)
    END (* LOOP FOR I *);

END (* E *);

FUNCTION V (K:TEST; PITPLUS1:OPINION; PSI2:INTEGER): REAL;
    (* VALUE OF BUYING TEST K GIVEN OPINIONS REPRESENTED BY PITPLUS1,
       LOOKING PSI2 STEPS (FURTHER) AHEAD *)

VAR       L : RESULT   (* INDEX OVER EACH POSSIBLE RESULT OF TEST K *) ;
  PITPLUS2 : OPINION  (* NEW OPINION IF TEST K WERE TO GIVE RESULT L *);
      VDUM : REAL     (* DUMMY FOR V *) ;

BEGIN (* FUNCTION V *)

VDUM := 0;
IF PSI2 > 0 THEN
    BEGIN FOR L := 1 TO NK DO
        BEGIN
            E(PITPLUS2, PITPLUS1, K, L, PSI1);
            VDUM := VDUM + U(PSI2,PITPLUS2) * PRODUCT(F[K,L],PITPLUS1);
        END (* SUMMATION OVER L *);
        VDUM := VDUM - C[K];
    END (* IF PSI2 > 0 *);

V := VDUM

END (* V *);

FUNCTION U (* (PSI2:INTEGER; PITPLUS2:OPINION):REAL; *) ;

    (* EXPECTED FINAL PAYOFF GIVEN OPINION PITPLUS2 OPINION VECTOR
       AND LOOKING PSI2 STEPS (FURTHER) AHEAD *)

VAR      K : TEST    (* INDEX TO CONSIDER POSSIBLE TESTS *);
         J : CHOICE  (* INDEX TO CONSIDER ALTERNATIVE FINAL CHOICES *);
      TEMP : REAL    (* VALUE OF LAST ACT CONSIDERED SO FAR *);
    UDUMMY : REAL (* DUMMY FOR U *) ;

BEGIN (* FUNCTION U *)
UDUMMY := -999.9 (* INITIALIZE U VERY LOW *) ;

FOR K := 1 TO PHI DO (* FIND VALUE OF BEST TEST *)
    BEGIN
        TEMP := V(K, PITPLUS2, PSI2 - 1);
        IF TEMP > UDUMMY THEN UDUMMY := TEMP
    END (* MAXIMIZATION OVER I *);

FOR J := 1 TO ALT DO (* FIND VALUE OF BEST FINAL CHOICE *)
    BEGIN
        TEMP := PRODUCT (Y[J], PITPLUS2);
    .   IF TEMP > UDUMMY THEN UDUMMY := TEMP
    END (* MAXIMIZATION OVER J *);

U := UDUMMY

END (* FUNCTION U *);

(*
    INITIALIZE:READ IN AND ECHO DATA
*)
BEGIN (* MAIN PROGRAM *)
TEMP := 1.0/S; WRITELN (? INITIAL PROBABILITY = ?, TEMP, ?EACH?);
FOR I := 1 TO S DO PIT[I] := TEMP; (* INITIALIZE PIT *)

FOR J := 1 TO ALT DO (* READ IN PAYOFF MATRIX *)
    BEGIN FOR I := 1 TO S-1 DO
        BEGIN (* INPUT LINE = PAYOFFS FOR ACT J *)
            READ (Y[J][I]);
            WRITE (Y[J][I])
        END;
        READLN (Y[J][S]); WRITELN (Y[J][S]);
    END; (* DONE READING PAYOFF MATRIX *)

WRITELN (? ***?);

FOR K := 1 TO PHI DO (* READ IN TEST CHARACTERISTICS *)
    BEGIN
        READLN (C[K]); WRITELN (C[K]) (* COST OF TEST K *);
        FOR L := 1 TO NK DO (* CONDITIONAL PROBABILITIES *)
            BEGIN FOR I := 1 TO S-1 DO
                BEGIN READ (F[K,L][I]); WRITE (F[K,L][I]) END;
                READLN (F[K,L][S]); WRITELN (F[K,L][S]);
            END (* DONE WITH OUTCOME L OF TEST K *)
    END (* DONE WITH TEST K *);

WRITELN (? ***?);
```

```
READLN (PSI1); WRITELN (? CONSERVATISM = ?, PSI1);
READLN (PSI2); WRITELN (? MYOPIA = ?, PSI2);

