2F

PNL-2862

UC-70

Assessment of Effectiveness of Geologic Isolation Systems

The Feasibility of Computer Interrogation of Experts for WISAP

L. H. Wight
Tera Corporation

May 1980

Prepared for the Office of Nuclear Waste Isolation under its Contract with the U.S. Department of Energy

Pacific Northwest Laboratory
Operated for the U.S. Department of Energy
by Battelle Memorial Institute

PNL-2862

Battelle

PACIFIC NORTHWEST LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
Under Contract DE-AC06-76RLO 1830

Printed in the United States of America
Available from
National Technical Information Service
United States Department of Commerce
5285 Port Royal Road
Springfield, Virginia 22151

Price: Printed Copy $_______*; Microfiche $3.00

*Pages	NTIS Selling Price
001-025	$4.00
026-050	$4.50
051-075	$5.25
076-100	$6.00
101-125	$6.50
126-150	$7.25
151-175	$8.00
176-200	$9.00
201-225	$9.25
226-250	$9.50
251-275	$10.75
276-300	$11.00

PNL-2862
UC-70

Assessment of Effectiveness of Geologic Isolation Systems

THE FEASIBILITY OF COMPUTER INTERROGATION OF EXPERTS FOR WISAP

L. H. Wight
Tera Corporation

May 1980

Prepared for the
Office of Nuclear Waste Isolation
under its Contract with the
U.S. Department of Energy

Pacific Northwest Laboratory
Richland, Washington 99352

ACKNOWLEDGMENT

This research was supported by the Waste Isolation Safety Assessment Program (WISAP) conducted by Pacific Northwest Laboratory. The program was sponsored by the Office of Nuclear Waste Isolation, managed by Battelle Memorial Institute for the Department of Energy under Contract DE-AC06-76RLO-1830. On 1 October 1979, WISAP became the Assessment of Effectiveness of Geologic Isolation Systems (AEGIS) Program and the Waste/Rock Interactions Technology (WRIT) Program. This report was issued by AEGIS.

FOREWORD

Associated with commercial nuclear power production in the United States is the generation of potentially hazardous radioactive waste products. The Department of Energy (DOE), through the National Waste Terminal Storage (NWTS) Program, is seeking to develop nuclear waste isolation systems in geologic formations. These underground waste isolation systems will preclude contact with the biosphere of waste radionuclides in concentrations which are sufficient to cause deleterious impact on humans or their environments. Comprehensive analyses of specific isolation systems are needed to assess the postclosure expectations of the system. Assessment of Effectiveness of Geologic Isolation Systems (AEGIS) program has been established for developing the capability of making those analyses.

Among the analyses required for isolation system evaluation is the detailed assessment of the postclosure performance of nuclear waste repositories in geologic formations. This assessment is concerned with aspects of the nuclear program which previously have not been addressed. The nature of the isolation systems (e.g., involving breach scenarios and transport through the geosphere) and the great length of time for which the wastes must be controlled dictate the development, demonstration, and application of novel assessment capabilities. The assessment methodology must be thorough, flexible, objective, and scientifically defensible. Furthermore, the data utilized must be accurate, documented, reproducible, and based on sound scientific principles.

The current scope of the Assessment of Effectiveness of Geologic Isolation Systems program is limited to long-term, postclosure analyses. It excludes the consideration of processes that are induced by the presence of the wastes, and it excludes the consideration of nuclear waste isolation in media other than geologic formations. The near-field/near-term aspects of geologic repositories are being considered by ONWI/DOE under separate programs. They will be integrated with the AEGIS methodology for the actual site-specific repository safety analyses.

The assessment of repository postclosure safety has two basic components:

- identification and analyses of breach scenarios and the pattern of events and processes causing each breach;
- identification and analyses of the environmental consequences of radionuclide transport and interactions subsequent to a repository breach.

The Scenario Methodology Development task is charged with identifying and analyzing breach scenarios and their associated patterns of events and processes.

The Scenario Methodology Development task is concerned with evaluating the geologic system surrounding an underground repository and describing the phenomena which alone or in concert could perturb the system and possibly casue a loss of repository integrity. Output from the Scenario Methodology Development task will establish the boundary conditions of the geology and hydrology surrounding the repository at the time of an identified breach. These bounding conditions will be used as input for the consequence analysis task, which will employ sophisticated hydrological transport models to evaluate the movement of radionuclides through the groundwater system to the biosphere.

CONTENTS

Page

ACKNOWLEDGEMENT iii
FOREWORD v
FIGURES viii
1.0 INTRODUCTION AND SUMMARY 1.1
2.0 SUBJECTIVE PROBABILITIES AND COMPUTER ASSESSMENT 2.1
2.1 Traditional Assessment Techniques 2.2
2.2 Computer-Aided Assessment 2.4
2.3 Existing Assessment Programs 2.7
3.0 STRUCTURE AND DECOMPOSITION 3.1
3.1 Recommendations 3.1
3.2 Existing Decomposition Computer Codes 3.2
3.3 Decomposition Methodology for WISAP 3.3
3.4 Decomposition: An Illustration 3.17
4.0 SCALING VARIABLES 4.1
4.1 Recommendations 4.1
4.2 Overview 4.1
4.3 Nature of Probability and Uncertain Judgment 4.6
4.4 Clinical Aspects of Scaling and Empirical Biases 4.8
4.5 Approaches to Scaling 4.14
4.6 Scaling Rare Events 4.17
4.7 Scaling: An Illustration 4.18
5.0 SYNTHESIS OF EXPERT OPINION 5.1
5.1 Recommendations 5.1
5.2 Existing Techniques for Synthesis 5.1
5.3 Synthesis Techniques for WISAP 5.10
5.4 Synthesis: An Illustration 5.21
6.0 REFERENCES 6.1
APPENDIX: DERIVATION OF OPTIMAL WEIGHTS IN SYNTHESIS A.1

FIGURES

Page

1.1 Overview of Encoding Procedure 1.2

1.2 General Assessment Code Logic 1.3

3.1 Decomposition of Logical Hierarchy 3.4

3.2 Decomposition of Hydrofracturing 3.5

3.3 Strength of Relationship x → y 3.9

3.4 Logic Hierarchy Branch Implied by Search of International Matrix 3.16

4.1 Scaling of Main Variables 4.2

4.2 Scaling of Subvariables 4.3

4.3 Gamblers' Fallacy Bias in Subjective Probability 4.10

4.4 Fraction of True Values Lying in Respective Regions of Respondent's Assessed Probability Distributions, Compared with Nominal Frequencies of 1, 24, 25, 25, 24, and 1% 4.11

4.5 Conservatism Bias in Subjective Probability Assessment. . . . 4.12

5.1 Idealized Decrease in Group Predictive Error as a Function of Group Size, Typical of Empirically Observed Errors 5.3

5.2 Reproducibility of Group Results as a Function of Group Size 5.4

5.3 Proportion of Respondents Shifting Their Positions as a Function of Distance from the Mode 5.5

5.4 Distribution of Estimates Tend to Lognormality Over Groups of Experts 5.7

5.5 Average Predictive Error for Group Estimates as Functions of Self Rating and Dispersion of Estimates Among Group Members 5.9

5.6 Synthesis 5.11

5.7 Individual and Synthesized Assessments 5.22

1.0 INTRODUCTION AND SUMMARY

Simulation of the response of a waste repository to events that could initiate a fault tree to breach and failure is currently a keystone to the Battelle Waste Isolation Safety Assessment Program (WISAP). The repository simulation, which is part of the Disruptive Event Analysis Task, models the repository for its entire design life, one million years.

This is clearly a challenging calculation, requiring input unlike any other response analysis (nuclear power plants, LNG storage terminals, etc.) by virtue of the long design life of the facility. If selecting a design earthquake is difficult for a 50 year design life power plant, what technology will provide design criteria for a million year design life?

Answers to questions like this can, to some extent, be based on data, but always require some subjective judgments. The subjectivity, which is sometimes driven by inadequate or incomplete data, or by a lack of understanding of the physical process, is therefore a crucial ingredient in an analysis of initiating events. Because of the variety of possible initiating events (glaciation, man-caused disruption, volcanism, etc.), many expert opinions will be solicited as input.

The complexity of the simulation, the variety of experts involved, and the volume of applicable data all suggest that there may be a more direct, economical method to solicit the expert opinion. This report addresses the feasibility of such a system.

In Section 2, we present background information that demonstrates the advantages of a computer interrogation system over conventional interrogation and assessment techniques. In the subsequent three sections we thoroughly review the three elements—structure and decomposition, scaling, and synthesis—that are basic to any interrogation and assessment technique. Figures 1–1 and 1–2 schematize the interrelationship between these three fundamental elements and, therefore, serves as a useful guide to these three sections. Each of these three sections begins with our recommended approach to the particular element and ends with an illustration of representative dialogue.

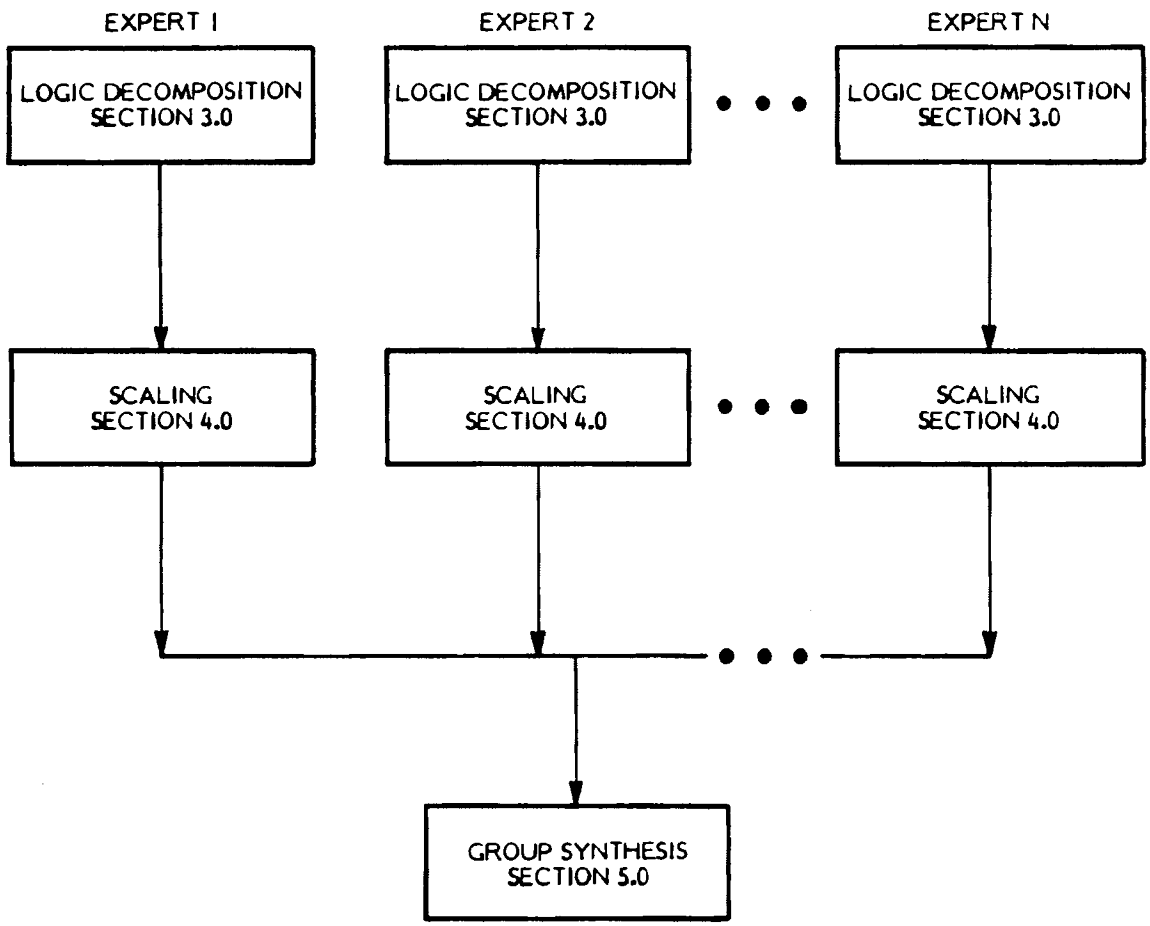


FIGURE 1-1

OVERVIEW OF ENCODING PROCEDURE

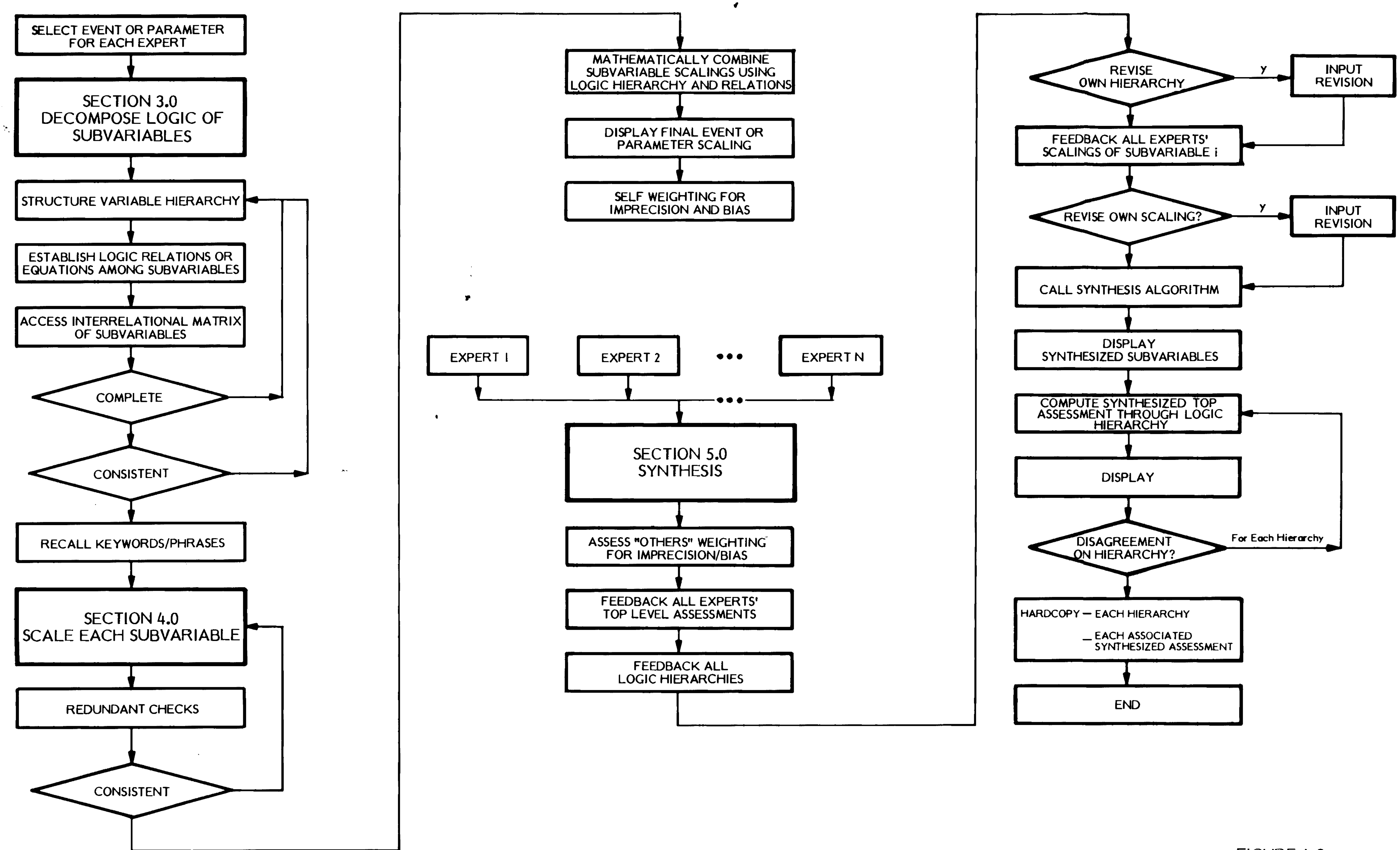


FIGURE 1-2

GENERAL ASSESSMENT CODE LOGIC

Briefly, we consider it not only feasible, but also highly desirable, to use an interactive computer system, perhaps linked with the Disruptive Event Analysis computer code, to interrograte and assess selected experts. Specific advantages of such a computer system include:

- Detailed, systematic interrogation
- Low cost interrogation
- Efficient analysis
- Significant feedback and iteration capabilities
- Consistency and logical response evaluation
- Unlimited data storage capabilities

Because it operates so efficiently, the system would be effective not only for developing input for the Disruptive Event Analysis, but also as a vehicle for public interest group participation during the licensing process.

Our estimates indicate that the computer-based techniques are cost-effective, as well as versatile and capable. We believe that assembling and applying the necessary interrogation software package would be less costly than developing and applying alternative (questionnaire, Delphi, frequent workshops, etc.) subjective probability assessment procedures: the entire programming effort could range from $100,000 to $200,000, depending on the detailed applicability of existing software packages. The existing computer hardware at the PNL would be more than adequate to accommodate our envisioned software.

2.0 SUBJECTIVE PROBABILITIES AND COMPUTER ASSESSMENT

Probabilities and probability distributions of many events and parameters associated with waste isolation are difficult to quantify. Some of these can be dismissed as so improbable as to be effectively impossible (e.g., glaciation in the Southwest), while some are probable enough to be considered certain (e.g., climate changes). Between these extremes, though, are variables for which not-so-easily-dismissed probabilities are required, such as evaluations for which normal empirical relations or theoretical models are difficult to develop. Quantification of these variables depends on expert judgment.

The current plan for developing input information for the WISAP Disruptive Event Analyses is to solicit the advice and judgment of selected experts, so the most important question is: How can we effectively elicit that advice and judgment? How can we avoid those arbitrary numbers which are too easily specified, and develop an assessment procedure that efficiently results in quantifications that reflect carefully evaluated judgment? There are a number of ways to do so, including workshops, questionnaires, personal reflection, interviews and Delphi techniques. Each of these has advantages and disadvantages. Since the early 1950's a pool of evaluated experience has been developing on subjective probability assessment. Work has been done on scaling techniques (e.g., Winkler, 1967; Stael von Holstein, 1970), psychological biases (e.g., Edwards, 1968; Tversky, 1974), group effects (e.g., Dalkey, 1970; Morris, 1974), and other facets of the problem. This work provides a background against which assessment schemes can be evaluated.

An assessment scheme should provide a quantification of probabilities or probability distributions, an indication of the precision of those quantifications, and an explicit description of the logical reasoning behind the numbers. This allows the numbers to be evaluated and compared by experts and to be aggregated across a group of experts. The logical reasoning behind the numbers is especially important because the probabilities of major interest are beyond mere intuition. Indeed, studies by Lichtenstein and Newman (1967) and Selvidge (1972), among others, have shown that people do not even have a vocabulary for describing the probability of truly rare events. Therefore, to assess such probabilities, the variables must be broken down into more manageable parts and recombined according to some set of logical relationships. To judge the numbers, one must also judge the logic through which they are calculated.

2.1 TRADITIONAL ASSESSMENT TECHNIQUES

For purposes of discussion, traditional assessment techniques might be divided into three groups:

- Questionnaires and personal interrogation
- Interviews and clinical procedures
- Workshops, group meetings, and Delphi techniques

These alternatives should be compared according to their cost, and the faith one has in their resulting numbers.

Questionnaires are the least expensive way to solicit and assess expert opinion. Well-written questionnaires consist of structured sets of redundant questions that force the person answering them to reflect on his opinion--that is, to interrogate himself. Major advantages of questionnaires are that a large number of experts can be interviewed and a broad range of topics can be covered with limited manpower and cost. An example of a carefully designed assessment performed using questionnaires is the resource appraisal carried out by the Oil and Gas Branch of the U.S.G.S., and published as U.S.G.S. Circular 725 (1975).

One problem with questionnaires, though, is that they are "static": the person answering them has no feedback to his answers, and no one pointing out logical inconsistencies. He is not given suggestions or asked why particular aspects of a problem fail to appear in his analysis. A related problem with questionnaires is that the depth of interrogation on particular issues is necessarily limited. It is difficult to design a sequence of questions that do not depend on the respondents' answers to develop detailed information. Second-round questionnaires are sometimes used to feedback the distribution of group opinion and see if answers tend toward a consensus, but in general, questionnaires are better suited to assessing broad distributions of opinion or, as in the U.S.G.S. case, opinions from many experts on different topics which combine to give an answer. Because there is little or no opportunity for feedback, precise wordings of questions often have more significant affect on questionnaire results than on the results of other assessment techniques.

Interviews require substantially more investment than questionnaires, but yield higher quality results. Yet, interviewing is a clinical skill, requiring considerable experience: the quality and consistency of assessments depend greatly on the skill of the interviewer. Even so, if higher cost and more manpower were not considerations, direct interviews would probably be a favored technique, because they provide extensive feedback and questioning of the logic behind assessments, and of course allow the interviewer to tailor the sequence and tone of questions to a particular expert. Interviews carry the disadvantage of not easily allowing reinterrogation after the response of an entire group has been obtained.

Delphi techniques—which vary widely—are essentially interview procedures in which anonymous feedback of group assessments is provided to the experts and second or even third rounds of questioning undertaken. The set of techniques included under various definitions of Delphi is so broad that it includes nearly all assessment procedures and, therefore, is not a technique in itself. The only unifying feature is anonymous feedback. Group assessments which are the goal of most Delphi studies are obtained by consensus.

Workshops and meetings are the other extreme of assessment. They are costly because a number of experts must be brought together in one place, but they allow open (and often heated) discussion of the bases for assessing probabilities. In workshops, logical structures are explicit, feedback is considerable, and new approaches to a problem are sometimes uncovered. On the other hand, careful reflection is often sacrificed to debate, and strong or persuasive personalities often have major influence on final results. Reconciling conflicting opinion or eliciting the opinion of reticent panel members is often difficult, and work on "brain-storming" during the late 1950's and early 60's has indicated that a group working together may be no more effective than an equally large group of people working individually (McGrath, 1966).