WRITELN (? ***?);

(*
   BEGIN SIMULATION
*)


ITERATION := 1 (* ITERATION = CHOICE OF AN ACT *);
WHILE ITERATION < 11 DO
    BEGIN
        WRITELN; WRITELN;
        WRITELN (? ITERATION = ?, ITERATION);

        BEST := -999.9;
        FOR K := 1 TO PHI DO
            BEGIN (* FIND BEST TEST AND ITS VALUE *)
                TEMP := V (K, PIT, PSI2);
                IF TEMP > BEST + 0.0000000001 THEN
                    BEGIN BEST := TEMP; KSTAR := K END;
                WRITE (K, ?:?, TEMP)
            END (* DONE EVALUATING TESTS *);
        WRITELN;

        FOR J := 1 TO ALT DO
            BEGIN (* SEE IF ANY FINAL CHOICE HAS HIGHER VALUE *)
                TEMP := PRODUCT (Y[J], PIT);
                IF TEMP > BEST THEN
                    BEGIN
                        ITERATION := 99 (* NO MORE TESTS *);
                        BEST := TEMP;
                        JSTAR := J
                    END (* DONE EVALUATING CHOICE J *);
            END (* DONE EVALUATING ALL POSSIBLE ACTIONS *);

        IF ITERATION < 11 THEN
            BEGIN (* PURCHASE TEST KSTAR *)
                WRITELN (? PURCHASE TEST?, KSTAR);
                IF ITERATION = 2 THEN
                    BEGIN (* MISLEADING DATA ON SECOND TEST *)
                        IF (KSTAR = 1) OR (KSTAR = 5) THEN L := 2
                            ELSE L := 1
                    END
                ELSE
                    BEGIN (* REPRESENTATIVE DATA ON OTHER TESTS *)
                        IF (KSTAR = 1) OR (KSTAR = 5) THEN L := 1
                            ELSE L := 2
                    END (* TEST K HAS GIVEN DATA L *);
                WRITELN (? RESULT = ?, L);
                E (PITPLUS1, PIT, KSTAR, L, PSI1) (* UPDATE OPINIONS *);
                FOR I := 1 TO S-1 DO
                    BEGIN
                        PIT[I] := PITPLUS1[I];
                        WRITE (PIT[I])
                    END;
                PIT[S] := PITPLUS1[S];
                WRITELN (PITPLUS1[S],? = NEW OPINION?)
            END (* DONE PROCESSING TEST *);

        ITERATION := ITERATION + 1 ;
    END (* ITERATION COMPLETED.  LOOP FOR NEXT ITERATION *);

WRITELN (? ***?);
WRITELN (? FINAL CHOICE = ALTERNATIVE NO. ?, JSTAR);
WRITELN (?       EXPECTED VALUE = ?, BEST)

END. (* PROGRAM *)
*EOR 00
*EOR 17
*EOR 17
*EOF
```

APPENDIX 3: Materials for the Pilot Experiment

Materials for Experiments 1, 2, and 3

This is an experiment in decision-making. I would like you to take the role of a veterinarian in an agricultural co-op. You are involved in raising a new variety of turkeys, and there is an unidentified disease causing losses among the flock. Assume that you have already eliminated all but four possible disease organisms as the cause of the epidemic, but so far you have no reason to suspect any one of the four over the others. However, you are quite sure that one and only one out of the four diseases is present. We will call the diseases "Virus A," "Virus B," "Bacterium C," and "Bacterium D."

Your job is to prescribe a treatment for the epidemic. There are just five possible treatments; the four "specific" treatments Drug A, Drug B, Drug C and Drug D, and one "broad spectrum" treatment Drug X.

If you prescribe Drug A and the disease is Virus A,
the co-op makes $1,000 profit.
If you prescribe Drug A and the disease is not Virus A,
the co-op makes no profit.

If you prescribe Drug B and the disease in Virus B,
the co-op makes $1,000 profit.
If you prescribe Drug B and the disease is not Virus B,
the co-op makes no profit.

If you prescribe Drug C and the disease is Bacterium C,
the co-op makes $1,000 profit.
If you prescribe Drug C and the disease is not Bacterium C,
the co-op makes no profit.

If you prescirbe Drug D and the disease is Bacterium D,
the co-op makes $1,000 profit.
If you prescribe Drug D and the disease is not Bacterium D,
the co-op makes no profit.

If you prescribe Drug X, it doesn't matter what disease is present; the epidemic is partly controlled, and the co-op makes $300 profit.

(NOTE: The drugs interact harmfully, so you can only prescribe one.)

If you choose to, you may also perform up to ten diagnostic tests before prescribing a treatment; however, each test will cost $100. The tests are subject to error; if desired, you may repeat the same test to increase its reliability either right away or after performing other tests in between. Tests may be performed as many times as you want in whatever order you want, or you may choose to avoid the cost involved by calling for no tests at all. Five kinds of tests are available to choose from: Test A, Test B, Test C, Test D, and Test V.

Test A is <u>positive</u> 90% of the time (and negative 10% of the time) if the disease is Virus A;
Test A is <u>negative</u> 80% of the time (and positive 20% of the time) if the disease is Virus B, Bacterium C or Bacterium D.

Test B is <u>positive</u> 90% of the time (and negative 10% of the time) if the disease is Virus B;
Test B is <u>negative</u> 80% of the time (and positive 20% of the time) if the disease is Virus A, Bacterium C, or Bacterium D.

Test C is <u>positive</u> 90% of the time (and negative 10% of the time) if the disease is Bacterium C.
Test C is <u>negative</u> 80% of the time (and positive 20% of the time) if the disease is Virus A, Virus B, or Bacterium D.

Test D is <u>positive</u> 90% of the time (and negative 10% of the time) if the disease is Bacterium D.
Test D is <u>negative</u> 80% of the time (and positive 20% of the time) if the disease is Virus A, Virus B, or Bacterium C.

Test V, "the virus test," is <u>positive</u> 80% of the time (and negative 20% of the time) if the disease is Virus A <u>or</u> Virus B.
Test V is <u>negative</u> 80% of the time (and positive 20% of the time) if the disease is Bacterium C <u>or</u> Bacterium D.

In this hypothetical situation, you will be given $1,000 (in play money) out of the co-op's operating fund for the year, which you can use to pay for diagnostic tests. If you prescribe the

(Experiment 1)

correct specific treatment (A, B, C or D) for the actual disease, I will give you another $1,000; if you choose treatment X, I will give you another $300, but if you pick treatment A, B, C or D and the disease is not the corresponding one, you get no more play money.

To make the experiment more realistic, when all these play money transactions are over with I will buy it back at the rate of 10¢ (real) for every $100 in play money, both profits from the turkeys and money left over from the original thousand in operating funds. This is in addition to the $4.00 payment for participating in the experiment. In other words, each test reduces your bonus by a dime but might help you to win an extra dollar.