2.2 COMPUTER-AIDED ASSESSMENT

In order for computer-aided assessment to be useful it must meet two criteria: it must be feasible to develop within a reasonable amount of time for a reasonable cost; and it either must provide better assessments than traditional techniques (e.g., interviews) or equally good assessments more easily. This section considers the advantages of computer-aided assessment. The feasibility of computer-aided assessment is discussed in detail in Sections 3 through 5 of this report.

The quality of computer-aided assessment can be judged only by comparing it to the quality of assessment resulting from interviews. The quality of computer assessment depends on the user being somewhat familiar with the goals and general procedure of assessment, and on the "interactibility" of the computer code. Therefore, in computer assessments, the subject must be made familiar with the goals, required precisions, and procedures of probability assessment. This can be accomplished in several ways: someone can tell the subject about the goals, etc.; an informative introductory text can be written into a code; or a briefing session can precede use of the code. Because the familiarization is so important, alternative procedures might be experimented with.

The advantages unique to computer-aided assessment are cost and speed, rapid consistency and sensitivity analysis, redundancy and feedback, and data storage. Data storage could prove of significant importance, particularly in a long program of research like WISAP, since previous assessments of expert opinion and the logic underlying those assessments must be reviewed at various stages of the program. Each of these advantages is elaborated upon below.

Cost and Speed of Assessment

The time required to quantitatively assess subjective probabilities can be much less than that required to verbally describe degrees of belief. This was reported in the earliest applications of subjective probability to industrial problems (e.g., Grayson, 1960). Many of the techniques of assessment, though, attempt to back-figure probabilities implied by choices among uncertain events and orderings. The computer can quickly do the calculations involved in these back-figurings so more questions can be asked in a given length of time, convoluted questions can be asked, and a greater number of redundancies can be introduced. Using computer-aided assessment, the user can proceed at his own pace, and stop when he is satisfied. Construction of logic diagrams is entirely automatic, and can be rapidly checked for consistency and completeness using techniques described in Section 3. Hard copies of scalings and logic diagrams are directly provided by the system. Although computer-aided assessment is only an additional tool to be used in obtaining input information and would not replace regular meetings or workshops, the capability of remote use offers extra freedom. Once initiated, the expert can work in his own office, and obtain a set of assessments required for new variables without having to spend time traveling.

Consistency and Sensitivity Analyses

Central to assessment is the development of the logical basis for quantification, testing it for consistency and completeness (i.e., removing any hidden logic), and varying the component probabilities to test sensitivity. All of these tasks are computationally intensive. Testing the logical structure is an inductive problem; it cannot be done automatically. However, if computation limits are not a problem--as they are when done by hand--then the interrelational properties of the entire set of identified variables can be explored using interaction matrices, conditional probabilities and correlations, network diagrams, and other techniques for summarizing interaction. These mathematical abstractions of the logic can be treated computationally to infer structures in the originally specified logic that are either inconsistent or incomplete. This provides redundant ways of checking logical structures which are not possible in an interview.

Sensitivity analysis in interviews is, in the same way, constrained by computational needs. Variables for geological modeling are often related through equations or correlations, and to propagate parametric uncertainty through three or more nested equations by hand is difficult: interview assessment cannot take advantage of the immediate feedback of output sensitivity to imprecision in variable scalings in the way that computer assessment can. This feedback of sensitivity calculations at potentially all levels of a logic hierarchy provides worthwhile information to the user and influences his assessments.

Redundancy and Feedback of Individual Scalings

Redundancy and feedback are important not only for developing the logical structure interrelating variables, but also for assessing probabilities or probability distributions over individual variables. Scalings are generated by asking questions in a number of ways, backfiguring the implied probabilities, and presenting the expert with conflicts that should be resolved. This redundancy and feedback are important parts of any scaling (Section 4), and are used in interviews as well as computer-aided assessments. The number of redundant checks one can make in an interview, however, is limited for several reasons. For one, if redundant sets of questions are to be valuable, they must not be so obvious that, recognizing the redundancy, an expert tailors his answer to appear consistent. This problem can be overcome with complicated sets of parallel questions written to camouflage redundancy. However, convoluted computations are then often required to develop implied probabilities. The available scope of redundant questioning is broadened by the use of computer processing.

Data Storage

The capacity for large data storage and rapid retrieval is unique to computer-aided assessment, and introduces operational capabilities which would otherwise not be available. When an expert uses the program to assess the probability or distribution for some variable, he identifies a set of subvariables and their logical relation to the upper event on parameters, a set of equations where appropriate, a matrix of interrelational properties among subvariables, and finally a probability scaling over each subvariable. With computer-aided assessment all of this information can be stored and quickly retrieved. Therefore, if a probability estimate is required for some future analysis, not only may the numbers themselves be retrieved, but also the entire reasoning that led to them and the assessment of variables on which the event or parameter depends.

The ability to retrieve this complete set of information provides new opportunities. First, it provides a basis for re-evaluating the numbers when they are used. Second, if new information becomes available at some later date the earlier assessment can be updated rather than a whole new assessment being required. Thus, if a new way for some event to occur is uncovered, it can be included in an original event hierarchy. If the new sequence can be decomposed into subvariables already scaled, implied modifications of probabilities or distributions can be directly calculated at the time. Third, if another expert disagrees with an assessment, comparisons can be made at each step of the reasoning that led to the numbers. Thus, the actual basis for disagreement is sharpened.

2.3 EXISTING ASSESSMENT PROGRAMS

The review of assessment needs for WISAP (Section 2) and an inventory of existing code indicate that direct transfers are impractical. Existing codes have, in general, been written for specialized tasks, somewhat different from those involved in the current analysis of needs. Nevertheless, there is much to be learned from existing codes and their applications, and in limited ways the logic of existing codes could be expanded to form the basis of service subroutines in a computer-aided assessment for WISAP.

Several programs exist for direct scaling of the probability of an event or probability distribution over a parameter (e.g., Schlaifer, 1964; Sickerman, 1975). These programs are more limited than that required for WISAP, but they have been successfully used for simple assessments and seem to yield quantifications not very different from interview assessments. Most existing programs are limited, however, in that they do not address the logical reasoning behind an assessment (Section 3), and therefore are inferior to direct interviews in which the underlying reasoning is extensively discussed. For WISAP, the structuring of event logic would have to be incorporated in the assessment program. Lapp and Powers (1977), in their work on computer-aided fault tree construction, report that computer construction is superior to manual construction, because fewer mistakes are made and constructions proceed more quickly. Computer-aided assessments can be equal in quality to interview assessments as long as the program forces the user into careful reflection, explicit decomposition of his logical reasoning, and internal consistency.

Existing codes related to probability assessment can be grouped in three classes: scaling codes for assessing event probabilities or probability distributions over parameters, which are found primarily in decision analysis; value function codes for assessing weighting and utility function, found in policy analysis and to some extent in decision theory work; and logic or fault tree construction codes, found primarily in reliability and safety analysis. In addition there are a number of interactive design codes in engineering and architecture that incorporate system-identification options, but these are not directly related to present needs.

Scaling codes have been in use for at least a decade. The earliest of these were developed at the Harvard Business School (e.g., Schlaifer, 1969), and were developed to aid in single assessments of "routine" events or parameters (i.e., not rare events, which would require logical structuring prior to assessment). In general these codes are written as assessment aids for Bayesian decision theory, and thus assume familiarity with subjective probability theory and typical formats of assessment. Within Bayesian theory, probabilities are taken as reflection of an individual's willingness to act on a belief, so these assessment codes are based on choices among idealized betting options (commonly called "lotteries"). Lotteries are structured to incorporate random events with which the user has familiarity—coins, dice, wheels-of-fortune. In principle, a user may not be able to directly associate numbers with his degrees of belief, but should be able to order lotteries on an intuitive level. More recent work has led to the development of assessment codes for multi-variable events and parameters, and limited stochastic variables (e.g., time series) typically modeled as multi-variate processes (e.g., Sicherman, 1975).

Value and utility function assessment codes have enjoyed broad popularity in recent years, particularly for participatory planning and multi-attributed problems. Existing codes encompass a spectrum of levels of sophistication. The goal of value function codes is a quantified objective function defined over the multi-attributed outcomes of decisions, usually public policy decisions. The objective function requires quantification of marginal rates of benefit or cost for each attribute of an outcome, and marginal rates of substitution among outcome attributes. In the special case of utility function assessment, measures of risk aversion (or proneness) must also be quantified. Many such codes exist, but due to recent public exposure, the most widely known is possibly that of Hammond and Adelman (1976) who have used their code to explore public sentiment in the case of the Denver Police Department bullet decision. Among other value or utility assessment codes are those of Sheridan (1975), Meyers (Richard Meyers, Harvard University, Personal Communication), and Sickerman (1975).

Most value and utility function assessment codes impose a structure on the assessments. For example, the objective function must be a weighted sum, or a weighted sum including cross-products (multi-nominal or so-called multiplicative forms). Although many existing codes are supposed to foster concensus and obtain group objective functions, in fact any synthesis of opinion is mostly exogenous to the code.

Logic and fault tree construction codes have been developed to assess inter-relational properties among variables in a complicated system and construct logic diagrams or hierarchies. These codes are still being developed. Typical of these codes are those of Lapp and Powers (1977), Fussel (1972), and Taylor (1973). These codes are not aimed at assessing probabilities or probability distributions for variables within the system, but rather at the structure of the variables. Numerical assignments are made subsequent to the structuring.

For present purposes, existing logic and fault tree construction codes are overly specialized. In analyzing systems reliability, 1,000 or more variables are not uncommon. These variables are by necessity interrelated through simple logical structures (e.g., "and/or" gates) and are usually considered to be statistically independent. Codes for handling multi-valued logic, in which components are not single zero-one variables, are only now in development. As discussed in Section 3, present purposes do not require the expansiveness of system structure analyzed by large fault trees. Therefore, an assessment code for WISAP can be and should be more detailed in the types of relationships and variables it has the capability to handle.

3.0 STRUCTURE AND DECOMPOSITION

Whether the probability assessment is done intuitively or with computer assistance, the variables must be broken into easily assessed or logically isolated components. This decomposition will be an important part of the proposed program, and therefore deserves particular attention. Decomposition involves identifying component variables, and specifying functional relations among them. The set of functional relations will then be used to recombine probabilities of more easily assessed components into probabilities or probability distribution for the variable of interest.

3.1 RECOMMENDATIONS

The most promising way to identify the structure of subvariables is a combined approach using logic hierarchies and interrelational matrices. The logic hierarchy is simply a generalized form of "event tree" in which functional and statistical relations as well as simple event chains describe the sequencing of variables. In its simplest form, hierarchical analysis is well known and widely discussed in the literatures of reliability, decision theory, and operations research. For application to probability assessment, however, hierarchical analysis will have to be broadened to incorporate functional relations among variables, and multivalued (or continuous) logic. Given the limited size of the hierarchies involved, this increase in complexity should not provide programming or computational difficulties.

To complement the logic hierarchy and to provide a vehicle for checking logical consistency and completeness, an "interrelational (IR) matrix" will be developed. The IR matrix is a square matrix of dimension equal to the number of subvariables in the hierarchy. Elements of the matrix encode the strength and direction of functional or statistical interrelation among subvariables, along a simple ordinal scale $(0, \pm 1, \pm 2)$. The IR matrix will not be directly used for computation, but for exploring the logical structure of the hierarchy and uncovering implied relationships not explicitly identified in the assessment.

Calculations of top-variable probabilities or probability distributions will be performed directly from the logic hierarchy. Gates in the hierarchy are specified either as functional relations (i.e., equations), as causal chains (i.e., zero-one event sequences), or as statistical correlations (i.e., conditional probabilities). Calculations and subvariable scalings will be structured so that each level in the hierarchy involves redundancy. This will provide an additional check on numerical values resulting from the analysis.

Currently available computer programs for fault-tree construction, cross-impact (including input-output) analyses, and simulation modeling provide only a starting point for the development of a hierarchy assessor, because the demands of the present problem are somewhat different from, and in ways more logically complicated than, the problems many available programs deal with.

Existing programs are not, therefore, directly transferable. Although relationships among variables in the present case are more involved than, say, in fault-tree construction, the number of variables and gates in the present case is limited; the difficult programming and computation problems associated with extensive logic branches do not have to be dealt with in developing simple interrogation programs.

3.2 EXISTING DECOMPOSITION COMPUTER CODES

Computer-aided construction of cross-impact and input-output matrices has been developed by Dalky (1975), Turoff (1975), and Gordon, et al (1970), among others. Much of the work in these efforts has been directed at reconciling inconsistencies in conditioned probabilities and at simple simulation studies. The use of interaction diagrams ("diagraphs") for constructing logic trees is considered by Lapp and Powers (1977). Of course, much work has been done in simulation modeling on relating interrelation matrices to "system" structure.

Although the logic hierarchy and interrelational matrix proposed here are related to fault-tree analysis and to input-output or cross-impact analysis, the required capabilities of the proposed methods are sufficiently unique that direct transfer of existing programs is not possible. Again, however, existing programs offer a foundation on which to build, as well as methodological insight.

Computer-aided fault-tree construction has been an intensively worked problem for at least five years. Although problems associated with automatic construction are not entirely solved, codes now exist for constructing fairly complicated fault-trees (e.g., Lapp and Powers, 1977, have constructed a 1,000-gate tree this way). Other fault-tree construction programs are discussed by Fussell (1972). These codes indicate that computer-aided construction of logic hierarchies is possible, that it is much quicker and more convenient than hand-construction, and that fewer mistakes and omissions are made than in hand constructed trees.

3.3 DECOMPOSITION METHODOLOGY FOR WISAP

Figure 3-1 presents a flowchart representation of the recommended decomposition portion of the code. This figure will serve as a focal point for the rest of the section, it can be folded out to provide easy reference to the rest of this section.

In attempting to assess any uncertain event, parameter, or variable, one always begins by decomposition, by dividing a complex or difficult assessment into somewhat isolated pieces each of which may be more easily assessed. Then, having obtained at least rough numbers for the components, logical relationships of the decomposition are used to deduce numbers for the more complicated event, parameter, or variable. A decompositional approach of one form or another is used even when assessments are made intuitively, based only on internal reflection. Thus a central part of computer-aided assessment must be directed at uncovering the logical structure of the event or parameter to be assessed, and the causal or statistical interrelationships among the variables contained in that logical structure.

The decomposition of most events or parameters in repository modeling is neither complicated nor extensive, as Figure 3-2 indicates. Direct intuitive

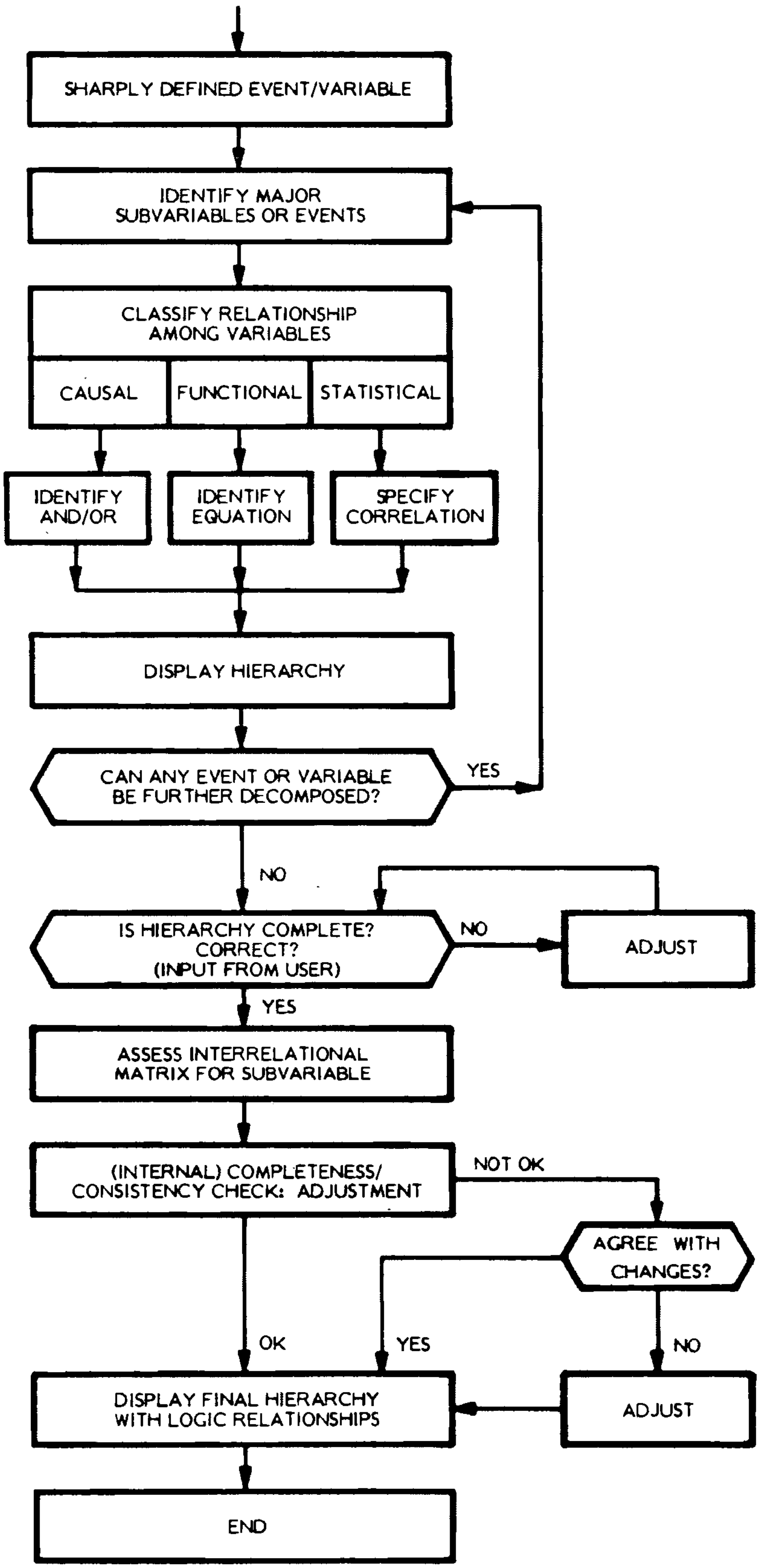


FIGURE 3-1

DECOMPOSITION OF LOGICAL HIERARCHY

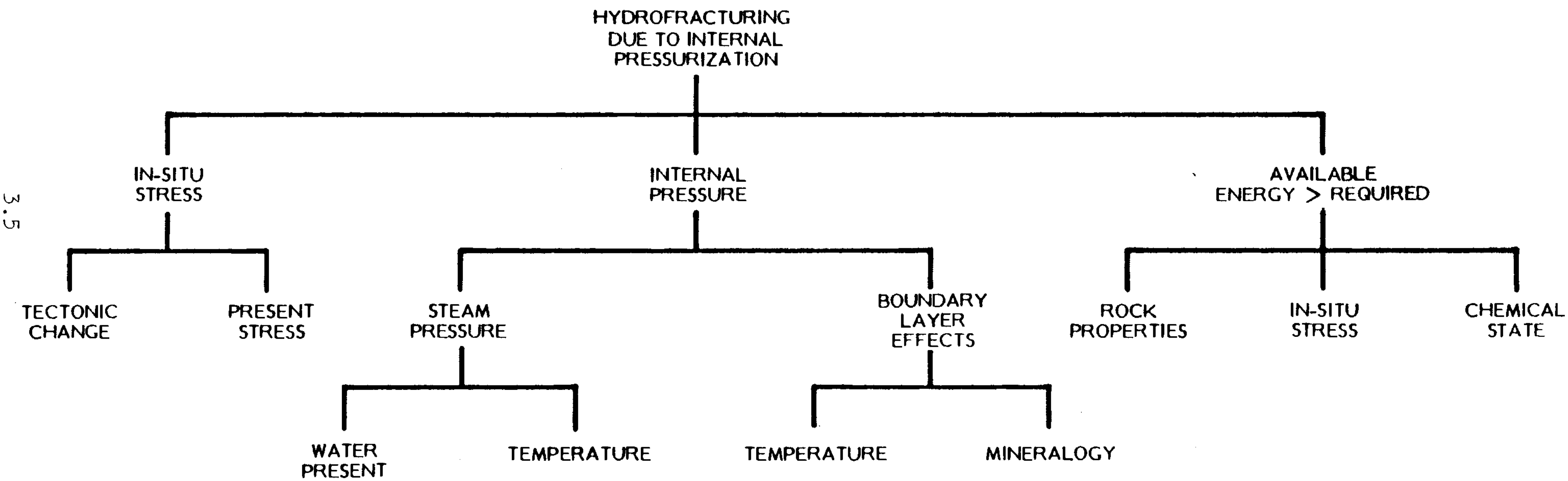


FIGURE 3-2

DECOMPOSITION OF HYDROFRACTURING

assessment, on the other hand, of the probability of hydrofracturing is at least difficult and maybe impossible. Furthermore, the resulting numbers are either very imprecise or not believable. To sharpen the estimate, one procedes by asking what hydrofracturing depends on, or what is correlated to it. Clearly, hydrofracturing depends on the minimum in-situ stress, and the internal pressures. If fractures are to propagate, energy availability is also important. Thus, the overall event can be decomposed into components, and in a similar way the components can be decomposed into subcomponents. Finally the decomposition leads to events that are isolated enough or predictable enough to be more easily assessed. Here, for example, the rock-temperature at a specified time after burial can be estimated within a reasonable range. Combining these estimates through the structure of the decomposition leads to an estimate of the probability of hydrofracturing which is more precise than a direct assessment would be.

Obviously, the decomposition of complicated events or processes into simpler parts is an underlying theme of all engineering. Fault and event tree analysis, decision theory, simulation modeling, analysis of variance, and many other techniques are all variants of decomposition analysis. The reason is clear: it is the way people solve problems intuitively, and the techniques force a logical orderliness on the analysis which might otherwise be wanting. The advantages of explicit techniques like computer-aided decomposition are that they foster logical consistency, mitigate omissions, and provide feedback so that more complete analyses can be performed. Because they maintain the logical structure of a decomposition they can be used to indicate structural relations which may not be intended by the user, but which are implied by the relation he explicitly identifies. The further advantage of computer-aided techniques is that redundant information can be quickly gathered for consistency checks.