In order to let me compare different people's decisions, I have determined all the test results in advance; they are as fair as I could make them -- if anything, they are rigged slightly in your favor. If you wish, I will explain the setup after the experiment is over. In the meantime, your best bet is to act as though the chances of error on the tests are as given and the four diseases are equally likely.

Please take time to re-read these instructions, familiarize yourself with the information cards, and ask any questions you may have. Your goal is to make as much money as you can for the co-op and for yourself, using as many or as few tests as you wish and any one of the five treatments.

(Experiments 2 and 3)

correct specific treatment (A, B, C, or D) for the actual disease, I will give you another $1,000; if you choose treatment X, I will give you another $300, but if you pick treatment A, B, C or D and the disease is not the corresponding one, you get no more play money.

To make the experiment more realistic, when all these play money transactions are over with I will buy it back at the rate of 10¢ (real) for every $100 in play money, whether profits from the turkeys or money left over from the original $1,000 in operating funds. In other words, each test reduces your winnings by a dime but might help you to win an extra dollar.

In order to let me compare different people's decisions, I have determined all the test results in advance, depending on what tests you purchase. These results are as fair as I could make them; if anything, they are rigged slightly in your favor. If you wish, I will explain the setup after the experiment is over. In the meantime, your best bet is to act as though the chances of error on the tests are as given and the four diseases are equally likely.

Please take time to re-read these instructions, familiarize yourself with the information cards, and ask any questions you may have. Your goal is to make as much money as you can for the co-op and for yourself, using as many or as few tests as you wish and any one of the five treatments.

INFORMATION CARDS

Possible Disease Organisms:

Virus A
Virus B
Bacterium C
Bacterium D

One and only one is present; any of the 4 is just as likely as the others. (25% probability)

DRUG A:

If the disease is Virus A, Durg A results in a complete cure and $1,000 profit.

If the disease is Virus B, Bacterium C, or Bacterium D, Drug A has no effect, the epidemic goes out of control, and no profit is made.

DRUG B:

If the disease is Virus B, Drug B results in a complete cure and $1,000 profit.

If the disease is Virus A, Bacterium C, or Bacterium D, drug B has no effect, the epidemic goes out of control, and no profit is made.

DRUG C:

If the disease is Bacterium C, Drug C results in a complete cure and $1,000 profit.

If the disease is Virus A, Virus B, or Bacterium D, Drug C has no effect, the epidemic goes out of control, and no profit is made.

DRUG D:

If the disease is Bacterium D, Drug D results in a complete cure and $1,000 profit.

If the disease is Virus A, Virus B, or Bacterium C, Drug D has no effect, the epidemic goes out of control, and no profit is made.

DRUG X:

Drug X results in partial control of the disease and a profit of $300, regardless of whether the disease is Virus A, Virus B, Bacterium C or Bacterium D.

TEST A — Cost = $100

If the disease is Virus A, Test A is positive 90% of the time and negative 10% of the time.

If the disease is Virus B, Bacterium C, or Bacterium D, Test A is negative 80% of the time and positive 20% of the time.

TEST B — Cost = $100

If the disease is Virus B, Test B is positive 90% of the time and negative 10% of the time.

If the disease is Virus A, Bacterium C, or Bacterium D, Test B is negative 80% of the time and positive 20% of the time.

TEST C Cost = $100

If the disease is Bacterium C, Test C is positive 90% of the time and negative 10% of the time.

If the disease is Virus A, Virus B, or Bacterium D, Test C is negative 80% of the time and positive 20% of the time.

TEST D Cost = $100

If the disease is Bacterium D, Test D is positive 90% of the time and negative 10% of the time.

If the disease is Virus A, Virus B, or Bacterium C, Test D is negative 80% of the time and positive 20% of the time.

TEST V Cost = $100

If the disease is Virus A or Virus B, Test V is positive 80% of the time and negative 20% of the time.

If the disease is Bacterium C or Bacterium D, Test V is negative 80% of the time and positive 20% of the time.

DETERMINATION OF TEST RESULTS AND DISEASES
(EXPERIMENTS 1 AND 2)

IF subject begins by purchasing Test A, B, C or D:
- The disease is selected to be the one tested for
- The result of the first test is "POSITIVE"
- The result of the second test is the result LESS likely given the selected disease
- The result of any subsequent test is the result MORE likely given the selected disease

IF subject begins by pruchasing Test V:
- The result of the first test is "POSITIVE"
- IF subject's second action is purchasing Test A, then
    - the disease is selected to be disease A
    - the result of the second test is "NEGATIVE"
    - the result of any subsequent test is the result MORE likely given disease A
- IF subject's second action is purchasing Test B, then
    - the disease is selected to be disease B
    - the result of the second test is "NEGATIVE"
    - the result of any subsequent test is the result MORE likely given disease B
- IF subject's second action is purchasing Test C or D, then
    - the result of the second test is "POSITIVE"
    - any subsequent purchases of Test C or Test D produce the result "NEGATIVE"
    - any subsequent puirchases of Test V produce the result "POSITIVE"
    - IF a treatment is selected before Test A or Test B is purchased, then the disease is selected at random, $P(A) = P(B) = 1/2$, $P(C) = P(D) = 0$
    - OTHERWISE, the diesase is the virus (A or B) tested for first, and all remaining test results are the result MOST likely given the selected disease
- IF subject's second action is purchasing Testt V, then
    - the result of the second test is "NEGATIVE"
    - the disease is selected and subsequent test results determined in the same manner as if the subject's second action were purchasing Test C or Test D
- IF subject's second action is prescribing a treatment, then
    - the disease is selected at random, $P(A) = P(B) = 1/2$, $P(C) = P(D) = 0$

IF subject begins by prescribine a treatment:
- the disease is selected at random, $P(A) = P(B) = P(C) = P(D) = 1/4$

#

Materials for Experiment 4

INTRODUCTION

This experiment consists of two different games which you will play against the computer. Your total score on Game I determines your chances in Lottery I, and your total score on Game II determines your chances in Lottery II; each lottery is for twenty dollars.

GAME I

In Game I, there are two imaginary bookbags full of pokerchips; Bag R has 70% red chips and 30% white chips, and Bag W has 70% white chips and 30% red chips.

You will play the game ten times; in each game, the computer will select one of the two bags at random with equal probability, and use that bag for the entirety of that game. Every time you type "?", the computer will draw a chip from the selected bag and tell you the color of the chip (using a random number generator with the probability determined by the selcted bag.) You end the game by typing "BAG R" or "BAG W". At this point, you will be asked to mark a line to show how sure you are of your answer. (This is for my information only, and has no effect on the outcome of the game or the lottery.) The computer then tells you if the bag you indicated was the one from which the chips had been drawn.

Everyone starts out with 200 free chances in Lottery I; each chip you call for will cost you one chance, but you win 200 more chances for each of the ten games in which you correctly identify the bag.

GAME II

The object of Game II is to maximize the profit of a turkey farmer by prescribing one of five possible treatments to stop the spread of an unidentified disease among the turkeys. There are four posible diseases (Virus A, Virus B, Bacterium C, and Bacterium D); each disease has a corresponding specific treatment (Treattment A, Treatment B, Treatment C, and Treatment D), and there is also a broad-spectrum treatment, Treatment X. If the specific treatment corresponding to the actual disease is applied, the profit is $1000, but if a specific treatment is applied and the actual disease is one of the other three, the profit is zero. If the broad-spectrum treatment is applied, it doesn't matter what disease is present; the disease is partially controlled and the profit is $300. (The treatments interact harmfully, so only one out of the five may be applied.)

At the beginning of each game, all you know is that the disease is one of the four; all four are equally probable. If you wish, you can increase your information by purchasing one or more diagnostic tests; however, these tests are both expensive ($100 each) and subject to random error. Five kinds of test are available; the four specific tests Test A, Test B, Test C and Test D, and the "virus test" Test V.