Hydrofracturing illustrates the extent of decomposition required for most geological and repository modeling: three or four levels of subevents, and two or three subvariables at each node. The problem is not computationally extensive, which means that a variety of straightforward interrogation techniques can be used. With a small number of nodes and branches, the logic structure can be highly over-determined by the parallel use of differing decomposition techniques.

This section outlines the requirements of a computer-aided decomposition procedure, and develops a set of techniques to satisfy those requirements.

Requirements of Decomposition Analysis

The decomposition program must have the capability to:

- Efficiently handle simple hierarchical logic structures and graphically display them to the user
- Incorporate both functional relations (i.e., mathematical equations) and statistical relations (i.e., correlation matrices) among variables
- Deduce relations among variables implied by but not explicitly stated in assessments, and test for logical consistency
- Treat multivalued logic (i.e., include event or parameters magnitudes as well as dichotomous occurrence/non-occurrence).

The central part of the decomposition program will be the development of hierarchy diagrams much like fault or event trees. The program must be written so that events and parameters are keyed to variable names, and stored as English sentences. A branching sequence of questions will be presented to the user, which will ask for a specification of the names of subevents and parameters for a specified top event, and the logical relation of those subevents or parameters to the top event (e.g., and/or relation, functional equation, statistical correlations, etc.).

The questioning will continue for each subevent until the user answers either that the subevent cannot be further divided or that it is isolated sufficiently to be assessed. A graphical display showing the logic diagram is developed, from the listings of subevents and their functional relations upward through the hierarchy, as illustrated in Figure 3-2. This diagram, presented on CRT display, can be altered with a light pen by the user to rearrange elements and relationships not fitting his ideas.

In decomposing an event or parameter for probability assessment, the relationships which join subvariables and lead to the event or parameter of interest can be more complicated than those usually allowed in fault or event tree analyses. To be realistic, the assessment program must allow for this complexity. In particular, the program must allow for functional (equations), causal (event chains), and statistical (correlation) relationships; it must allow for subvariables that act jointly and subvariables that act independently; and it must allow for interdependencies among subvariables and subvariables common to more than one branch.

The logic hierarchy developed this way will be similar to, but distinct from a fault tree. The hierarchy will be much less complicated than common fault trees and therefore need not be constrained by the normal limitations of fault trees (e.g., 0/1 variables, direct causal linkage, etc.). Variables may be multivalued and relationships among variables may be causal or statistical.

The structuring program must also be capable of testing consistency and searching for implicit relationships. In many cases the logic hierarchy will be too simple to require complicated checks for consistency, but even moderately large hierarchies may tax the user's ability to recognize inconsistencies merely by inspecting of the diagram--particularly when there are dependencies among lower level variables or common variables entering at several places in the hierarchy.

Precise scaling of conditional probabilities will be performed for upward sequences of events or parameters. However, precise scalings, at least in the structuring stage, will not be possible "horizontally" across lower level variables. Nevertheless, information on dependencies among these variables will be required. So in the structuring stage, an interrelational matrix will be assessed for the entire set of lower level variables, as schematized in Figure 3-3. This matrix

X ↓ \ Y →	1	2	3	4	5	6	7	8	9
1. IN-SITU σ_3		0	0	0	0	-1	0	-1	-1
2. INTERNAL PRESSURE	0		0	0	-2	0	+1	0	+1
3. PRESENT σ_3	X	0		0	0	0	0	-1	0
4. TECTONIC $\Delta\sigma_H$	X	+1	0		0	-1	0	-1	0
5. STEAM GENERATED	0	X	0	0		0	-1	0	0
6. WATER PRESENT	0	X	0	0	X		-1	+1	0
7. TEMPERATURE	+2	X	0	0	X	+1		X	+1
9. ENERGY AVAILABLE	0	0	0	0	+1	0	+1	0	

"X" = LINKED RELATION IN HIERARCHY

"±2" = CAUSAL OR FUNCTIONAL RELATION

"±1" = MODEST CORRELATION

"0" = INDEPENDENT

FIGURE 3-3

STRENGTH OF RELATIONSHIP X ⟶ Y

will scale the "strength and direction" of interdependence to a simple ordinal measure--perhaps -2, -1, 0, +1, +2, where "$\pm$2" indicates high positive or negative correlation or functional dependence, "$\pm$1" indicates modest positive or negative correlation or dependence, and "0" indicates independence. The interrelational matrix generated this way will be used with a simple graph theory approach and the original hierarchy to search for implied relationships and inconsistencies. This is considered in more detail later in this report.

In sum, an assessment program must be realistic. To be realistic the decomposition analysis must allow for a variety of relational forms through which subvariables lead to events or parameters, and it must account for the interdependencies, functional or statistical, among subvariables. This would not be computationally possible where the logic hierarchy is not limited in size.

Algorithms for Decomposition

The basic algorithms for eliciting logic structures and calculating probability distributions will be quite similar to those now used for computer assisted fault tree construction (Fussell, 1975; Lapp and Powers, 1977) and cross-impact analysis (Dalkey, 1975; Turoff, 1975), and will be based on concepts, approaches, and special techniques used in current codes. Of course, the structuring code envisioned must be more comprehensive than fault tree and cross-impact codes, since it must admit causal and statistical relations, interdependencies among subvariables, and multi-level variables. Essentially all fault tree and most cross-impact codes treat larger and more complex sets of variables, and of necessity cannot be designed with the capabilities proposed here.

Enumeration of Subvariables and Their Relationships

Because the extent of the logic trees is limited, the elicitation itself can proceed by direct enumeration of subvariables. The user will be required to provide three types of information: an enumeration of subvariables, the gate type linking subvariables, and the form of predictive relation leading from subvariable to top event.

For each set of subvariables, a two-way categorization of the logic gate connecting them to the top event will be possible. These involve determining: (a) the functional form, and (b) whether subevents act jointly or individually in leading to the top event. The program will require the user to select one from among each of two sets:

Functional Form

Functional (F) — Meaning subevents are related to the top event through a functional equation. Subevents may be either continuous or discrete variables. If "functional" is elected, the user will be required to specify the form of the relationship as an equation, which may be empirical, theoretical, or simply an intuitive heuristic of the user.*

Causal (C) — Meaning that the occurrence or non-occurrence of the top event is a result of the occurrence or immediate non-occurrence of subevents. Causal relations are those commonly found in fault trees, and treat only discrete events. In fact "causal" relations are a special case of "functional" relation, but the distinction facilitates assessments, as people tend to think of the two somewhat differently.

Statistical (S) — Meaning that top events relate to subevents through statistical correlation. Events and subevents are either not causally related, or the nature of the relationship is imperfectly known and therefore summarized by conditional probabilities.**

---

* The relation of "water present" and "temperature" to internal steam pressure is an example. Water present is a dichotomous variable; temperature is continuous. Internal steam pressure is related to these variables through the saturated steam tables. The program should have the flexibility to accept the relationships either in equation form, or graphically via the CRT and light pen in order to allow simple use of tabulated data.

** For example, the density of fractures (i.e., number per rock volume) caused by hydro fracturing can only be related in an imprecise way to rock properties and energy sources. This relation might best be described through correlation. Usually, statistical relations will not have a directional nature, in that correlations and conditional probabilities must satisfy common probabilistic properties.

Interaction

Jointly (J) — Meaning that the level of each subevent is important in predicting the top event. The effect on the top event may be of an additive type in which each subevent contributes independently; for example, the energy requirement for hydrofracturing depends somewhat independently on stress-state and fracture surface energy (additively). Or the effect can be interdependent; for example energy requirement for hydrofracturing is interdependent with stress-state and rock deformation properties. For "causal" relationships, "joint" interaction would be equivalent to an "and" gate in fault tree analysis.

Independent (I) — Meaning that the top event depends only on the independent occurrence or extreme level of any of the subevents. For example, "internal pressure" depends either on steam pressure or on boundary layer effects, whichever is more severe. Again, for "causal" relations, "independent" interaction would be equivalent to an "or" gate in fault-tree analysis. "Independent" interaction with a "functional" relationship would take the set of functional equations, evaluate it for each subevent, and the effect select the maximum or minimum.

Thus, for any one gate in the tree, one of six possible combinations is possible: F/J, F/I, C/J, C/I, S/J, S/I. The sequence of further questions and the form of the computations with the tree depend on which combination is selected. Therefore at each gate the program branches to a sequence of questions aimed at eliciting the precise form of the functional relation and the form of aggregation (for which "joint" variables may already be contained in the functional equation).

Calculation of Top Events

When the logic hierarchy is formed and gates are identified by explicit relationships, the scaling subroutine is called to assess probabilities or probability distributions over lowest level subevents. When scaling is completed, the program procedes to a calculation phase in which subevent scalings are propagated through the logic hierarchy to arrive at a scaling for the top event. In the process, scalings for intermediate level events or variables are stored for later redundancy checks.

The simplest calculation algorithm is that associated with C/J and C/I gates. These are simple Boolian operations on the zero-one variables, and have extensive precedence in reliability studies (e.g., Vasely and Narum, 1970). The result of these calculations is a zero-one variable representing the intermediate or top variable. Given the simplicity of the envisioned logic hierarchies, C/J and C/I gates should present little programming difficulty.

Calculations for S/J and S/I gates will follow standard statistical methods using conditional probabilities. Because conditional probabilities in this case must satisfy the whole probability theorem, Bayes' theorem and other inverse techniques can be used either to improve efficiency or to provide redundancy checks. In most cases, lower level variables will themselves be described by distributions, making numerical integrations or related numerical techniques necessary. Given the limited number of variables, this presents little difficulty. For S/I gates, only the set of conditional probability distributions given each individual subvariable and the marginal distributions of each subvariable are required to obtain the upper variable distribution ($(y/x_1, \ldots, x_n)$ can be shown to equal $(y/x_1) \ldots (y/x_n)$). However, for S/J gates the joint distributions of the subvariables are also needed. Thus, S/J gates will require both more scaling and more computation than S/I gates.

F/J and F/I gates will require calculation of functions with random variable parameters. There are well developed techniques for performing variable transformations either exactly or approximately. In most cases, scaling of variable distributions will not follow analytical distributions and will not have explicitly identified moments, therefore simulation techniques (e.g., Monte Carlo) seem a natural way of performing the calculations. From the scalings, inverse transformation can be obtained numerically so that a uniform pseudo-random number generator can be used to perform simulations. Computational efficiency will be a key ingredient in these simulations because feedback times to the user should be short.

The result of the entire computational phase will be graphical displays of probability distributions over upper level variables, or ranges of probabilities for discrete events.

Logic Checks

A strength of computer-aided assessment is that the user immediately sees the implication of his assessments, and can assess variables in redundant ways as a check on consistency.

Two types of structural information are assessed: a direct logic hierarchy or tree, and an interaction matrix. Taken together, these are used to (1) check for internal consistency of the logic, and (2) uncover implied branches of the hierarchy. Both of these functions are performed interactively with the user deciding whether inconsistencies or implied structure are significant, and altering logic relationships to balance his perception of the variables with the underlying implications.

In the logic hierarchy of Figure 3-2, variable assessments would be made at all three or four levels, and the resulting implications checked either at the uppermost or intermediate levels. If these assessments are internally consistent, one can have increased confidence in the numbers obtained. If the assessments are inconsistent, the user goes back through the hierarchy to see the implications for subvariable scalings. With the computer acting as accountant, cycles of feedback and adjustment procede quickly, and a balancing among subvariable scalings is finally obtained.

If no balancing seems satisfactory, further work on the logic hierarchy is proposed by the program. One possibility is that hinting or reminder phrases be stored by key words in the program. These phrases would briefly describe physical or geological processes related to the variables being structured. For example, in the hydrofracturing case, one reminder phrase would ask, "DOES THE EXISTING JOINT SYSTEM INFLUENCE (HYDROFRACTURING DUE TO INTERNAL PRESSURE)?" As long as the subject matter under consideration has limited scope, as in the geological factors case, storage of such phrases presents few problems, and may be simply the collection of subvariables entered by other users for the same variable. Clearly, though, any feedback of other users' hierarchies or scalings should come after a complete attempt has been made by the current user, so as not to prejudice the results beforehand.

The idea of searching a logic hierarchy for inconsistencies and using interrelational (input-output) matrices to infer branches of logic trees is not new. Attempts at computer-aided fault-tree construction, cross-impact analysis and statistical decomposition analysis have all to one extent or another addressed this problem. Techniques from these works can, of course, be brought to bear in actually building the present program.

Implied structures, to a first approximation, can be found by searching the interrelational matrix for chains of variables connected by non-zero matrix entries. An exhaustive search of small matrices is computationally simple. For example, in the matrix of Figure 3-2, "available energy balance" is affected by "temperature" (i.e., entry (7, 9) is non-zero). "Temperature," in turn is affected by "water present" and "internal pressure," and "water present" is affected by "boundary layer properties," "in situ stress," "and tectonic stress changes." The latter two are primary variables because their respective column entries are entirely zero. "Boundary layer properties" has non-zero column elements, but the user may elect to truncate the tree at that point. This leads to the implied logic branch shown in Figure 3-4. The user may or may not agree that the implied branch is important. Entries in the interrelational matrix do not specify absolute magnitude, only direction and relative strength. If he thinks it unimportant he may truncate it. If he thinks it may be important, he can call the routine to identify gate properties, then recall stored scaling for the variables, and recalculate implied probability distributions for upper variables. In this routine the program acts as a classical interrogator, pointing out non-obvious implication of the user's intuition and forcing a confrontation.

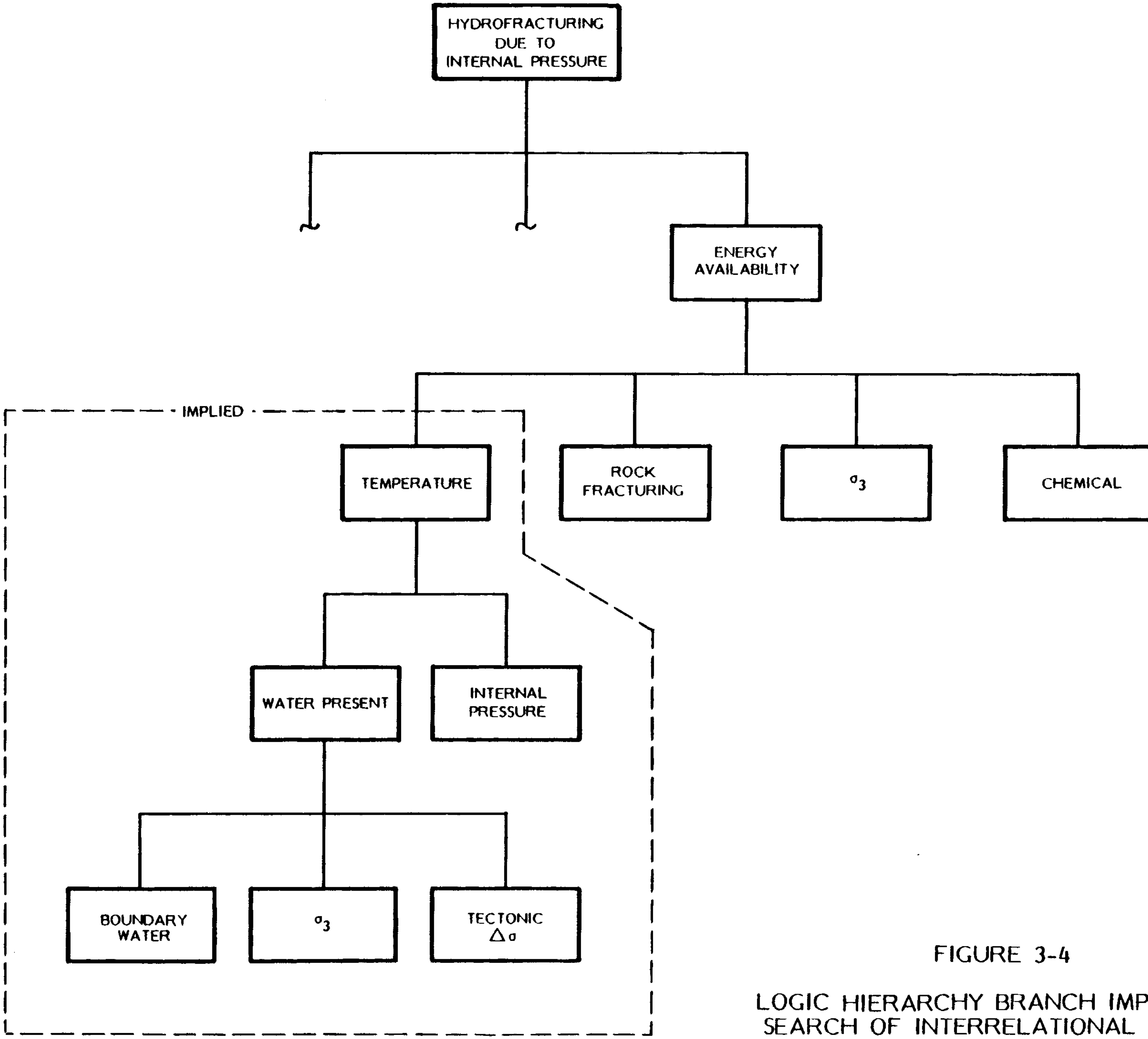


FIGURE 3-4

LOGIC HIERARCHY BRANCH IMPLIED BY SEARCH OF INTERRELATIONAL MATRIX

Finding logical inconsistencies may be somewhat more difficult than uncovering implied structures, but it is equally important. At least these directions must be considered. The first would use the interrelational matrix to generate existing branches of the logic hierarchy and compare the generated branches with those directly assessed. The second would check that completed feedback loops, through which a variable is functionally related to itself, do not exist in the tree. The third would identify "common-mode" variables through which branches of the hierarchy are de facto correlated. In itself the last is not an inconsistency unless the calculation phase treats them as independent. If inconsistencies are found (for example, if the assessed hierarchy and interrelational matrix do not agree) the user is presented with them so that he can obtain a resolution.

3.4 DECOMPOSITION: AN ILLUSTRATION

The purpose of the decomposition routines is to develop three aspects of the logic structure underlying an event or parameter: the hierarchical tree of subvariables, the relational form through which subvariables at any mode lead to upper variables, and an interrelational matrix that codifies the degree of interrelationship between each pair of subvariables. The form of questioning of this interactive process is exemplified in the following small pieces of dialogue.

In the course of interrogation, the program or the user identifies, for example, "the probability of undetected faults" as the variable to be assessed. The code then asks:

> *Can you decompose the (probability of undetected faults) into major component events or parameters on which it depends? Please indicate each on a separate line.*
>
> Spatial fault density in the region.
>
> Amount of exploration effort.
>
> *How are these variables related to the (probability of undetected faults)? Is the relationship causal, mathematical, or statistical?*
>
> Mathematical.

What is the equation relating (spatial fault density in region) and (amount of exploration effort) to the (probability of undetected faults)? Use the following variable labels:

```
Y     =    (probability of undetected faults)

X(1) =    (spatial fault density)

X(2) =    (amount of exploration effort)

Y     =    (1 - E ** (- X(1))) (1 - X(2))/(1 - E** (- X(1)))
           (1 - X(2)) + E** (- X(1)).  (D)
```

Based upon present information, the logic hierarchy is as follows:

```
              probability of undetected fault
                            *
                            *
                          F(1)
                            *
                            *
       - - - - - - - - - - - - - - - - - - - - -
       *                                       *
       *                                       *

Spatial Fault Density            Amount of Exploration Effort
```

```
F(1): Y = (1 - E** (-X(1))) (1 - X(2))/(1 - E** (-X(1))
(1 - X(2)) + E** (-X(1)).
```

Can you decompose (amount of exploration effort) into major component events or parameters on which it depends? Please indicate each on a separate line.

Type of exploration.

Density of exploration.

(D) This is Bayes' Theorem in which amount of exploration effort is taken to mean the conditional probability of finding an existing fault with the exploration program.

How are these variables related to the (amount of exploration effort)? Is the relationship causal, mathematical, or statistical?

Mathematical.

What is the equation relating (type of exploration) and (density of exploration) to (amount of exploration effort)? Use the following variable labels:

X(2) = (amount of exploration effort)

Z(1) = (type of exploration)

Z(2) = density of exploration)

*X(2) = (1 - E** (-Z(1)Z(2))) * (E)*

Can you decompose (spatial fault density) into major component events or parameters on which it depends? Please indicate each on a separate line.

In situ stress

Geological medium

How are these variables related to (spatial fault density)? Is the relationship causal, mathematical, or statistical?

Statistical

What is the correlation between (spatial fault density) and (in situ stress) and (geological medium)? Use the following variable labels:

X(1) = (spatial fault density)

Z(3) = (in situ stress)

Z(4) = (geological medium)

Statistical relationships can be specified as correlation coefficients, regression equations, or data tables (graphs). Select one.

Data table

(E) This is a simple exploration saturation function.

The user can now specify the relationship using a light pen or by typing in variable pairs. The code would respond with a graphical display of the current logic hierarchy, then try to further decompose the current lowest variables, and so on.

Having exhausted direct interrogation, the code moves to assessing an inter-relational matrix among the set of subvariables.

The following subvariables have been identified as contributing to the (probability of undetected faults):

X(1) = (spatial fault density)

X(2) = (amount of exploration effort)

Z(1) = (type of exploration)

Z(2) = (density of exploration)

Z(3) = (in situ stress)

Z(4) = (geologic medium)

Rank the degree of interrelation between each pair of subvariables, as they appear, on the scale:

-2 = functionally related in a negative direction

-1 = negatively correlated

0 = no interrelationship

+1 = positively correlated

+2 = functionally related in a positive direction

(Spatial fault density) and (amount of exploration effort)?

0

(Type of exploration) and (exploration density)?

+1

(Type of exploration) and (in situ stress)?

0

. . . and so on, until a matrix is developed. The matrix is used to explore for implicit logic structures and inconsistencies, and changes computed internally are fed back to the user for confirmation and agreement. The final hierarchy and IR matrix are stored. The code moves on to other top level events, or to scaling the individual subvariables.

4.0 SCALING VARIABLES

4.1 RECOMMENDATIONS

The scaling subroutines will present the expert with sequences of questions and from his answers back figure implied probabilities or distributions. Many mathematical and procedural techniques for scaling uncertainties have been proposed in the literature. The issue is to extract the most useful techniques and coalesce them into a practical scheme. To the extent possible, the logical structuring developed in the decomposition stage must be extended to the point that subevents over which probabilities are scaled do not have diminishingly small probabilities.

A three-phase procedure is suggested, consisting of pre-conditioning, scaling, and verification. Pre-conditioning identifies the dimension and range of a subvariable and presents key words which might trigger further introspection by the expert. Scaling introduces relative ranking of subevent probabilities and back figures probabilities implied by the expert's choices between dichotomous alternatives. Verification checks the internal consistency of probability assessments and feeds back inconsistencies for resolution. Figure 4-1 presents a flowchart representation of the scaling procedure for a global variable. A detailed flow chart focusing on the scaling of a subvariable is presented in Figure 4-2. These figures will serve as the focal point of this section.

4.2 OVERVIEW

Decomposition of an upper level event as discussed in Section 3 leads to a set of subvariables which can be logically related to the upper event. Thus, to obtain a probabilistic description of the upper event, probabilities or probability distributions over the subevents can be mathematically combined according to the logical structure assessed in the decomposition. The question at present, therefore, is how to obtain those probabilistic statements on subvariables.

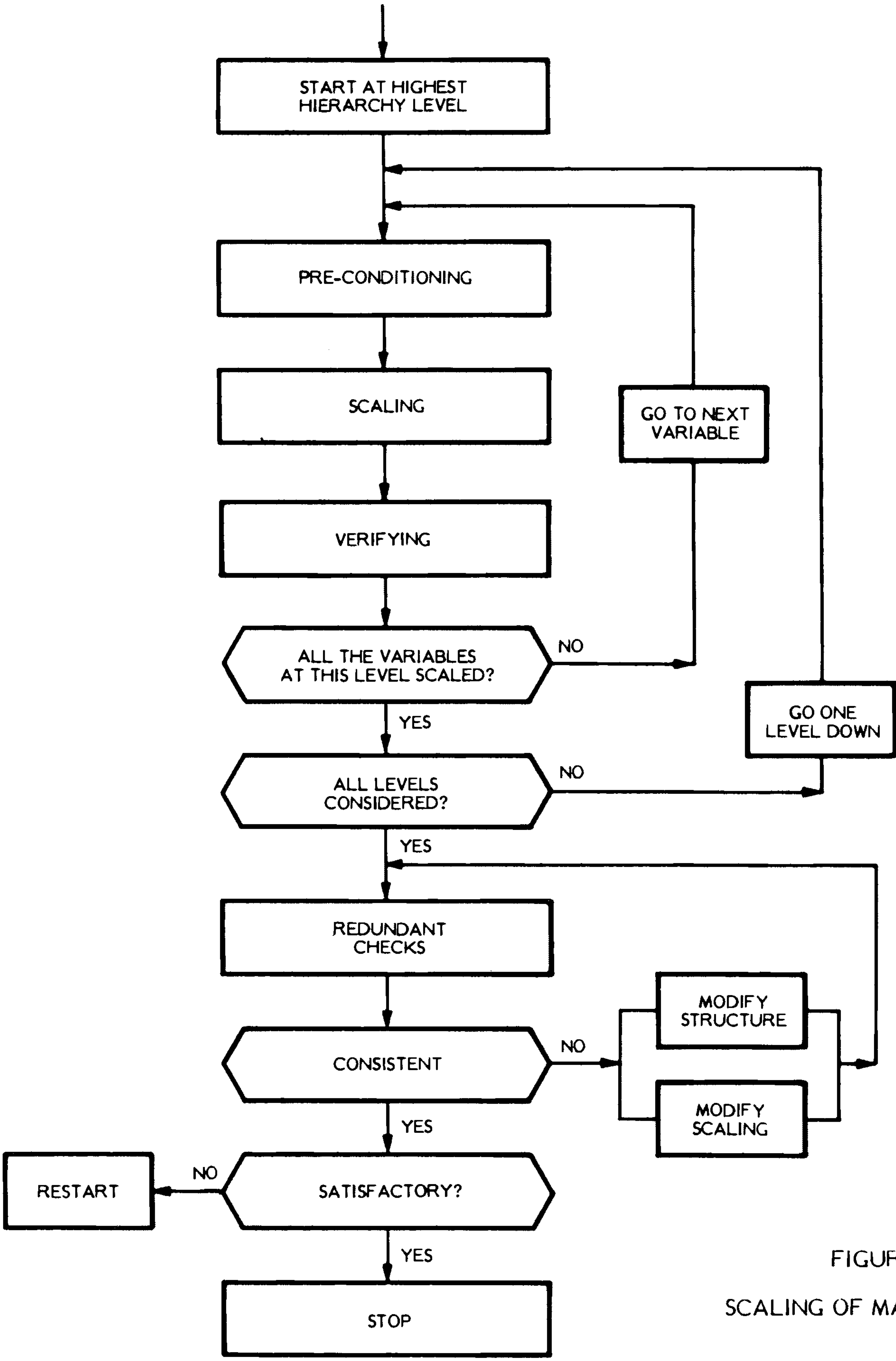


FIGURE 4-1

SCALING OF MAIN VARIABLES

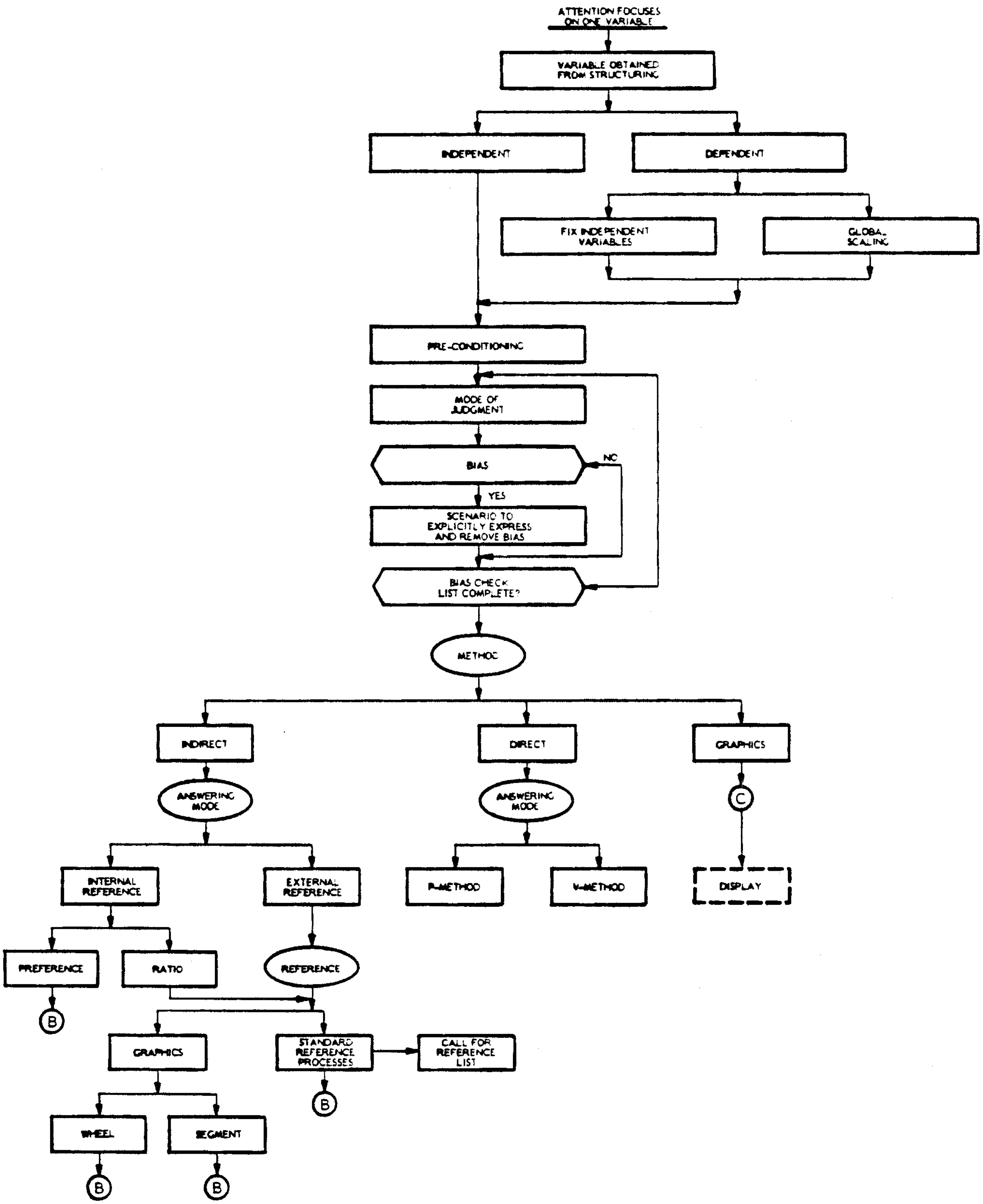


FIGURE 4-2

SCALING OF SUBVARIABLES

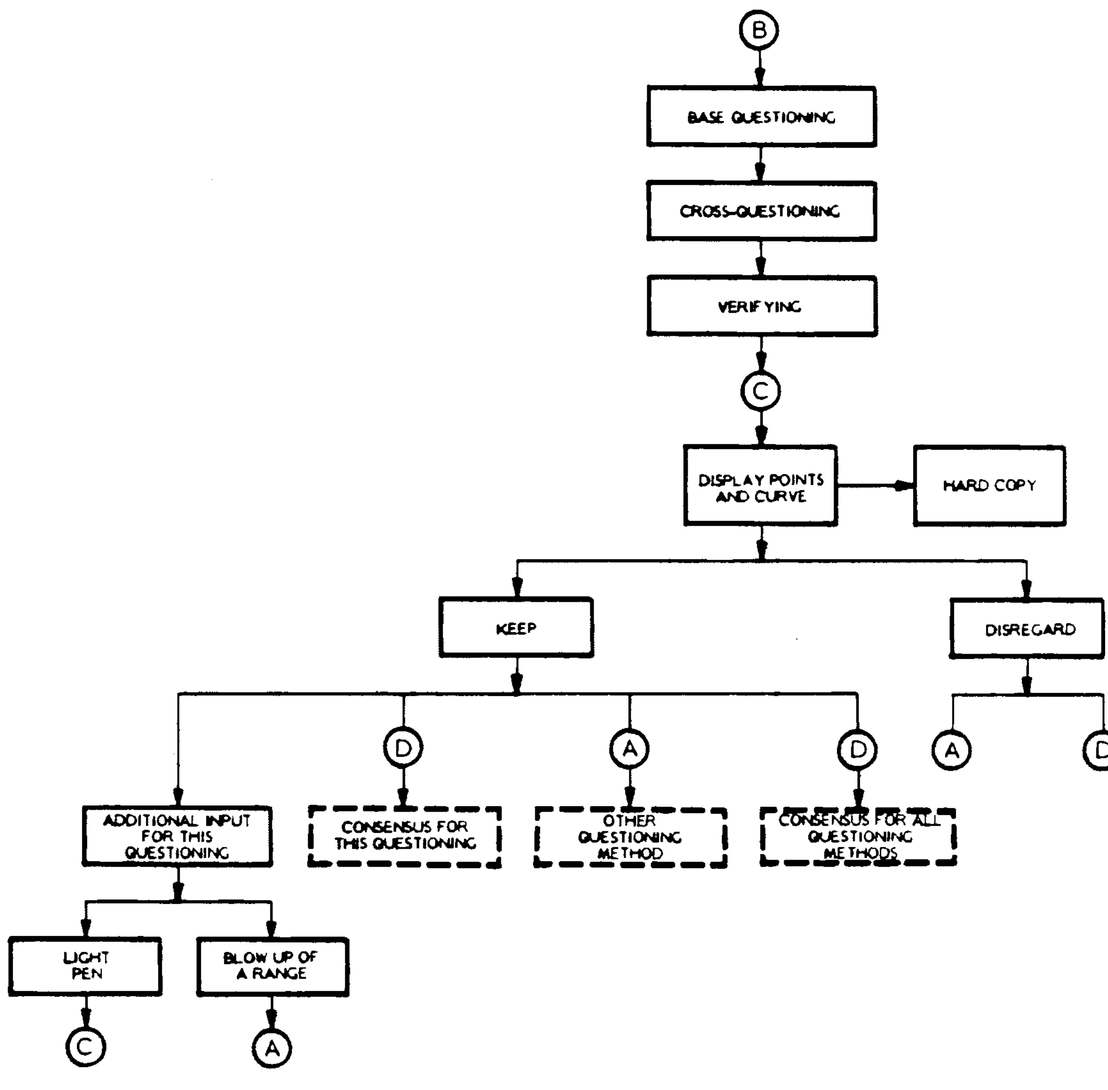


FIGURE 4-2
(con't.)

SCALING OF SUBVARIABLES

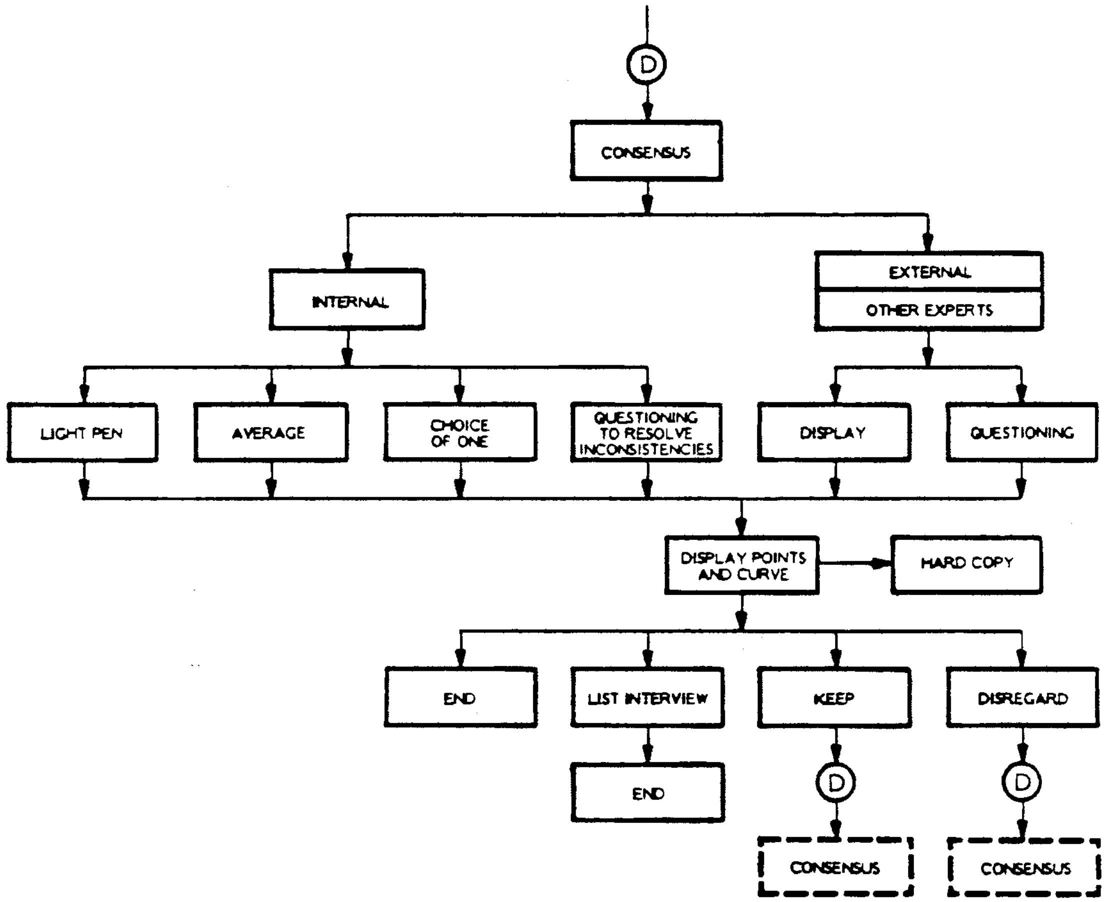


FIGURE 4-2
(con't.)

SCALING OF SUBVARIABLES

Fundamental to the entire idea of quantitative scaling of expert opinion is that "snap" judgments about nebulous variables are replaced by carefully thought out responses about precisely defined events. In quantifying those "carefully thought out" responses, it is of course much better to develop vague but true reflections of uncertainty rather than precise but wrong numbers. There are many competing, but in some ways complementary, methods for assessing and scaling uncertainties. Some work well in certain situations and for certain people, others work well for different situations and for different people. In all cases, responses to opinion assessments depend heavily on psychological variables. Assessment has strong "clinical" facets. To the extent possible, assessment methods should reflect what is known of psychological trends and biases. As the issue resolves to obtaining accurate reflections of opinion, a dogmatic approach of using only scaling techniques from one field (e.g., decision analysis, psychophysical scaling, Delphi) must be avoided in favor of a more eclectic and pragmatic approach.

The single most important result of probability scaling over subvariables is an accurate reflection of uncertainty and ranges within which some parameter might be realized. This bounding is much more important than "best estimates" or other measures of central tendency. Any assessment program must concentrate attention on ensuring, to the extent possible, that uncertainties and ranges of variables are not erroneously constrained by the form of questioning. As discussed in Section 4.4, the sequence of "what is asked" may strongly influence biases in measures of uncertainty.

4.3 NATURE OF PROBABILITY AND UNCERTAIN JUDGMENT

The nature of probability based on expert opinion is philosophically different than relative frequency concepts underlying classical statistics. Judgmental assessments result in what might be called "inductive" probability, or the degree to which a set of propositions or evidence supports or lends confirmation to some other proposition(s); that is, the degree to which full belief in some set of propositions leads to partial belief in others.

Mathematical probability theory can be based on a finite set of axioms within which the term *probability* is primitive: its properties are defined but its meaning is not.* This means the definition of probability is a question of philosophy, not mathematics, and has led to various schools of thought. The most well-known of these schools might be broadly categorized as relative-frequency and degree-of-belief, the latter defining probability as the degree to which one believes in the truth of some proposition or occurrence of an event. An objectivist view of degree-of-belief thought has been propounded in the literature by Keynes and Jeffreys, but today most degree-of-belief theory is subjectivist: degree-of-belief is unique to the individual, conditioned on his own unique experience, but modified formally upon new and enumerable information (i.e., evidence). As the formal modification of belief is accomplished through Bayes' Theorem, this school is often called "Bayesian." While the distinction between frequency and belief is widely argued in the literature, the important point at present is that the epistemological underpinnings of subjective probability or quantification of expert opinion are well developed and not *ad hoc*.

Within subjectivist theory the "goodness" of a probability assessment is reflected only in how accurately the assessment portrays an individual's judgment. Different people (experts) have different probabilities for the same events and all of them are "right." The extent to which some people have probabilities that are more externally valid than others (i.e., predictive accuracy as manifest in the real world) only reflects that some people have better judgment than others. So, the key in scaling is to accurately reflect true (personal) uncertainties . . . not to approximate reality. The latter is not possible *except by selecting people with high substantive expertise*. Well-designed methods of assessment cannot compensate for technical ignorance, nor can they increase external validity beyond that inherent to the judgment of the expert. Miracles are not allowed.

---

* For example, probability is a number between 0 and 1; the sum of the probabilities of exclusive and exhaustive events is 1; the probability of the joint occurrence of independent events is the product of the individual probabilities.

Opinions, and thus subjective probabilities, are often changed by group discussion, advocacy, and argumentation. Peer review through confrontation strengthens and clarifies the chains of assumptions leading to predictions and, in principle, increases the external validity of projections. The history of science, however, is full of examples of persuasive, brilliant, dynamic scientists who have been fundamentally wrong about important physical processes, and who through dint of personality have led the course of science astray. The first task is to have an expert argue with himself, and only then to compromise differences among experts. For, whatever the failings or successes of Delphi techniques, this principle of avoiding, at least initially, the interpersonal dynamics of small group discussion often proves fruitful in eliciting full ranges of opinion. Discussion of the problem of coalescing individual probabilities into group distributions is presented in Section 5. This is a question with a substantial literature.