For each of the specific tests Test A, Test B, Test C and Test D, the probability of a "POSITIVE" result is .90 if the corresponding disease is present, and .20 if the corresponding disease is not present. Thus, the probability of a "NEGATIVE" result on any specific test A, B, C or D is .10 if the corresponding disease is present, and .80 if the corresponding disease in son present.

For the virus test Test V, the probability of a "POSITIVE" resut is .80 if the disease is Virus A or Virus B and .20 if the disease is Bacterium C or Bacterium D. Thus, the probability of a "NEGATIVE" result on Test V is .20 if the disease is Virus A or Virus B, and .80 if the disease is Bacterium C or Bacterium D.

The above information is summarized in the table.

TABLE 1: COSTS, PAYOFFS AND PROBABILITIES
(TASK ENVIRONMENT FOR THE PILOT STUDY)

TEST A: COST = 100

PROBABILITY OF::	VIRUS A	VIRUS B	BACTERIUM C	BACTERIUM D
	IF ACTUAL DISEASE IS:			
POSITIVE RESULT:	.90	.20	.20	.20
NEGATIVE RESULT:	.10	.80	.80	.80

TEST B: COST = 100

PROBABILITY OF::	VIRUS A	VIRUS B	BACTERIUM C	BACTERIUM D
	IF ACTUAL DISEASE IS:			
POSITIVE RESULT:	.20	.90	.20	.20
NEGATIVE RESULT:	.80	.10	.80	.80

TEST C: COST = 100

PROBABILITY OF::	VIRUS A	VIRUS B	BACTERIUM C	BACTERIUM D
	IF ACTUAL DISEASE IS:			
POSITIVE RESULT:	.20	.20	.90	.20
NEGATIVE RESULT:	.80	.80	.10	.80

TEST D: COST = 100

PROBABILITY OF::	VIRUS A	VIRUS B	BACTERIUM C	BACTERIUM D
	IF ACTUAL DISEASE IS:			
POSITIVE RESULT:	.20	.20	.20	.90
NEGATIVE RESULT:	.80	.80	.80	.10

TEST V: COST = 100

PROBABILITY OF::	VIRUS A	VIRUS B	BACTERIUM C	BACTERIUM D
	IF ACTUAL DISEASE IS:			
POSITIVE RESULT:	.80	.80	.20	.20
NEGATIVE RESULT:	.20	.20	.80	.80

PAYOFFS FOR DISEASE/TREATMENT PAIRS:

	VIRUS A	VIRUS B	BACTERIUM C	BACTERIUM D
	IF ACTUAL DISEASE IS:			
TREATMENT A	1000	0	0	0
TREATMENT B	0	1000	0	0
TREATMENT C	0	0	1000	0
TREATMENT D	0	0	0	1000
TREATMENT X	300	300	300	300

When the computer says "TEST OR TREAT", you may respond "TREAT A", "TREAT B", "TREAT C", "TREAT D" or "TREAT X", in which case you will be asked to indicate on a line what disease you think is most likely and how likely you think it is. (This is for my information only and has no effect on the outcome of the game or of the lottery). The computer will inform you of the actual disease, your profit (if any), and your new balance; the game then ends. If you have not completed the ten scheduled games, the next game begins with a new chioce of actual disease.

Alternatively, you may respond to "TEST OR TREAT" with "TEST A", "TEST B", "TEST C", "TEST D", or "TEST V". In this case, the computer deducts $100 from your balance, informs you of the result of the test, and then asks "TEST OR TREAT" once again.

NOTE: You may purchase as many or as few tests as you want; in particular, you may purchase the same test more than once in a given game, either right away or after seeing the results of other tests. Test results are generated independently according to the probabilities in the table, so a repetition of any test will usually but not always give the same answer within the same game.

You will play this game five times just for practive, then five more times for chances in Lottery II. Everyone starts out with ten free chances in Lottery II; at the end of the second five games, you get one additional chance for every $100 in your account (or lose one chance per $100 if your account is neagtive). In other words, every test you call for costs you one chance ($100) in Lottery II, but might improve your channces of winning ten chances ($1000).

Please re-read these instructions and familiarize yourself with the table. You may take as long as you want to think about each move in the game, but you must do all the figuring in hour head; no paper-and-pencil calculations are allowed.

Also note that you start out with the same chance of winning Lottery II as you did in Lottery I; because the twenty participants get fewer "chances" in Lottery II than in Lottery I, each chance is worth that much more.

REFERENCES

REFERENCES

Abelson, Robert P. et al, editors: THEORIES OF COGNITIVE CONSISTENCY: A SOURCEBOOK, Rand McNally, Chicago, 1968

Barnett, G. Octo, John B. Bailleul, and Barbara B. Farquhar: "The Testing of Clinical Judgement - An Experimental Computer-Based Measurement of Sequential Problem-Solving Ability,"in Jacquez, John A. (ed.): COMPUTER DIAGNOSIS AND DIAGNOSTIC METHODS, Charles Thomas, 1972

Baumol, William J.: ECONOMIC THEORY AND OPERATIONS ANALYSIS, Prentice-Hall, 1977