4.4 CLINICAL ASPECTS OF SCALING AND EMPIRICAL BIASES

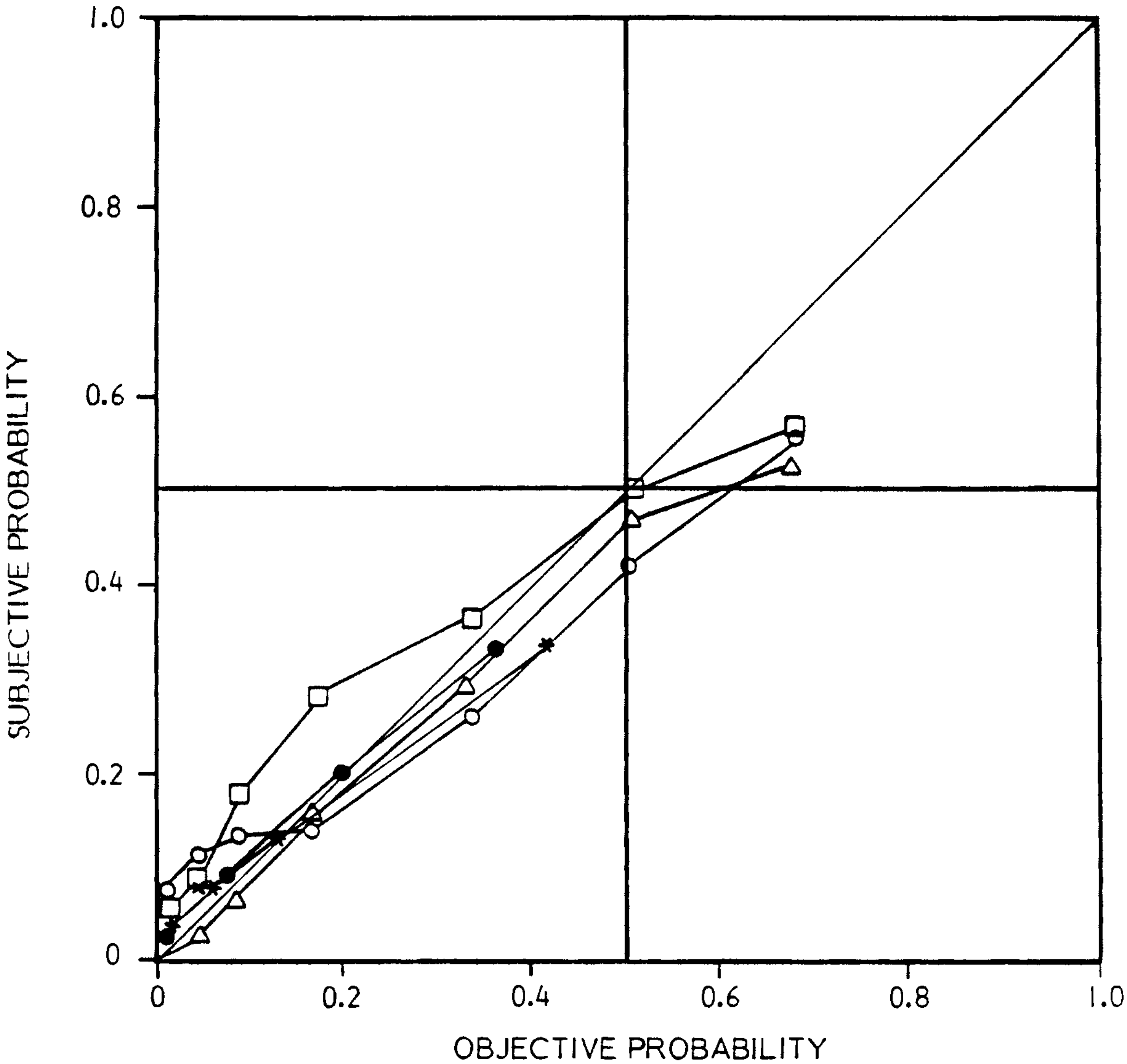
The consistency of probability assessment requires only that opinions expressed as probabilities satisfy the axioms of probability. In other words, that probability distributions integrate to 1.0 and the like. Thus the internal check for consistency, redundancy, feedback and related techniques do not ensure external validity. However, considerable work has been done in attempts to evaluate the validity of assessed probabilities, and that work indicates rather consistent biases in the opinions people give. These biases must be recognized in computer scaling routines. Some methods do exist for reducing such biases.

To evaluate past work on biases in probability estimation one might ask first "how well" people assess their uncertainties, and then "how" they assess. In other words, what is the empirical external validity of probabilities and what rules of thumb do people rely on in answering questions about uncertainties.

Briefly, at least three major and consistent biases are commonly observed in subjective probability distributions: gambler's fallacy, overconfidence, and conservatism. When asked to estimate the probability of a discrete event, or when subjective probabilities are inferred from risky decisions, a trend of overestimating small probabilities and underestimating large probabilities emerges (Figure 4-3). The crossover probability reported varies from study to study, but seems to be in the range 0.2 to 0.5. An implication for scaling routine development is that when asked to compare some uncertainty with an objective random device (e.g., roll of a die), subjects "misperceive" the objective probability. This bias is sometimes called gambler's fallacy. Similarly, when assessing probability distributions over continuous or multinomial variables, subjects tend to consistently underestimate their actual uncertainty. That is, assessed probability distributions tend to be too tight. Tail regions of an assessment, which should only see very small percentages of realizations (e.g., values of the uncertain variable outside the five and 95 percentiles should be realized in only 10 percent of the cases), empirically see up to half of the realizations (Figure 4-4). An implication for scaling is that techniques encouraging dispersion in assessed probabilities should be promoted. This bias is usually called overconfidence. Finally, when shown data and asked to modify their subjective probabilities, subjects tend not to change their opinions as much as Bayes' Theorem would specify. Subjects tend not to give as much credence to data as statistical theory would specify (Figure 4-5). An implication for scaling is that subjects should not be asked to express uncertainties as equivalent sample sizes or otherwise asked to perform intuitive updating from which *a priori* subjective probabilities are inferred. This bias is usually called conservatism.

In each of the above cases the conclusion of a consistent bias across subjects rests on considerable empirical verification. However, the interpretation of underlying causes or of the sources of the biases, whether in the subject or in the experimental procedure, is hotly debated. Nevertheless, it is clear that biases do exist when probabilities are scaled using common techniques.

The way subjects evaluate uncertainties, and, therefore, the source of observed biases, has been the subject of work by Tversky, Kahneman, and their colleagues, and has led to three broad heuristics. Tversky and Kahneman (1974) in their well-known paper call these <u>representativeness</u>, <u>availability</u>, and <u>anchoring</u>. <u>Representativeness</u> is the tendency to assign the probability of an event according to the degree of similarity it has with a broader group of events. This



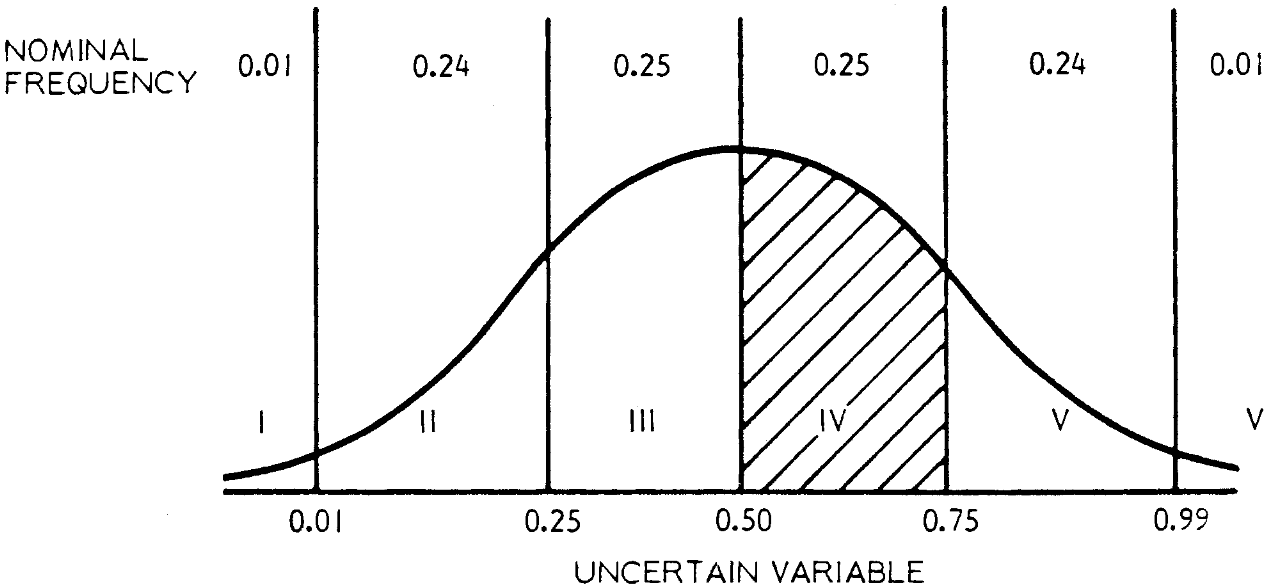
△ □ Mosteller, F., and P. Nogee (1951). "An Experimental Measurement of Utility," Journal of Political Economics, 59: pp. 371-404.

○ Preston, M. G., and P. Baratta (1948). "An Experimental Study of the Aution-value of an Uncertain Outcome," Am. Jour. Psychology, 61: pp. 183-193.

● * Griffith, R. M. (1949). "Odds adjustment by American horse race bettors," Am. Jour. Psychol., 62: pp. 290-294.

FIGURE 4-3

GAMBLERS' FALLACY BIAS IN SUBJECTIVE PROBABILITY



Variable Number	REGION I	II	III	IV	V	VI	TOTAL
1	3	16	20	40	11	10	100
2	15	12	35	19	10	9	100
3	11	8	28	29	13	11	100
4	51	41	6	1	1	0	100
5	1	1	13	28	29	28	100
6	24	14	12	13	10	27	100
7	1	3	11	9	15	61	100
8	9	2	13	10	8	58	100
9	25	15	18	9	7	26	100
10	18	8	8	12	16	38	100
Total	158	120	164	170	120	268	1000

FIGURE 4-4

FRACTION OF TRUE VALUES LYING IN RESPECTIVE REGIONS
OF RESPONDENT'S ASSESSED PROBABILITY
DISTRIBUTIONS, COMPARED WITH
NOMINAL FREQUENCIES OF
1, 24, 25, 25, 24, AND 1%

AFTER ALPERT AND RAIFFA (1969)

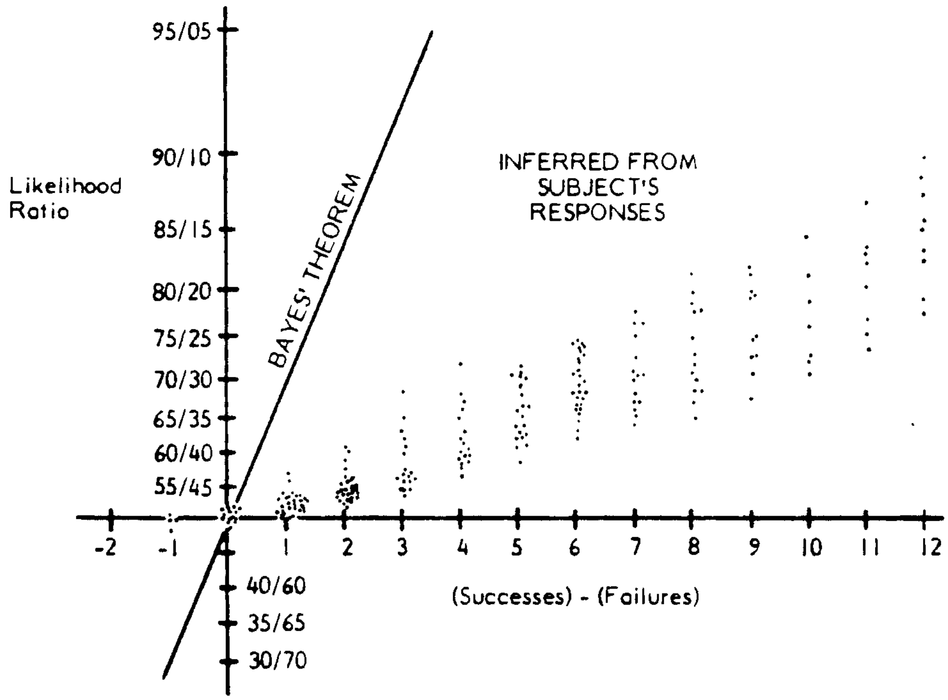


FIGURE 4-5

CONSERVATISM BIAS IN SUBJECTIVE PROBABILITY ASSESSMENT. INFERRED LIKELIHOOD RATIOS IN BERNOULLI TRIALS COMPARED WITH BAYES' THEOREM.

AFTER PHILLIPS AND EDWARDS (1966).

leads to a cluster of biases including (i.) insensitivity to prior or base frequencies when data is evaluated, (ii.) insensitivity to sample sizes in evaluating the weight of information, (iii.) misconceptions of chance self-correcting itself to cancel random fluctuations, (iv.) insensitivity to the predictability of information, and (v.) illusion of the validity of projections. An implication for scaling is that events should be structured in as much detail as possible. Availability is the tendency to assign greater probability to events that readily come to mind than to events that do not. Availability biases can be due to (i.) the ease of retrievability of instances of an event, (ii.) the relative effectiveness of methods of mentally searching for past examples of events, (iii.) the comparative ease of imagining occurrences, and (iv.) intuitive associations between events leading to illusory correlations. An implication for scaling is that the subject should enumerate and broadly survey his information at the beginning and then throughout the assessment. Anchoring is the tendency to focus on one piece of information or hypothesis and then insufficiently adjust the assessment to encompass the full range of uncertainty about that point. Anchoring can manifest itself in at least the following three ways, (i.) given a "starting point" or best estimate, subjects tend not to adjust their assessments sufficiently away from that point, thus estimates of central tendency are influenced by the starting point, (ii.) given a best estimate, subjects tend to assess their distribution of uncertainty about that point too tightly, and (iii.) conjunctive and disjunctive events are assessed with a bias toward the individual probability (i.e., compound events requiring the joint occurrence of multiple events are given probabilities which are too high, and compound events that require only the occurrence of one or more of a set of subevents are given probabilities which are too low). An implication for scaling is that bounds and ranges should be assessed first and central tendencies should only be developed by iterative bounding.

Thus, several perversions of quantitative judgment exist, and these apply whether subjective probabilities are quantified or not. The quantification itself is not the source of the biases explained in the Tversky and Kahneman's heuristics. Nevertheless, by identifying biases a scaling procedure can be developed which at the very least recognizes their existence.

One last bias must be mentioned, the bias due to motivation. Unlike the biases discussed above (usually called cognitive), motivational biases may be conscious, or nearly so. A subject may, for whatever reasons, wish to influence the outcome of a decision or the results of modeling. On the other hand, he may think that, as an "expert," he should make predictions about his subject with a strong degree of confidence. These biases can only be dealt with by trying to convince the subject to be honest in his assessments, and by impressing him with the importance of accurate statements of uncertainty. From the perspective of scaling methodology, motivational biases are difficult to deal with. However, this is not a problem unique to quantification.

4.5 APPROACHES TO SCALING

The scaling procedure within computer assessment might be divided into three phases: pre-conditioning, scaling, and consistency verification. In pre-conditioning the subject is directed to explicitly present those considerations which are determinants of his opinion on the variable. In scaling, the subject's subjective probability distribution is quantified. Finally, in verification, the quantified scalings are checked for internal consistency.

The events for which probabilities are scaled are those at the lowest level of decomposition in the event hierarchy. Thus, further decomposition into logically related subevents only nests another level of structuring. Rather, the pre-conditioning phase should be one in which the general dimension and range of the variable is identified, and questions and key words which might trigger further introspection by the subject are provided. These questions and key words would be provided exogenously in the development of the assessment code, and would be expanded by subjects' response as the code is implemented. For example, the subject would be asked for an "upper bound" on in-situ stress, and then asked if he can build a scenario consistent with an even higher level. He would also be asked to state in short phrases those pieces of evidence he thinks might bear on his judgment. These would, of course, probably not be quantitative. The intent would be to lay a groundwork for the subsequent scaling, and broaden the immediate perception at the time of scaling. This will require careful thought and planning when the code is written, but is important enough to warrant the effort. A number of scaling techniques have been developed in the literature of psychological measurement, decision analysis, operations research, policy analysis, and other fields. Torgerson (1958) and Pfanzagel (1971), among others, present techniques from psychological measurement, including such things as conjoint scaling, paired comparisons, and canonical representation. Stevens (1976) presents a number of techniques from psychophysics, including cross-modality matching and partitioning. Stael von Holstein (1970) reviews the decision analysis techniques, including inference from betting behavior. Quade (1976) discusses techniques from policy analysis as developed at RAND. To the extent that better assessments can be obtained, techniques from any or all of these disciplines should be used.

In essentially all scaling techniques the subject is asked to rank relative probabilities (i.e., develop an ordinal scale), and to make dichotomous choices between uncertain alternatives. From the answers to these questions implicit probabilities are calculated. Questions are asked in several ways so that redundant information is obtained, and cumulative probability distributions are developed. The scalings are usually performed by direct responses, in which probabilities are directly asked for; by indirect responses, in which probabilities are back figured; by graphs, in which the subject sketches his cdf or plots points using a light pen; or by relating probabilities of events to semantic variables (e.g., "likely," "unlikely," "probable," "improbable"). In the latter case great care must be taken to properly encode the quantitative meaning of the descriptive phrases because substantial variability exists across subjects (see, e.g., Lichtenstein and Newman, 1967; or Selvidge, 1972).*

---

* In general, semantic variables are at best a crude vehicle for encoding probability, and should be used only in conjunction with other techniques. Even then, great care must be used in their interpretation. Some workers would even hold that semantic variables can only be considered as ordinal.

In the initial work on developing an assessment, code techniques from subjective probability theory and decision analysis might be used. These techniques have the advantage of evaluated use, directness, and mathematical simplicity. Furthermore, codes now exist which might be straightforwardly adapted to the present purpose (Section 2.3). An illustration of this approach is given in the example following this section. In the course of initial implementation and testing, the adequacy of this approach should be evaluated, and modifications making use of other or related scaling techniques introduced.

Among the more difficult problems of scaling is correlation among variables. When the number of variables is limited (e.g., two or three dimensions), conditional distributions can be assessed. Scaling proceeds by fixing the level of one variable and assessing cdf's over the remaining variables. However, if more than two or three variables are statistically dependent, this approach rapidly becomes too detailed and expansive to be practical. Approximation techniques have been worked on for multiple variables by which joint cdf's are inferred from the set of bivariate assessments (e.g., Dalkey, 1975). Such techniques might be used. However, the degree to which the approximation introduces error is a function of the dimension of the joint cdf and the level of statistical correlation among the variables.

Other procedures for assessing joint cdf's include searching for variables upon which conditional independence is approximated, introducing transformations which lead to independence (e.g., the sum and difference between two variables are often independent even if the variables themselves are not), adopting restrictive families of distributions which have parameters specifying correlation (e.g., multivariate Normal distributions), and bounding correlations and testing for sensitivity in upper level event probabilities. The applicability of such techniques is discussed by Robinson (1971).

Redundant information should be collected in the scaling phase and evaluated in a verification stage. This last stage encompasses calculations by the code of conflicting or inconsistent probabilities, and resolution of conflicts or inconsistencies by the subject. At present, resolution is almost always accomplished by presenting the subject with the inconsistency and having him reflect introspectively and correct it. This seems to work in practice.

4.6 SCALING RARE EVENTS

The purpose of decomposition is to structure rare events into a series of subevents the probabilities of which are not either diminishingly small or extraordinarily large. People have a very difficult time thinking about the probabilities of rare events, and may not even have a language with which to describe them. The findings of Lichtenstein and Newman seem to substantiate this last point. Here "rare" is taken to mean outside the range (0.01 to 0.99). There is even some debate over whether people's assessments can be refined to precisions as small as one percent (e.g., can subjects actually distinguish between 25 and 26 percent, or even 25 and 30 percent?). However, precision may be a function of relative and not absolute magnitude. That is, a subject might be able to distinguish between one and two percent, but not between 50 and 51 percent. However, most of the subevents will not be rare.

If, in fact, a rare event probability must be directly scaled, some approximation technique will be required. To date, the only comprehensive work on scaling rare events is a Ph.D. thesis by Selvidge (1972)--see also Selvidge (1973)--coming out of the decision analysis literature. However, other work on this topic is currently under way. Selvidge presents a three-step procedure for assessing rare events: description and decomposition (conceptually similar to the intent of Section 3), relative ranking of probability, and numerical assignments. The latter is accomplished with a variety of ad hoc devices (e.g., 10^3 by 10^3 square grids with varying numbers of squares darkened). In overall concept, Selvidge's approach is to rely heavily on description and structured introspection rather than direct scaling. The approach uses such devices as including an explicit event, "other things overlooked," and specifying populations at risk and exposure to risk. In this way some rare events can be approached by considering larger collectives of risks and dividing by the size of the larger--sometimes hypothetical--population. It is not at all apparant that present techniques for treating direct scalings of rare events are satisfactory. The best way of dealing with such events, as argued in Section 3, is to exhaustively structure the logical relations of subevents and continue the decomposition until either non-rare subevents are identified, or the required interrelational properties among subevents exceed our physical understanding of natural mechanisms.

4.7 SCALING: AN ILLUSTRATION

The scaling routines are divided into the three parts previously mentioned: preconditioning, scaling, and verification. Continuing with the example of Section 3, and focusing on in-situ stress (σ_3), the assessment might proceed as follows:

> *We are now ready to scale probabilities for the elements identified in the decomposition. The first of these is (in-situ stress). List elements or references that may bear on (in-situ stress). For definitions, call HELP followed by the label.*
>
> HELP, elements, references.
>
> *ELEMENTS are quantities or variables upon which events may depend.*
>
> *REFERENCES are published studies.*
>
> *List elements or references that may bear on (in-situ stress).*
>
> Element: tectonic province, evidence of deformation, depth of burial, water pressure.
>
> Reference: Steinberg and Schwer, survey of North American rock bursts in mining.
>
> *Summarize your general thoughts based on these elements and references. This information will be maintained.*
>
> This region lies near a tectonically active area. It is thought to be undergoing fairly high rates of uplift. The formations are dry, therefore effective stresses will not be influenced by pore pressures. Steinberg and Schwer report considerable rock burst activity in deep mines within the same general area; thus, in-situ lateral stresses are probably high. As far as I know there is not much field evidence of large deformations in recent geological time. If there were, this would further confirm my

suspicions of high lateral stress. Because there is considerable soil overburden, existing evidence of deformations might be obscured.

Do you want to see a check list of other elements bearing on (in-situ stress), or the responses of other users?

No.

What level of (in-situ stress) would you be very surprised to see exceeded?

6,000 psi.

Suppose you were told that (in-situ stress) were in fact greater than (6,000 psi). Could you build a scenario consistent with this? If yes, describe briefly.

Yes. The region may be buttressing the tectonically active zone to the east.

Would you like to reevaluate your upper bound of (6,000 psi)?

Yes. 8,000 psi.

The code would procede in a similar way to establish a lower bound on stress. A central result of the pre-conditioning phase is to establish elements which, while not leading to "best"estimates, bear on what the variable might be in its extremes. In any judgmental task people tend to ignore the exceptional in favor of the "bulk" of the information. Thus, assessments often do not reflect true ranges of uncertainty.

After pre-conditioning, the code turns to scaling.

We are now ready to attempt to quantify your feelings about (in-situ stress). If you wish a description of the procedure for scaling call HELP, otherwise return.

Return.

What answering mode do you wish to use?

(1) probability wheel
(2) probability segment
(3) external reference list
(4) internal reference: preference
(5) internal reference: ratio
... ...

1.

Consider the following two bets. With one bet we spin the wheel in the left corner of the screen and if it stops with the pointer in the shaded sector you win \$50; otherwise, you win nothing (the wheel is set at 10 percent shaded). With the other bet you win \$50 if the (in-situ stress) is less than σ_i. *Which bet do you prefer? Type 1 if you prefer the wheel, 2 if you prefer the second bet.*

2.

Which one do you prefer now? (the wheel has been set at 90 percent shaded).

1.

The size of the shaded area is varied until the expert is indifferent between the two bets. The expert is then asked to determine the probability that the in-situ stress will be smaller than a couple of σ_i, then greater than a few $\sigma_j \neq \sigma_i$ always using the probability wheel as a mode of comparison. The results are then displayed on the screen and the following question is asked:

This is the information you provided by answering the above question. Choose from the following:

1. additional questioning
2. consensus for these answers
3. other questioning method
4. general consensus for this variable
5. disregard this set of answers

3.

Which method would you like?

5.

The variable range (1,000 psi, 8,000 psi) has been divided into seven equal intervals:

1: (1,000, 2,000)
2: (2,000, 3,000)
3: (3,000, 4,000)
4: (4,000, 5,000)
5: (5,000, 6,000)
6: (6,000, 7,000)
7: (7,000, 8,000).

Can you order them by increasing order of probability?

1, 2, 7, 3, 6, 4, 5.

Can you determine the relative probability of these intervals? Answer as the ratio of their probabilities where a negative multiple (-) indicates the first to be less favorable than the second.

6:1?

-3.

4:7?

2.

. . .

The intervals are each covered once, and some are covered more than once to allow cross-checking. In general, the redundant information comes not by repeated questioning on the same pair but by the probabilistic implication of the set of answers.

The code attempts to resolve inconsistencies by direct confrontation (i.e., "feedback").

The implication of your answers is inconsistent. This is natural, you should not consider it a flaw. However, the inconsistencies must be resolved. Your answers imply probabilities for the range (1,000, 2,000) of (0.1), (0.05), and (0.01). Can you pick the most appropriate from this set?

0.05.

The code then updates all the probabilities and searches for new or additional inconsistencies. The feedback continues until either consistency is achieved or the subject replies that he cannot refine his answers any further. In the latter case a best fitting cdf is calculated by least squares or a similar criterion.

When the subject is satisfied with his answers and final cdf is graphically displayed and stored in memory the code moves on to the next variable. While this procedure may seem tedious, experience has shown that with use such direct quantifications can be even quicker than personal introspection, since the computer acts as an accountant in maintaining logical consistency. In any event, the only alternative to careful and enumerated thinking about one's feelings is quick answers, which are probably not what one would like as input to further modeling.

5.0 SYNTHESIS OF EXPERT OPINION

Expert opinions as input to modeling or decision-making are not probabilities in the sense of recurring natural phenomena. They are not guesses at "true" probabilities, which in the case of non-recurring events are undefined, or defined only as a degree-of-belief (e.g., Savage, 1954). Rather, expert opinions are quantifications of individual beliefs, and are good or bad only to the extent that they have been properly assessed and represent the expert's actual belief.

How to use such expert opinion in modeling or decision-making has long been a problem. Among the more intriguing questions raised are: What is a "good" expert? How should expert opinion be incorporated with the analyst's own opinion and with parallel empirical data? How should conflicting expert opinion be reconciled? Is there an appropriate quantitative measure of the knowledge of an expert and of a group of experts? For these questions there are only ad hoc answers, or philosophical paradigms too complex or nebulous to be implemented.

5.1 RECOMMENDATIONS

A technique for synthesizing group opinion that is neither too ad hoc, and therefore improper, nor too sophisticated, and therefore unuseable, is needed. Several have been proposed. For example, a technique can be developed based on a two-part analysis: the first part would be a feedback and consensus phase with emphasis on reconciling logic hierarchies; and the second would be an error theory approach to synthesizing conflicting assessments. The error theory part is based on simple scalings of bias and random error, and correlation among experts.

5.2 EXISTING TECHNIQUES FOR SYNTHESIS

Synthesizing expert opinion involves the precision with which a probability is assessed, the confidence the expert has in the number, the knowledge the expert has on the particular variable, and the consistent bias errors which may stem from the means of assessing or from the expert. A limited number of tools have been developed for combining the above considerations and coming to a synthesis of quantified opinion. These are of four types:

- Consensus building
- Weighting schemes
- Error theory
- Likelihood methods

Each of these describes a related collection of tools, techniques, and tricks.

Any technique for synthesizing opinion must not simply take numbers as given, and operate on them. Rather, it should force an exploration of the basis of differences, identify important discrepancies among experts, and reassess each expert's opinion in light of the variation of opinion across the group (i.e., the method should include group feedback). Strong emphasis on identifying the logic structure or hierarchy within the assessment code allows computer-aided exploration of the differences among experts.

Consensus Building

Consensus building schemes are a central part of Delphi methods. The Delphi method feeds back group opinion in an anonymous summary form to each expert and asks the expert to re-evaluate his own assessments in light of the group opinion. If a consensus is reached through this cycling, assessment ends. If a consensus is not reached, another feedback cycle can be undertaken. Sometimes each expert is asked to briefly justify the reasoning behind his assessment, and this information is also provided in the feedback cycle. There has been considerable work in Delphi techniques on predictive error (Figure 5-1), reproducibility (Figure 5-2), convergence (Figure 5-3), and other methodological variables.

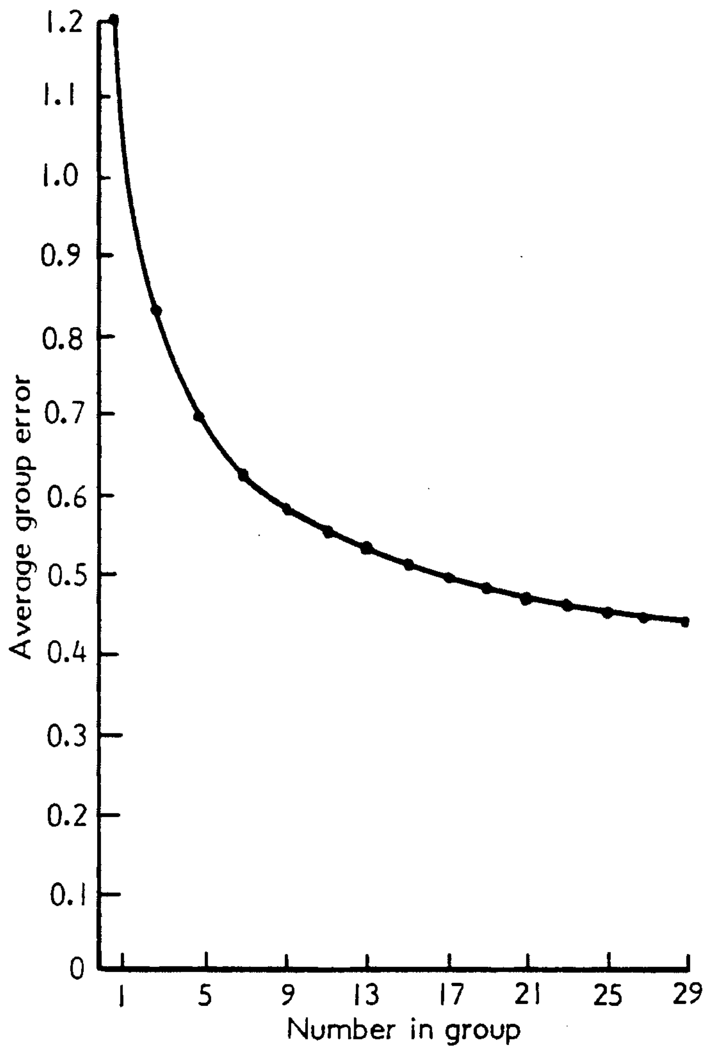


FIGURE 5-1

IDEALIZED DECREASE IN GROUP PREDICTIVE ERROR AS A FUNCTION OF GROUP SIZE, TYPICAL OF EMPIRICALLY OBSERVED ERRORS.

AFTER FUSFELD (1971)

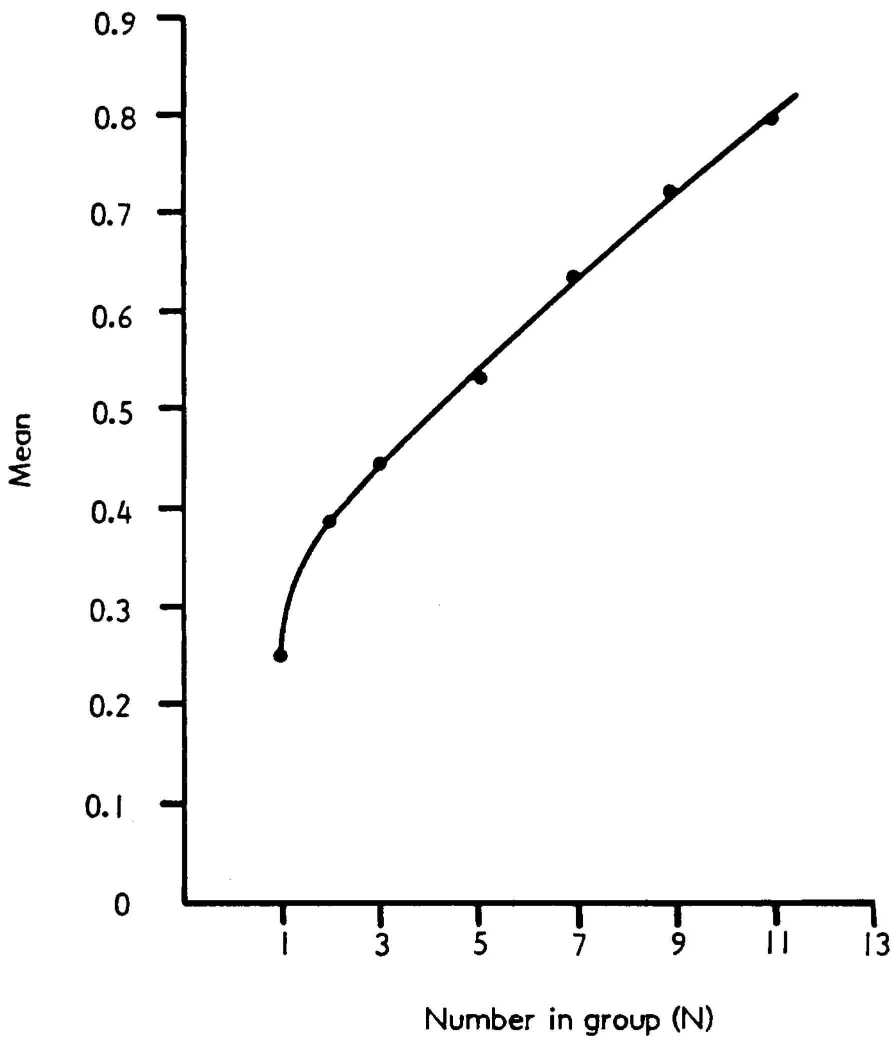


FIGURE 5-2

REPRODUCIBILITY OF GROUP RESULTS AS A FUNCTION OF GROUP SIZE. ORDINATE SHOWS CORRELATION COEFFICIENT AMONG GROUP RESPONSES.

AFTER FUSFELD (1971)

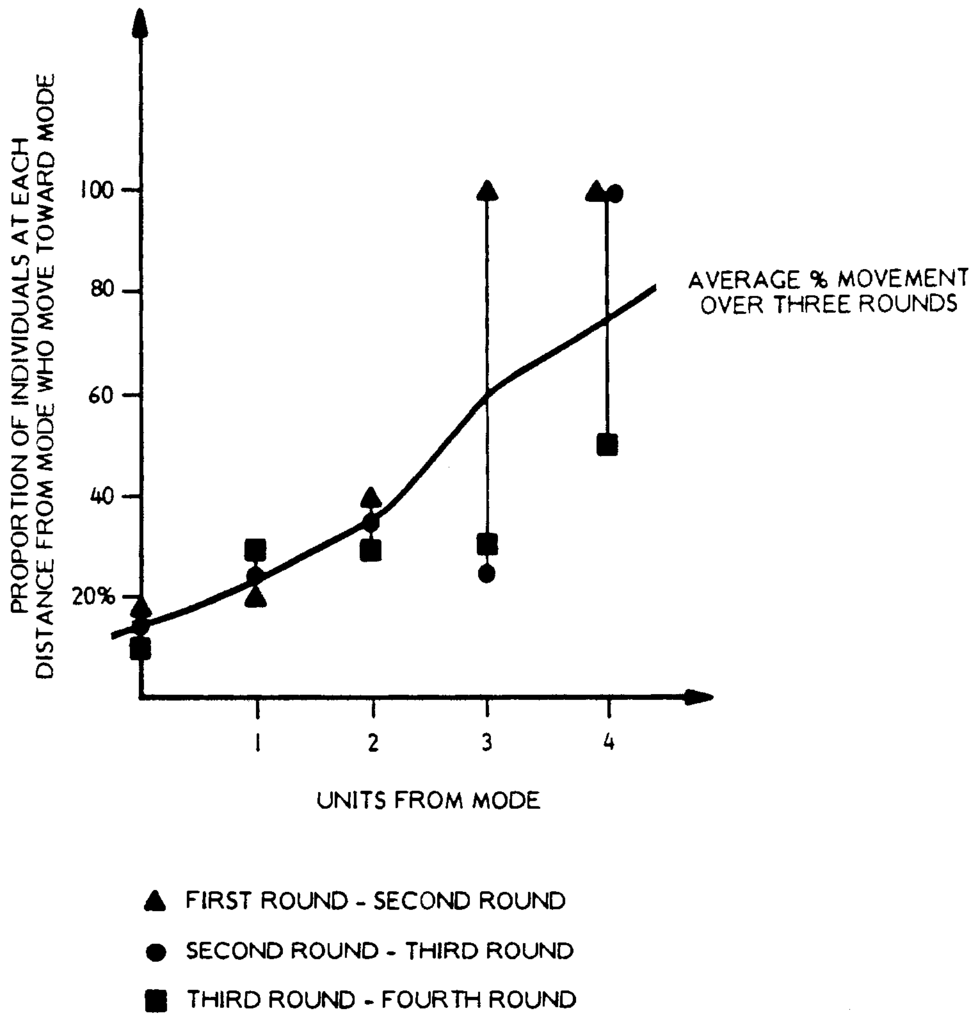


FIGURE 5-3

PROPORTION OF RESPONDENTS SHIFTING THEIR POSITIONS
AS A FUNCTION OF DISTANCE FROM THE MODE

(Adapted from The Delphi Method)

A problem with consensus building is that there may be very good reasons why opinion varies, so the consensus may not be a better estimate than the original distribution. Also, a consensus may not be attainable, and even if it is, it may be artificial. Work by Dalkey (1975) (Figure 5-4) and others suggests that the distribution of group opinion tends to be lognormal, so in cases where a consensus is not reached, an estimator like the median is usually suggested. The rationale for using the median is that it minimizes the "expected error" for variables generated by the obtained distribution (and in the lognormal case, coincides with the mode and geometric mean). This approach implicitly weights each expert's opinion the same and does not consider bias errors. Furthermore, no attempt is made by the analyst or user of the experts' opinions to interpret differences and reach a conclusion of his own.

Weighting Schemes

Weighting schemes take a "weighted average" of experts' opinions. The implicit assumption in weighting schemes is that one of the experts is "correct," but it is not a priori clear which one. Given experts' predictions $(x_1, \ldots, x_n)$ and weights $(w_1, \ldots, w_n)$, the synthesized prediction is:

$$\hat{x} = \sum_{i=1}^{n} w_i x_i \qquad (4.1)$$

The set of weights can be interpreted as a probability mass function over the experts, in which case the estimator $\hat{x}$ would have the properties of an expectation. Except for this special interpretation of weighting schemes, the procedure is ad hoc: weights are usually generated by rules such as, "attach the most weight to the assessment of the better or more knowledgeable expert."

The dimensions for measuring "better" or "more knowledgeable" experts are not immediately apparent. These qualities are probably multidimensional, which means reducing them to a single weight is difficult. In practice, weights are usually generated subjectively, either by asking each expert how he would weight

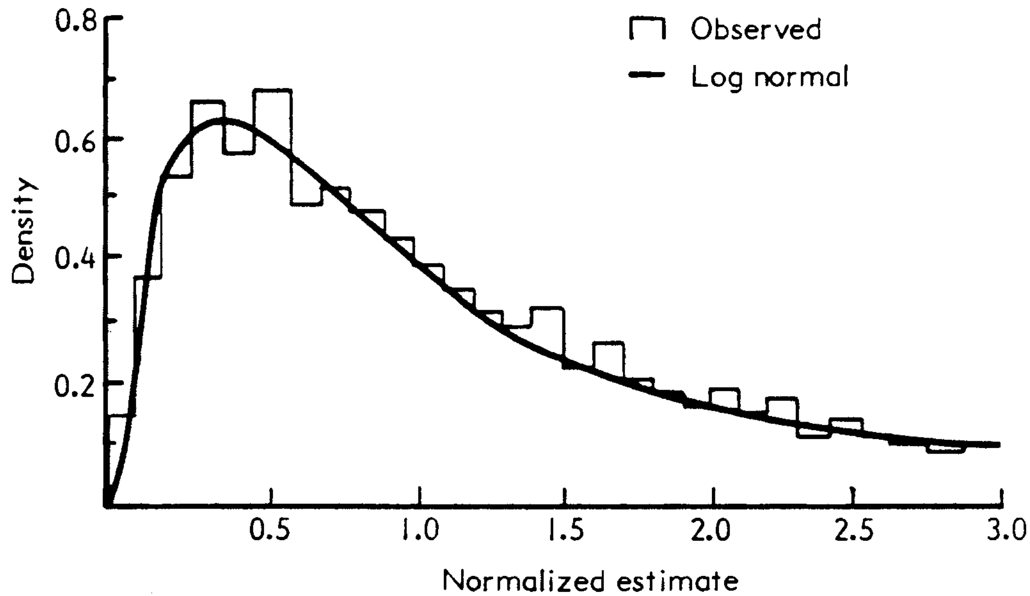


FIGURE 5-4

DISTRIBUTION OF ESTIMATES TEND TO
LOGNORMALITY OVER GROUPS
OF EXPERTS

AFTER FUSFELD (1971)

himself relative to others, or by asking experts to rank one another. Of course, the analyst could assign weights to the experts himself. Dalkey, Brown, and Cochran (1973) (Figure 5-5) have investigated the use of self-evaluated weights in forecasting, and Winkler (1968) has presented weighting schemes within a Bayesian framework. However, there is still work to be done.

A difficulty with weighting schemes is that they do not account for biases in experts responses. They take no account of optimism, pessimism or other systematic errors which are known to exist (e.g., Martino, 1970), and which may be suspected in a particular application.

Error Theory

Error theory is based on the assumption that experts are noisy transducers for "measuring" reality. Each assessment may contain both random error and bias error, and once estimates are made of the statistical properties of these errors, normal error propagation theory can be used to draw aggregate conclusions.

Error propagation satisfies some of the objections to weighted averages, but appropriate procedures for obtaining the error variances, etc., are still difficult to identify. Suggestions are sometimes made that errors can be statistically inferred from experts' past performances. However, there simply may never be an entirely objective way to evaluate expert opinion. Long track records by which statistical calibration can be made simply do not exist for most real assessment problems: "What has expert X's success rate been for predicting ice ages?" Calibration histories that can be collected necessarily deal with different types of variables than long-term forecasts deal with. Therefore, weights, error variances or other evaluation techniques will have to be partly subjective.

Likelihood Methods

Likelihood methods for synthesis have been suggested by Morris (1977) in a Ph.D. dissertation and subsequent papers. Likelihood methods treat experts' opinions as information in the Bayesian sense, then use the familiar updating procedures to

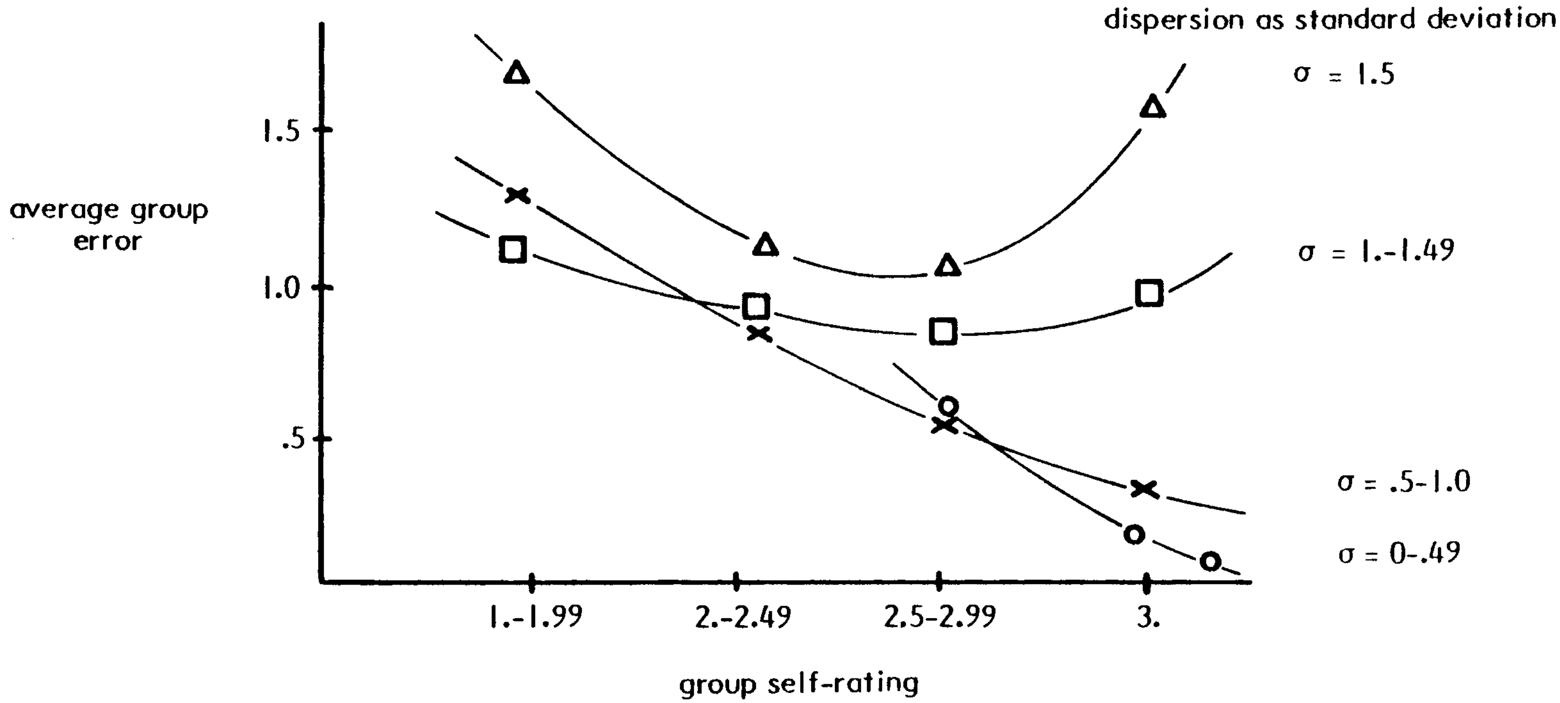


FIGURE 5-5

AVERAGE PREDICTIVE ERROR FOR GROUP ESTIMATES AS FUNCTIONS OF SELF RATING AND DISPERSION OF ESTIMATES AMONG GROUP MEMBERS

AFTER FUSFELD (1971)

combine experts' assessments with the analyst's own prior assessment to calculate a unique posterior probability or distribution. Evaluations of the experts' opinions are entirely contained within a joint likelihood function describing the conditional probability of each experts' assessments given the "true" state of nature.