Chan, Sui-Wah: "DOCS: The Doctor Simulation System," Technical paper, Medical Information Systems Theory Project, Michigan State University, 1974

Chan, Sui-Wah, "Toward A Structural Theory of Medical Decision Making," Technical Paper, Medical Information Systems Theory Project, Michigan State University, 1974

Chan, Sui-Wah: "A Structural Analysis of Problem Complexity in Information Processing Behaviors as Related to Human Problem Solving," Proposal to the Information Science Research Program of the National Science Foundation, 1979

Chan, Sui-Wah and Thomas Whalen: "Problem Solving, Decision Making and Computer Based Information Processing: Annotated Bibliography," Technucal Paper, Medical Information Systems Theory Project, Michigan State University, 1978

Edwards, Ward: "Subjective Probability Inferred From Decisions," Psych. Review, vol. 69, #2, p 109-115, 1967

Edwards, Ward: "Conservatism in Human Information Processing," in Kleinmuntz, Benjamin (ed.): FORMAL REPRESENTATION OF HUMAN JUDGEMENT, Wiley, New York, 1968 p.17-52

Edwards, Ward: "N = 1: Diagnosis in Unique Cases," in Jacquez, John A. (ed.): COMPUTER DIAGNOSIS AND DIAGNOSTIC METHODS, Charles Thomas, 1972

Fried, Lisbeth S. and Cameron R. Peterson: "Information Seeking: Optional Versus Fixed Stopping," J. Experimental Psychology 30, No. 3, 525-529, 1969

Glaser, R., D.E. Damrin, and F.M. Gardner: "The Tab Item: A Technique For The Measurement of Proficiency in Diagnostic Problem Solving Tasks," Educational and Psychological Measurement, 14, 283-293, 1954

Gorry, G. Anthony, Jerome P. Kassirer, Alvin Essig, and William B. Schwartz: "Decision Analysis As the Basis for Computer-Aided Management of Acute Renal Failure," Amer. J. Med., 55 475-484, October 1973

Howard, Ronald A. (Guest Editor): IEEE TRANSACTIONS ON SYSTEMS, MAN AND CYBERNETICS, vol SMC-4 #3, September, 1968 (Special issue on decision analysis)

Kleinmuntz, B.: "The Processing of Clinical Information by Man and Machine," in Kleinmuntz, B. (ed.): FORMAL REPRESENTATION OF HUMAN JUDGEMENT, Wiley, New York, 1968, p149-186

McGuire, C. B. and R. Radner (eds.): DECISION AND ORGANIZATION: A VOLUME IN HONOR OF JACOB MARSCHAK, North-Holland, Amsterdam, 1972

Newell, Allen and H. Simon: HUMAN PROBLEM SOLVING, Prentice Hall, 1972

Nichols, Arthur C. and Brian Schott: SIMQ, A BUSINESS SIMULATION GAME FOR DECISION SCIENCE STUDENTS, Kendall/Hunt Publishing Company, Dubuque, 1972

Pau, L. F.: "Sequencing Laboratory Tests for Improved Sequential Diagnosis," MEDINFO 1974, pp. 569-573

Raiffa, Howard: DECISION ANALYSIS, Addison-Wesley, Reading, Mass. 1968

Rimoldi, H.J.A.: "Testing and Analysis of Diagnostic Skills," in Jacquez, John A. (ed.): THE DIAGNOSTIC PROCESS, University of Michigan, 1964

Scandura, Joseph: PROBLEM SOLVING: A STRUCTURAL/PROCESS APPROACH WITH INSTRUCTIONAL IMPLICATIONS, Academic Press, New York, 1977

Schum, David A. and David W. Martin: "Human Processing of Inconclusive Evidence From Multinomial Probability Distributions," Organizational Behavior and Human Performance 3, 353-365, 1968

Schwartz, William B., G. Anthony Gorry, Jerome P. Kassirer, and Alvin Essig: "Decision Analysis and Clinical Judgement," Amer. J. Med., 55, 459-472, October, 1975

Simon, Herbert A.: ADMINISTRATIVE BEHAVIOR, Free Press, 1957

Simon, Herbert A.: THE NEW SCIENCE OF MANAGEMENT DECISION, Prentice-Hall, 1977

Slovic, Paul and Sarah Lichtenstein: "Comparison of Bayesian and Regression Approaches to the Study of Information Processing in Judgement," Organizational Behavior and Human Performance, Vol.6, 649-744, 1971

Sprosty, Paul J.: "The Use of Questions in the Diagnostic Problem Solving Process," in Jacquez, John (ed.): THE DIAGNOSTIC PROCESS, University of Michigan, 1964

Sussman, Gerald Jay: "Electrical Design: A Problem for Artificial Intelligence," MIT Artificial Intelligence Memo 425, June, 1977

Thompson, William J.: "SIMQ Decision Support System," In-house Manual, Georgia State University Department of Quantitative Methods, Atlante, 1978

Vroom, Victor H.: WORK AND MOTIVATION, Wiley, New York, 1964

Von Neumann, John and Oskar Morgenstern: THEORY OF GAMES AND ECONOMIC BEHAVIOR, Wiley, New York, 1947

Wallsten, Thomas S.: "Using Conjoint-Measurement Models to Investigate A Theory About Probabilistic Information Processing," J. Math Psych 14, pp144-185, 1976

Whalen, Thomas and Sui-Wah Chan: "The Decision Analysis Problem As An Experimental Paradigm For The Study of Human Information Processing," Proceedings of the 1979 International Conference on Cybernetics and Society, IEEE, 1979 (in press)