As with the previous techniques, the joint likelihood approach points out an important consideration for synthesizing opinion: experts' opinions are generally not statistically independent. Opinion, and therefore assessments, are predicated on similar theories or logical structurings of a problem. Empirical evidence is often common or at least partially shared by experts. If this common dependence is not accounted for in aggregating assessments, the results become more precise than they should be. The implication is that the reasoning behind different experts' assessments should be studied, so that fundamental sources of disagreement are brought forth, and shared opinions, either on the way variables affect one another or on assessments of individual subvariables, can be partitioned off. Morris (1976) suggests techniques for calibrating experts' assessments (i.e., likelihood functions) using hypothetical assessments, but this procedure suffers the limitation discussed above: the assessments are hypothetical and of perhaps only marginally related variables.

5.3 SYNTHESIS TECHNIQUES FOR WISAP

Figure 5-6 presents a flowchart representation of synthesis techniques. This figure will serve as a focal point for the following discussion.

There must be two phases to the reconciliation and synthesis of differing assessments by different experts: an exploratory phase in which primary disagreements are identified; and a synthesis phase in which disagreements are aggregated at the primary level.

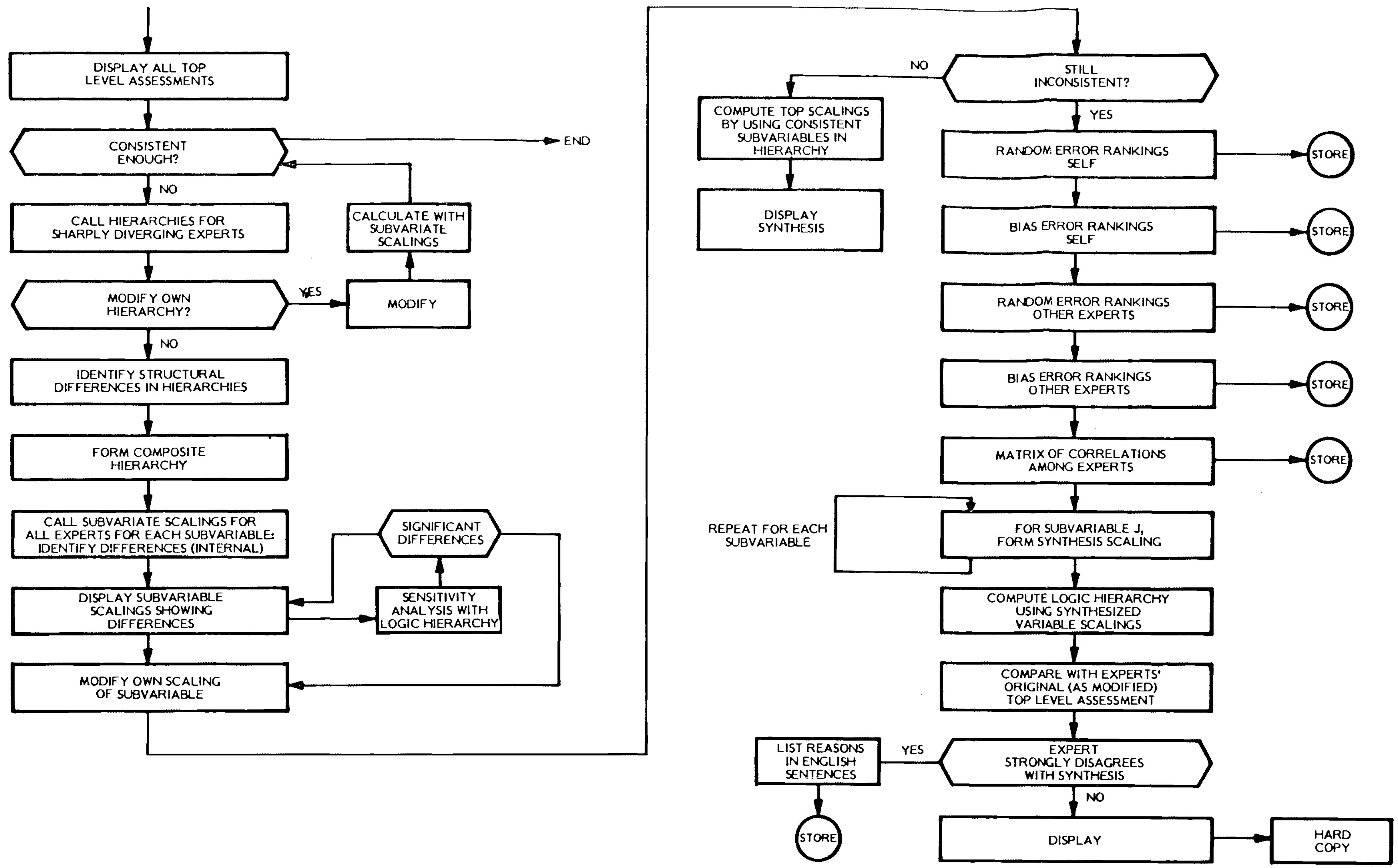


FIGURE 5-6

SYNTHESIS

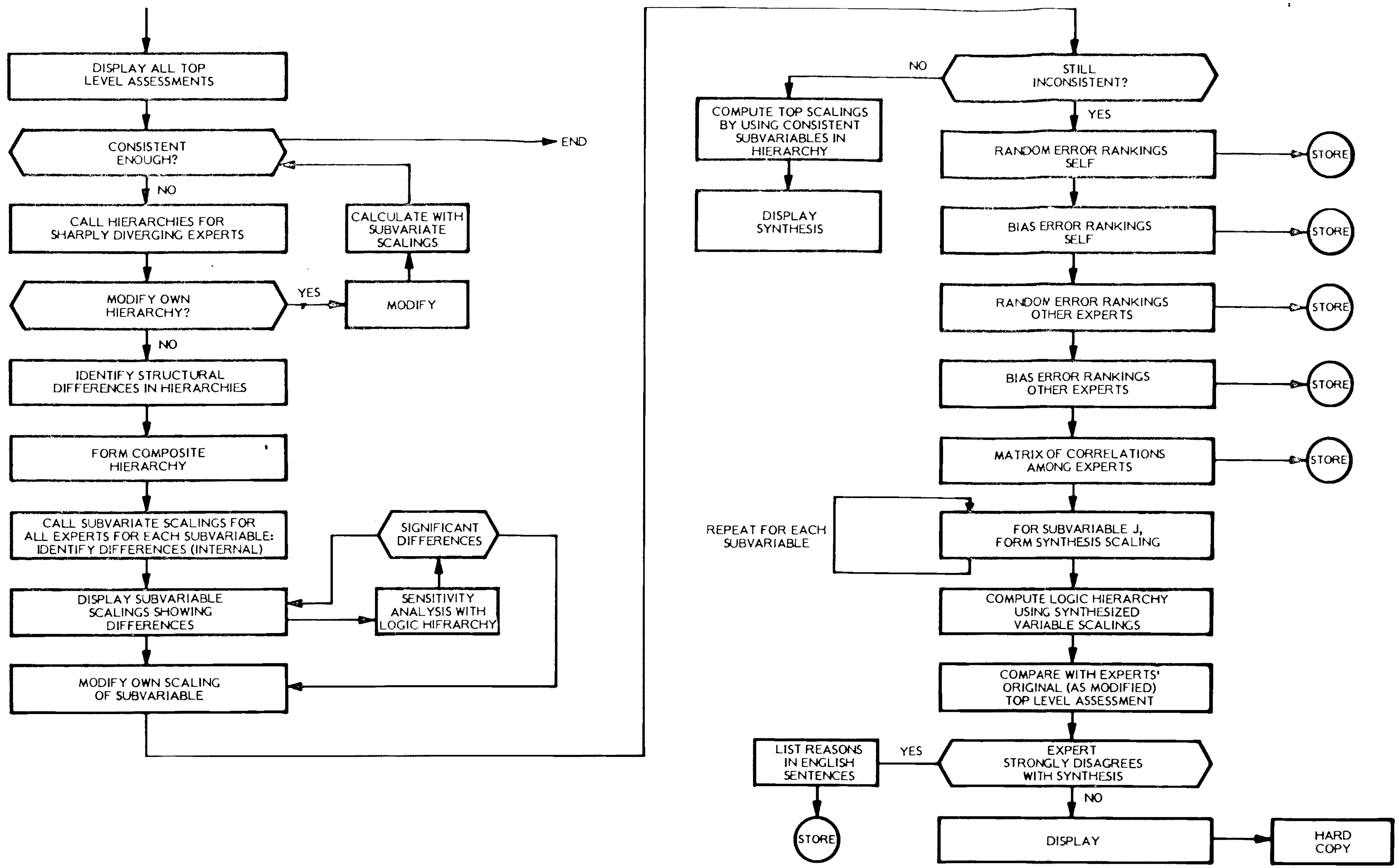


FIGURE 5-6

SYNTHESIS

A fundamental difference between the assessment code required by WISAP, and existing assessment codes is that the WISAP code would assess and store structural information in the form of a logic hierarchy underlying the reasoning leading to an assessment. Therefore, there is information other than simply the final numbers to be considered in synthesizing differing opinions. The first step in synthesis will be to recall logic hierarchies for a particular variable, feed all of these back to the experts, and have each expert re-evaluate his own hierarchy in light of the entire set of hierarchies. This may show that some variation is due to overlooked scenarios (i.e., branches) or overlooked interrelationships among subvariables. To the extent this is so, this first feedback cycle will lead to some convergence in final assessments. Further, to the extent that variations are due to differing perceptions of the logical structuring of events or parameters, the discussion is reduced to a fundamental rather than derived level. Basic differences are shown to exist and can be argued at the appropriate level.

After the logic hierarchies are reviewed by the set of experts, individual subvariable scalings can be reviewed. If the experts agree fairly well on the logical structure, then differences must be caused by differing scalings of subvariables or by differing assessments of the interrelations of subvariables. Both sets of information can, of course, be retrieved. If individual subvariable scalings differ significantly among experts, methods of the type summarized in Section 5.2 must be used to form a synthesis. If the interrelations of subvariables differ, some other technique for synthesis must be used.

The first step in reconciling interrelational differences should, again, be feedback and reconsideration by the group of differing experts. Beyond this strategy, a new set of methods will have to be developed. Any reconciliation must satisfy basic requirements of logical consistency, and consistency with the adopted logic hierarchy. The first step is sensitivity analysis on the differing components of subvariable interaction. It may be that differences that appear significant in the interrelational matrix are not, in fact, causing significant differences in top-level probabilities. In this case, the discrepancies are being caused by differences in subvariable scalings. If sensitivity analyses fail--that is, if top level probabilities are sensitive to differences in the interrelations of subvariables--then the range of top level variation must be determined and some form of weighting or calibration introduced to estimate a central tendency within that range.

Thus far the difficult problem has been skirted. If differences either in subvariable scalings or interrelations of subvariables cannot be reconciled, then some scheme must be introduced to establish (1) a measure of central tendency in the top level assessment, (2) a measure of "reasonable" imprecision in the estimate, and (3) the total range through which the assessment might vary. The third of these is easy; the first two are not. However, based on the previous discussion, a method can be divised. Clearly, though, epistemological constraints limit the objectivity that can be achieved in such a method. There is no way around this problem, because any realistic method will be in part ad hoc.

For theoretical and operational reasons, the most favorable way of synthesizing assessments would appear to be a combination of error theory and calibration. Feedback and consensus building will not work because they would have been already tried. Weighted averages seem inappropriate because they do not allow for systematic biases or incorporate estimates of random error (i.e., imprecision). Joint likelihood methods involve difficult conditional (inverse) probability statements, and since there probably will be no analyst's prior probability to update (or if there is, it will be fairly diffuse), they appear to carry no advantage to error theory calibrations.[1]

Error theory calibrations require an estimate of each expert's imprecision, or the random error in each expert's assessment, an estimate of potential bias error and the direction of the bias (this estimate need not be sharp), and a measure of

---

[1] Likelihood approaches use the conditional distribution $p(x_i/Y)$, where x_i is the expert's assessment and Y is the predicted variable, through Bayes' Theorem to infer $p^1(Y/x_i) \propto p^o(Y)\, p(x_i/Y)$. The "prior" $p^o(Y)$ is the analyst's opinion. Error theory approaches attempt to directly establish $p(Y/x_i)$.

correlation with other experts' assessments. Let the random error or imprecision be denoted e_i, the bias error be denoted b_i, and the correlation of assessments by experts i and j be denoted $_{ij}$. The imprecision in the estimate b_i will be denoted by a variance term $V(b_i)$.

A best estimate of the variable is obtained by a minimum variance linear estimator

$$\hat{y} = \sum_{i=1}^{n} w_i x_i \tag{4.2}$$

where $\hat{y}$ is the estimate, x_i are the experts' best estimates and (w_i) is a set of optimal weights subject to the constraint $\Sigma\ w_i = 1.0$. Define a new variable,

$$\sigma_i^2 = V(x_i) + V(b_i) \tag{4.3}$$

where $V(x_i)$ is the variance of expert i's assessment. $V(x_i)$ can be approximated by inspection or calculated using a simple numerical algorithm. Define the covariance of experts' assessments as

$$Cov(x_i, x_j) = \rho_{ij}\ \sigma_i\ \sigma_j \tag{4.4}$$

If the bias in each expert's estimate is removed by the transformation[2]

$$\hat{x}_i = x_i - b_i \tag{4.5}$$

---

2 The bias b_i could either be defined as a proportion or as a difference (as above). The choice depends on which is more easily used by the expert. If defined as a proportion, equation (4.3) and (4.5) would be modified appropriately.

then the optimal set of weights can be obtained as that which minimizes the estimate variance

$$\min_{\underline{\underline{w}}_i} \; E\left[(\hat{y} - w_i x_i)^2 \right] \qquad (4.6)$$

Introducing a Lagrange multiplier λ , the optimal weights can be shown to be (Appendix A),

$$\left\{ \begin{array}{c} \{w_i\} \\ \hline \lambda \end{array} \right\} = \left\{ \begin{array}{c} \underline{\underline{\emptyset}} \\ \hline 1 \end{array} \right\} \left[\begin{array}{c:c} \underline{\underline{C}} & \begin{array}{c} 1 \\ \vdots \\ 1 \end{array} \\ \hdashline 1 \; . \; . \; . \; 1 & \emptyset \end{array} \right] \qquad (4.7)$$

where

$$\underline{\underline{C}} = \begin{bmatrix} Cov(\overset{\sigma}{x}_1, \overset{\sigma}{x}_1) & \cdots & Cov(\overset{\sigma}{x}_1, \overset{\sigma}{x}_n) \\ \vdots & \ddots & \vdots \\ Cov(\overset{\sigma}{x}_n, \overset{\sigma}{x}_1) & \cdots & Cov(\overset{\sigma}{x}_n, \overset{\sigma}{x}_n) \end{bmatrix} \qquad (4.8)$$

The variance in the estimate is found by substitution into equation (4.6)

$$V(\hat{y}) = \sum_{i=1}^{n} \; \sum_{j=1}^{n} w_i w_j \; Cov(x_i x_j)$$

With a mathematical procedure for synthesizing assessments the question becomes, how can the required calibrations be established. This will be done through self-evaluation by each expert, and evaluations by the experts of one another.

An assessment procedure, as any analytical technique, benefits from simplicity. It is usually better to know how an analytical procedure is working and therefore clearly recognize its limitations, than to inaugurate complex procedures which result in outputs that are difficult to interpret (hence, the old adage "it is better to be vaguely right than precisely wrong"). Given the inability to specify an objective or empirically based calibration methodology, the procedure for calibrating and aggregating experts' opinions should be simple. The critical test of such a procedure is its applicability in practice. Therefore, the procedure discussed below is a starting point, which must be refined and modified in initial testing phases of the assessment code.

Each expert is asked to rate his knowledge and therefore the precision of his scaling of each subvariable on a simple 1-5 scale. "1" indicates very little knowledge about the variable and very low confidence in the precision of the scaling; "3" indicates familiarity with the variable and confidence that the precision of the estimate is "as high as for most experts in the field"; "5" indicates extensive knowledge (equal to the most knowledgable experts) and high confidence that the scaling is precise (there is little potential error in the scaled probabilities). Numerical levels of precision associated with the ranking (e.g., "2" corresponds to $\pm$30% error) are not explicitly stated in the questioning, but at the user's option may be input as associated with each rank along the scale. Also, each expert is asked to rank the other experts along the same scale. The timing of ranking other experts will have to be experimented with. It would seem that the "others" ranking should not take place until an initial assessment has been made, and possibly not until at least one cycle of feedback has occurred.

From these rankings of knowledge and assessment precision, an estimate of random error must be made for each expert. There is some experience with self and others ranking in technology assessment (Dalkey, Brown, and Cochran, 1970; Dalkey, 1969; and North and Pyke, 1969, among others). This must be combined with initial experiments to develop a correlation between rankings and error estimates. Obviously, these error estimates will be rough, but the purpose at

---

Dalkey, N. C. (1969). "The Delphi Method: An experimental study of group opinion." RAND Corporation, RM-5888-PR.

North, H. P., and D. L. Pyke. "Probes of the technological future." Harvard Business Review, v. 3: 68-76.

hand is to form a synthesis of opinion and rough estimates of error differences are better than assuming all experts are equally precise. The mean ranking for an individual expert cannot be used to summarize his expertise, because the 1-5 scale is only ordinal. However, the median ranking does have mathematical meaning, and might, therefore, be used instead. Even so, attention must be given to the question of balancing self rankings with others ranking.

Bias errors (e.g., optimism or pessimism) are probably best estimated through ratings of others. Again, a simple rating is given over an ordinal scale, and that scale is then correlated to percent bias errors. The median rating is used as the summary statistic. At the user's option the scale may be directly associated with input bias errors specified in percent, or may be specified, for example, on a -5 to +5 ranking, with zero indicating no bias.

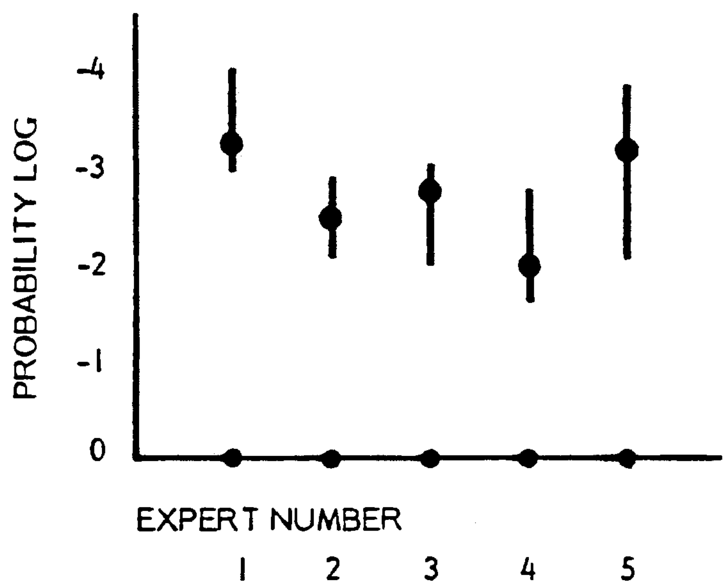
Correlations among experts' opinions reflect similarities in the schools of thought the experts represent. Certainly, if three experts all represent the same general philosophy on a particular scientific issue, their assessments will contain redundant information. If a fourth expert represents a different philosophy, then a priori his opinion should be given more weight than suggested by his 1:3 minority position on the expert panel, because his opinion is more independent than the others' opinions are. The correlation among experts is easily and compactly represented by a correlation matrix of the terms ρ_{ij}. These correlations could be directly requested from each expert for each pair of experts. This would be a scaling on $-1, +1$. However, statistical (i.e., quantitative) correlation coefficients are foreign to the intuition of most people, at least in their appropriate association with, say, data scattergrams, and a less direct ranking of interrelation would seem more conducive to valid responses. This might be done using an interrelational matrix of the type discussed in Section 2, in which perhaps a broader scale (-5, +5) is used and later associated with the appropriate correlation coefficient.

As in all the synthesis problems, reasonable, rather than precise, solutions are sought since precise solutions are likely to be red-herrings. A simple ranking device is developed, against which experts can easily and straightforwardly state their beliefs. Opinions on the credibility or knowledge of other experts can be strong, but they are seldom precise. Yet inability to specify precise numbers is no reason to abandon otherwise correct procedures. To obtain statistical measures for use in calculations, the ranking scale is correlated or associated with the statistical measures on the basis of experimentation in early implementation phases of the code.

The output of synthesis calculations would be a graph showing individual assessments, the ranges of those assessments, the synthesized assessment, variance, and range, as illustrated in Figure 5-7.

5.4 SYNTHESIS: AN ILLUSTRATION

Displayed below are the assessments of the entire group of experts for the variable (probability of undetected faults). Are these sufficiently consistent simply to take their average?



No

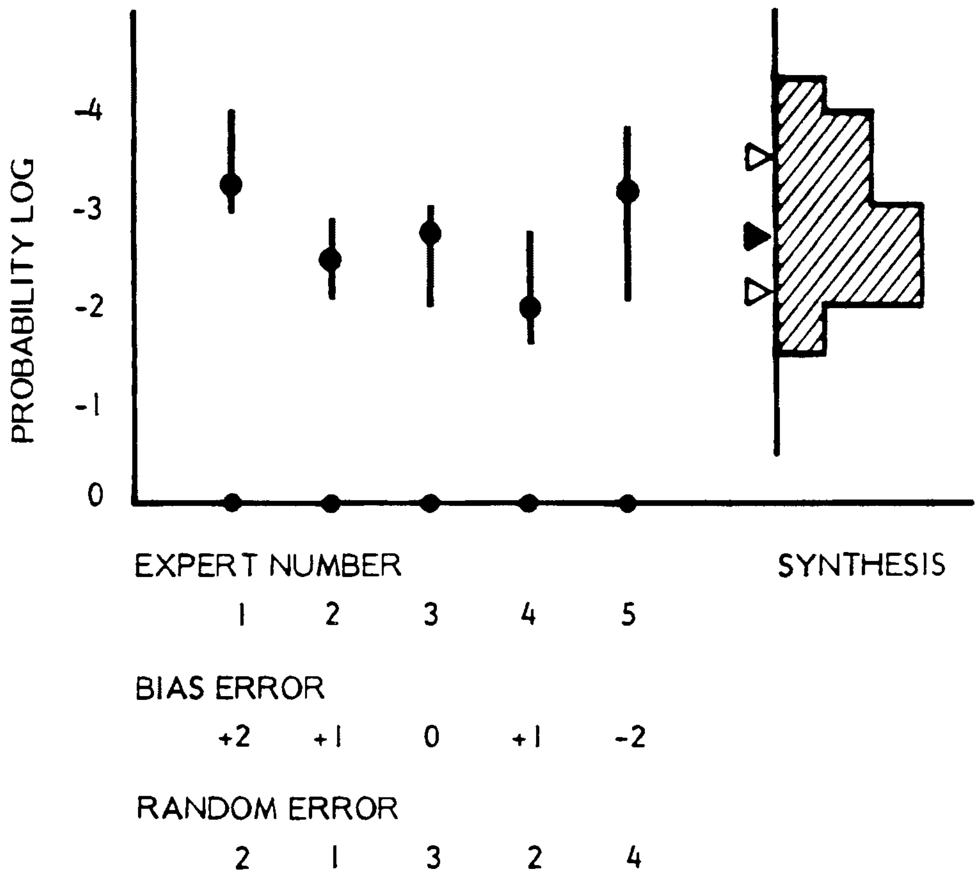


FIGURE 5-7

INDIVIDUAL AND SYNTHESIZED ASSESSMENTS

Shown below (for hard copies return HC) are the logic hierarchies of the (4) experts for the variable (probability of undetected fault). Please study these.

HC

Having seen the group of hierarchies, would you modify your own?

Yes

How? Indicate by "add" or "delete" branch at node indicated by variable number.

Add (existing seismicity) at node Y

Your hierarchy is now:

```
              probability of undetected fault (Y)

                              *
                              *
                             F(1)
                              *
                              *
           - - - - - - - - - - - - - - - - - - - -
           *                  *                  *
           *                  *                  *
   spatial fault      amount of            existing
   density x(1)       exploration       seismicity x(3)
                      effort x(2)
           *                  *
           *                  *
          S(1)               F(2)
           *                  *
           *                  *
 - - - - - - - - - -    - - - - - - - - - - - -
 *                 *    *                     *
 *                 *    *                     *
in situ   geological  type of              density
 stress     medium    exploration            of
  Z(3)       Z(4)       Z(1)            exploration
                                            Z(2)
```

How should relationship F(1) be modified?

Y = (1 - E** (-X(1))) (1 - X(2) X(3))/(1 - E**(-X(1)))
(1 - X(2) X(3)) + E** (- X(1)).

C existing seismicity increases probability of detecting fault. The surrogate variable X(3) is defined as an enhancement factor which will be subjectively correlated to microseismic activity.[3]

Would you make other modifications?

No

Indicate subvariables common to the set of hierarchies. Use variable labels prefixed by expert number.

1X(1) = 2A(3) = 3Z(1) = 4B(2)

.
.
.

Scalings for variables 1X(1), 2A(3), 3Z(1), 4B(2) are shown below. 1X(1) and 3Z(1) fail the similarity test based on overlap. Do you think this difference is significant?

Yes

In a similar way the code leads to an identification of differences among the scalings of individual variables. Internally, the code tests the sensitivity of top variable scalings to these differences through a Monte Carlo subroutine or related numerical technique. The code then proceeds to precision and knowledge ratings for subvariable scalings still not consistent.

[3] "Comment" notes might be permissible to be stored with logic hierarchy.

On a scale of 0 to 10, rate your own knowledge and the precision of your scaling of (spatial fault density). For reference, this scale should correspond to:

0 = little or no knowledge

5 = average familiarity of expert in general field

10 = highly knowledgable

7

Previous calibration studies using this code have shown a rating of (7) to roughly correspond to an imprecision range of (70%) to (130%) of your estimate. Would you agree? If you disagree, please specify an imprecision range.

Agree

A similar rating is made for the other experts.

Do you consider yourself optimistic or pessimistic? In other words, might your answer be somewhat biased? Rate your own bias on the scale -10 to +10. This scale should correspond to:

-10 = very biased to the low side

0 = unbiased

+10 = very biased to the high side

If more convenient, use a best estimate of your own percent bias. Indicate by symbol %.

2

Self ratings of bias are probably not very good, so each expert is next asked to rate every other expert on the same scales.

Experts often base their opinions on similar schools of thought. To this extent, errors in experts' predictions would be correlated. For each pair of experts, as they appear, indicate the degree of *correlation that you think will exist between their answers. Use the following scale:*

-5 = perfect inverse correlation

-2 = modest inverse correlation

0 = totally independent

+2 = modest simultaneous correlation

+5 = perfect simultaneous correlation

Expert 1 (George Smith) and Expert 2 (Robert Johnson)?

+3

. . . and so on, until a matrix is completed. This matrix may be 5 x 5, or a similar dimension. Based upon the synthesis equations in the text, this ranking information is used to form a synthesized estimate of the subvariable scaling. This synthesis includes a "best" estimate, variance and range, and is presented as shown schematically in Figure 4-3.

6.0 REFERENCES

Alpert, M., and H. Raiffa, 1969. "A Progress Report on the Training of Probability Assessors," Mimeo., Harvard Graduate School of Business Administration.

Dalkey, N. C., 1967. "The Delphi Method: An Experimental Study of Group Opinion." RAND Corp. P-3704.

Dalkey, N. C., 1969. "The Delphi Method: An Experimental Study of Group Opinion." RAND Corp. RM-5888-PR.

Dalkey, N. C., B. Brown, and S. Cochran, 1970. "Use of Self Ratings to Improve Group Estimates." TFSC <u>1</u> (3).

Dalkey, N. C., 1975. "An Elementary Cost Impact Model" in <u>The Delphi Method</u>. Ed. H. A. Lindstone and M. Turoff, Addison Wesely.

Dalkey, N. C., 1975. "Toward a Theory of Group Estimation," in <u>The Delphi Method</u>.

Davidson, D., P. Suppe, and S. Siegel, 1957. "<u>Decision Making: An Experimental Approach</u>, Stanford University Press.

Edwards, W., 1968. "Conservatism in Human Information Processing," in <u>Formal Representation of Human Judgment</u>, B. Kleinmonts (ed). Wiley.

Fusfeld, A. R., 1971. "The Delphi Technique: Survey and Comment," Sloan School of Management, Massachusetts Institute of Technology.

Fussel, J. B., 1972. "Synthetic Tree Model--A Formal Methodology for Fault Tree Construction." Ph.D. Thesis, Georgia Inst. Tech.

Gordon, T. J., R. Rochberg and S. Enzer, 1970. "Research on Cross Impact Techniques with Applications to Selected Problems in Economics, Political Science and Technology Assessment." Institute for the Future, R-12.

Grayson, C. J., 1960. <u>Decisions Under Uncertainty, Drilling Decisions by Oil & Gas Operators</u>. Howard Univ. Grad. Sch. Bus.

Hammond, K. R., and L. Adelman, 1976. "Science, Value and Human Judgment." Science Magazine, 194, pp. 389-396.

Hampton, J. M., P. G. Moore and H. Thomas, 1973. Subjective Probability and Its Measurement. Journal of the Royal Statistical Society, Sec. A., Vol. 136, pp. 21-42.

Lapp, S. A., and G. Powers, 1977. "Computer-Aided Synthesis of Fault Trees." IEEE Transactions in Reliability, April.

Lichtenstein, S., and J. R. Newman, 1967. Empirical Scaling of Common Verbal Phrases Associated with Numerical Probabilities. Psychon. Sci., Vol. 9.

Lipinski, A. J., H. M. Lipinski and R. H. Randolph, 1973. Computer-Assisted Expert Interrogation: A Report on Current Methods Development. Technological Forecasting and Social Change 5, pp. 3-18.

McGrath, J. E., and I. Altman, 1966. Small Group Research. Holt, Rinehart and Winston.

Morris, P. A., 1974. "Bayesian Expert Resolution." Ph.D. Thesis, Engineering Econ. Systems, Stanford Univ.

Morris, P. A., 1977. "Combining Expert Judgment: A Bayesian Approach." Mgt. Sci., 23: 679-693.

North, H. P., and D. L. Pyke, 1969. "Probes of the Technological Future." Harvard Business Review, v. 3: pp. 68-76.

Pfanzagl, J., 1971. Theory of Measurement. Springer Verlag.

Phillips, L. D., and W. Edwards, 1966. "Conservatism in a Simple Probability Inference Task," Journal of Experimental Psychology, No. 72: pp. 254-64.

Powers, G. J., and F. C. Thompkins, Jr., 1976. "Computer Aided Synthesis of Fault Trees for Complex Processing Systems," in Generic Techniques in Systems Reliability Assessment. Henley, E. J., and J. W. Lynn (ed.).

Powers, G. J., and F. C. Thompkins, Jr., 1974. "Computer Aided Fault Tree Synthesis for Chemical Processes." AIChE Convention, Philadelphia.

Pratt, J. W., H. Raiffa and R. Schlaifer, 1965. Introduction to Statistical Decision Theory (preliminary edition). McGraw-Hill, New York.

Quade, E. S., 1975. Policy Analysis for Public Decisions. Elsevier.

Robinson, W. O., 1971. "Multivariate Probability Assessment," Massachusetts Institute of Technology Operations Research Center Report TR 60.

Savage, L. J., 1954. Foundations of Statistics. Wiley.

Schlaifer, R. L., 1964. Computer Programs for Decision Analysis. Harvard Univ. Grad. Bus. Sch.

Schlaifer, R., 1969. Analysis of Decisions Under Uncertainty. McGraw-Hill, New York.

Selvidge, J., 1972. Assigning Probabilities to Rare Events. Ph.D. Thesis, Harvard Business School.

Sicherman, A., 1975. "An Interactive Computer Program for Assessing and Using Multiattribute Utility Functions." MIT OR Center, TR-111.

Sheridan, B., 1975. "Technology for Group Dialog and Social Choice," in The Delphi Method.

Spetzler, C. S., and C. von Holstein, 1975. Probability Encoding in Decision Analysis. Management Science, Vol. 22, No. 3.

Stael von Holstein, C-A. S., 1970. "Assessment and Evaluation of Subjective Probability Distribution." Stockholm: The Economic Research Institute at the Stockholm School of Economics.

Stanford Research Institute International, 1973. An Interactive Computer Interview: A Decision Analysis Technical Memorandum. SRI International, Menlo Park, California.

Stanford Research Institute International, 1977. Readings in Decision Analysis, Second Edition. SRI International, Menlo Park, California.

Stevens, S. S., 1975. Psychophysics. Wiley-Interscience.

Tani, S. N., P. Milgrom and C. von Holstein, 1977. Design of an Experirnent in Probability Encoding. Technical Memo. Stanford Research Institute.

Taylor, J. R., 1973. "A Formalization of Failure Mode Analysis of Central Systems." Danish Atomic Energy Commission. R150-M-1654.

Theil, H., 1971. Statistical Decomposition Analysis. Elsevier, Holland.

Torgerson, W. S., 1958. Theory and Methods of Scaling. McGraw-Hill.

Turoff, M., 1975. "An Alternative Approach to Cross Impact Analysis," in The Delphi Method.

Tversky, A., 1974. "Assessing Uncertainty." Jour. Roy. Stat. Soc., 36: pp. 148-160.

Tversky, A., and D. Kahneman, 1974. "Judgment Under Uncertainty: Heuristics and Biases," Science, No. 185: pp. 1124-31.

U. S. Geological Survey Circular 725. Estimation of Undiscovered Resources in the U. S. June, 1975.

Vesely, W. E. and R. E. Narum, 1970. "PREP and KITT: Computer Codes for Automatic Evaluation of a Fault Tree." Idaho Nuclear Report IN 1349.

Winkler, R. L., 1967. "The Assessment of Prior Distributions in Bayesian Analysis," Journal of the American Statistical Association, Vol. 62, pp. 776-800.

Winkler, R. L., 1968. "The Consensus of Subjective Probability Distributions." Mgt. Sci, 15.

APPENDIX

DERIVATION OF OPTIMAL WEIGHTS IN SYNTHESIS

$$\min_{w_i} E\left[(\hat{y} - \Sigma w_i x_i)^2\right] = \min_{w_i} \hat{y}^2 - 2\ \Sigma w_i \hat{y} E(x_i) + \Sigma\Sigma w_i w_j\ E(x_i x_j) \qquad \text{(A.4.1)}$$

subject to the constraint

$$\Sigma w_i = 1.0 \qquad \text{(A.4.2)}$$

Taking the derivation of A.4.1 with respect to w_i and equating to zero yields the n equations,

$$0 = \Sigma w_i\ (\mathrm{Cov}(x_i, x_j)) + \lambda, \quad i = 1, \ldots, n \qquad \text{(A.4.3)}$$

where λ is the Lagrange multiplier entering through the constraint. The (n + 1) st equation is the derivative w.r.t. λ,

$$0 = \Sigma w_i - 1 \qquad \text{(A.4.4)}$$

Solving in matrix format,

$$\left\{\begin{array}{c} \{ w_i \} \\ \hline \lambda \end{array}\right\} = \left\{\begin{array}{c} \underline{\emptyset} \\ \hline 1 \end{array}\right\} \left[\begin{array}{c|c} \underline{\underline{C}} & \begin{array}{c} 1 \\ \vdots \\ 1 \end{array} \\ \hline 1 \ldots 1 & \emptyset \end{array}\right]^{-1}$$

where C is the expert covariance matrix,

$$\underline{\underline{C}} = \begin{bmatrix} \mathrm{Cov}(\hat{x}_1, \hat{x}_1) & \cdots & \mathrm{Cov}(\hat{x}_1, \hat{x}_n) \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}(\hat{x}_n, \hat{x}_1) & \cdots & \mathrm{Cov}(\hat{x}_n, \hat{x}_n) \end{bmatrix} = \begin{bmatrix} \mathrm{Cov}(x_1, x_1) & \cdots & \mathrm{Cov}(x_1, x_n) \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}(x_n, x_1) & \cdots & \mathrm{Cov}(x_n, x_n) \end{bmatrix}$$

DISTRIBUTION

No. of
Copies

OFFSITE

UNITED STATES

A. A. Churm
DOE Chicago Patent Group
9800 South Cass Avenue
Argonne, IL 60439

27 DOE Technical Information
Center

2 Argonne National Laboratory
Reference Library
9800 South Cass Avenue
Argonne, IL 60439

20 Battelle Memorial Institute
Office of Nuclear Waste
Isolation
Attn: Beverly Rawles
505 King Avenue
Columbus, OH 43201

John Bredehoeft
U.S. Geological Survey
Reston, VA 22092

John Bird
Geogology Department
Cornell University
Ithaca, NY 14853

2 Brookhaven National Laboratory
Reference Section
Information Division
Upton, Long Island, NY 11973

Wayne A. Carbiener
Office of Nuclear Waste
Isolation
Battelle Memorial Institute
505 King Avenue
Columbus, OH 43201

No. of
Copies

Neville G. W. Cook
Dept of Materials Science and
Mineral Engineering
Hearst Mining Building
University of California
Berkeley, CA 94720

Carl R. Cooley
DOE Office of Nuclear Waste
Management
Washington, DC 20545

Environmental Protection Agency
Office of Radiation Programs
Technology Assessment Division
M/S AW559
Washington, DC 20460

Robert M. Garrells
Dept of Geologic Sciences
Northwestern University
Evanston, IL 60201

Colin A. Heath
DOE Division of Nuclear Waste
Management
Washington, DC 20545

William M. Hewitt
Office of Nuclear Waste Isolation
Battelle Memorial Institute
505 King Avenue
Columbus, OH 43201

Peter L. Hofmann
Office of Nuclear Waste Isolation
Battelle Memorial Institute
505 King Avenue
Columbus, OH 43201

Muzaffer Kehnemuyi
Office of Nuclear Waste Isolation
Battelle Memorial Institute
505 King Avenue
Columbus, OH 43201

No. of Copies	
	John F. Kircher Office of Nuclear Waste Isolation Battelle Memorial Institute 505 King Avenue Columbus, OH 43201
2	Lawrence Berkeley Laboratory Reference Library University of California Berkeley, CA 94720
2	Lawrence Livermore Laboratory Reference Library P.O. Box 808 Livermore, CA 94550
2	Los Alamos Scientific Laboratory Reference Library P.O. Box 1663 Los Alamos, NM 87544
	J. B. Martin Asst. Director for Radioactive Waste Management Branch NRC Division of Materials and Fuel Cycle Facility Licensing Washington, DC 20555
	John T. McGinnis Office of Nuclear Waste Isolation Battelle Memorial Institute 505 King Avenue Columbus, OH 43201
2	Barry N. Naft NUS Corporation 4 Research Place Rockville, MD 20805
2	J. O. Neff Department of Energy Columbus Program Office 505 King Avenue Columbus, OH 43201

No. of Copies	
2	Neil A. Norman Environmental Sciences Department Bechtel National Inc. P.O. Box 3965 San Francisco, CA 94105
2	Oak Ridge National Laboratory Central Research Library Document Reference Section Oak Ridge, TN 37830
	Frank L. Parker Dept. of Environmental Eng. Vanderbilt University Nashville, TN 37235
	George F. Pinder Dept. of Civil Engineering Princeton University Princeton, NJ 08540
2	Savannah River Laboratory Reference Library Aiken, SC 29801
	Howard P. Stephens Sandia Laboratories P.O. Box 5800 Albuquerque, NM 87115
	David B. Stewart U.S. Geological Survey National Center 959 Reston, VA 22092
	Robert Williams Electric Power Research Inst. 3412 Hillview Avenue P.O. Box 10412 Palo Alto, CA 94304
5	L. H. Wight Tera Corporation 2150 Shattuck Avenue Berkeley, CA 94704

No. of
Copies

FOREIGN

D'Allessandro Avogadro
Commission of European Communities
Joint Research Centre
I-21020 Ispra (Varese)
ITALY

V. K. Barwell
Environmental Research Branch
Atomic Energy of Canada Limited
Chalk River, Ontario KOJ1JO
CANADA

Center for Atomic Energy
Documentation (ZAED)
ATTN: Dr. Bell
Postfach 3640
D-7500 Karlsruhe
F.R. of GERMANY

Ferruccio Gera
Radiation Protection and Waste
Management Division
Nuclear Energy Agency/OECD
38 Boulevard Sucket
F-75016 Paris
FRANCE

K. H. Hubenthal
Bundesministerium fur Forschung
und Technologie
Stressemannstrasse 2
Postfach 200706
D-5300 Bonn
F.R. of GERMANY

2 INIS Clearinghouse
International Atomic Energy Agency
Karntner Ring 11
P.O. Box 590
A-1011 Vienna
AUSTRIA

No of
Copies

Klaus Kuhn
Institut fur Tiefagerung
Wissenschaftliche Abteilung
Berliner Strasse 2
D-3392 Clausthal - Zellerfeld
F.R. of GERMANY

Hans W. Levi
Hahn-Meitner-Institut fur
Kernforschung
Glienicker Strasse 100
D-1000 Berlin 39
F.R. of GERMANY

Library
Studsvik Energiteknik AB
S-611 01 Nykoping
SWEDEN

Franz Peter Oesterle
Physikalisch-Chemische
Bundesanstalt
Bundesallee 100
D-3300 Braunschweig
F.R. of GERMANY

Tönis Päpp
Kärnbränslesäkerhet
Fack. 10240
Stockholm
SWEDEN

Tjalle Vandergraff
Atomic Energy of Canada Limited
Whiteshell Nuclear Research
Establishment
Tinawa, Manitoba ROE 1LQ
CANADA

Glenn Merrett
Storage and Disposal Branch
Whiteshell Nuclear Research
Establishment
Tinawa, Manitoba ROE 1LQ
CANADA

No. of
Copies

Egbert Schapermeier
Battelle-Institute e.V.
Am Romerhof 35
D-6000 Frankfurt am Main 90
F.R. of GERMANY

ONSITE

4 DOE Richland Operations Office

O. J. Elgert
H. E. Ransom
J. J. Schreiber
F. R. Standerfer

2 Rockwell Hanford Operations

R. A. Deju
Rockwell Document Control

45 Pacific Northwest Laboratory

A. Brandstetter (20)
F. H. Dove
M. A. Harwell
J. S. Light
G. M. Petrie
A. M. Platt
R. J. Serne
J. T. Zellmer
Technical Information Library (5)
Publishing Coordination (2)
Water and Land Resrouces Dept.
Library (10)