OCT 30 1989

MEMORANDUM FOR: Mark A. Cunningham, Branch Chief
Probabilistic Risk Analysis Branch
Division of Systems Research, RES

FROM: Ronald L. Ballard, Chief
Geosciences & Systems Performance Branch
Division of High-Level Waste Management, NMSS

SUBJECT: REQUEST FOR REVIEW OF SANDIA REPORT ON EXPERT JUDGMENT

I have enclosed a copy of the draft report submitted by Sandia National Laboratories entitled "Elicitation and Use of Expert Judgment in Performance Assessment for High-Level Waste Repositories," prepared under contract FIN A-1165. Although Lee Abramson has already provided review comments, P.K. Niyogi has also been following the development of this report. Any additional review comments are needed by November 7, 1989. We appreciate your cooperation in this matter.

Ronald L. Ballard, Chief
Geosciences & Systems Performance Branch
Division of High-Level Waste Management, NMSS

Enclosure:
As stated

cc: L. Abramson, RES
P.K. Niyogi, RES

DISTRIBUTION:

Central Files
JOBunting, HLEN
PDR & LPDR
PBrooks, HLGP
REBrowning, DHLWM
JLinehan, HLPM
CNWRA
SCoplan, HLGP
BJYoungblood, DHLWM
HLGP r/f
RLBallard, HLGP
NMSS r/f
ACNW

OFC	HLGP	HLGP	HLGP			
NAME	PBrooks/cj	SCoplan	RLBallard		NH 8/1	
DATE	10/26/89	10/27/89	10/27/89			

delete T-35

DRAFT

NUREG/CR-5411
SAND89-1821

ELICITATION AND USE OF EXPERT JUDGMENT IN PERFORMANCE ASSESSMENT FOR HIGH-LEVEL RADIOACTIVE WASTE REPOSITORIES

Evaristo J. Bonano
Stephen C. Hora[1]
Ralph L. Keeney[2]
Detlof von Winterfeldt[2]

July 1989

Sandia National Laboratories
Albuquerque, NM 87185
Operated by Sandia Corporation
for the
U.S. Department of Energy

Prepared for
Division of High-Level Waste Management
Office of Nuclear Material Safety and Safeguards
U.S. Nuclear Regulatory Commission
Washington, DC 20555
NRC FIN A1165

---

[1] University of Hawaii, Hilo, HI
[2] University of Southern California, Los Angeles, CA

DRAFT

ABSTRACT

This report presents the concept of formalizing the elicitation and use of expert judgment in the performance assessment of high-level radioactive waste repositories in deep geologic formations. The report outlines aspects of performance assessment in which the elicitation and use of expert judgment should be formalized, discusses existing techniques for formalizing the elicitation and use of expert judgment, and presents guidelines for applying these techniques in performance assessment.

DRAFT

CONTENTS

DRAFT

CONTENTS (Continued)

DRAFT

CONTENTS (Continued)

DRAFT

CONTENTS (Concluded)

DRAFT

LIST OF TABLES

DRAFT

FOREWORD

This report presents the concept of formalizing the elicitation and use of expert judgment in the performance assessment of high-level radioactive waste repositories in deep geologic formations. The report outlines aspects of performance assessment in which the elicitation and use of expert judgment should be formalized, discusses existing techniques for formalizing the elicitation and use of expert judgment, and presents guidelines for applying these techniques in performance assessment.

DRAFT

1. INTRODUCTION

The use of expert judgment permeates all scientific inquiry and decision making. The choice is not whether to use expert judgment, but whether to use it in an explicit and disciplined manner or in an ad hoc manner. For significant technical, environmental, and socioeconomic problems, it is often useful to formalize the elicitation and use of expert judgment. One such problem is the long-term disposal of high-level radioactive waste (HLW) in repositories mined into deep geologic formations. In siting and designing a safe, environmentally sound, and legally acceptable repository, many of the analyses must use expert judgment.

The Environmental Protection Agency (EPA) has mandated quantitative analyses in its Standard 40 CFR Part 191 for the disposal of spent nuclear fuel and high-level and transuranic radioactive wastes. In particular, the EPA requires a so-called "performance assessment" in the containment requirement of this standard. (The other requirements are individual and groundwater protection requirements that concern only the undisturbed behavior of the repository system.) Performance assessment refers to "quantitative analyses that (1) identify the processes and events that might affect the disposal system; (2) examine the effects of these processes and events on the performance of the disposal system; and (3) estimate the cumulative releases of radionuclides, considering the associated uncertainties, caused by all significant processes and events" (EPA, 1985). EPA further requires that performance-assessment estimates be represented by an overall probability distribution of cumulative releases. Furthermore, these probability distributions are to be used to determine whether the release standards in 40 CFR Part 191 are met. The Nuclear Regulatory Commission (NRC) has been charged with implementing this standard and examines the quality of a performance assessment when evaluating a license submitted by the Department of Energy (DOE) to construct and operate an HLW repository.

Obviously expert judgment is extensively used in any responsible analysis of potential health impacts from a repository and particularly in performance assessments. Expert judgment is required in identifying and screening events and scenarios, in developing and selecting models that characterize the geology and hydrology of the repository system, in assessing model parameters, in collecting data, and in making strategic decisions about the repository that could affect its performance. While it is desirable to use data and modeling extensively in performance assessment, it is nevertheless clear that these data and models can never substitute for the many crucial expert judgments in the assessment.

The quality of a performance assessment rests on its foundation of expert judgments. Consequently, to demonstrate that an HLW repository meets regulatory requirements, all significant expert judgments should be documented and supported with sound logic and the best information. This is particularly important because of the need for multiple scientific disciplines to address the long-term disposal of HLW and because of the intense scrutiny that all decisions will likely receive. Responsibility and accountability can be enhanced by a formal elicitation and use of judgment, which is a well-documented, systematic process whereby experts make inferences or evaluations about a problem using available information as well as accepted scientific methods. This allows for traceability of the procedures, techniques, and methods, assumptions, and physical principles relied on in any inferences or evaluations.

1.1 Objective of this Report

This report discusses the formal elicitation and use of expert judgment in performance assessment of HLW disposal systems. More specifically, professional knowledge about the analysis of HLW disposal systems and about the elicitation and use of expert judgment is combined to develop insights on the formalization of expert judgments applicable to HLW repositories. The report (1) discusses the role of expert judgment in performance assessment of HLW repositories, (2) identifies areas needing formal expert judgment in HLW disposal, (3) describes the formal elicitation and communication of expert judgment, and (4) provides suggestions for the use of expert judgment in HLW disposal.

1.2 Expert Judgment in Performance Assessment of HLW Repositories

Experts are used to design and implement activities to understand present site conditions and predict the behavior of the disposal system. Expert judgment will be used in (1) setting priorities for data collection, (2) designing site data-collection activities, (3) determining the level of resources for reduction of uncertainties, (4) quantifying the uncertainty in numerical values for key parameters, (5) developing scenarios and assigning corresponding probabilities of occurrence, and (6) formulating approaches for validating conceptual and mathematical models as well as verifying computer codes. These important tasks need to be addressed before using models and computer codes to predict behavior of the disposal system. Expert judgment is also used with the models and codes to estimate the system's performance for comparison with the numerical criteria in the regulations. For example, expert judgment is required to screen insignificant scenarios, select methods for propagating uncertainty through the models and codes, quantify uncertainty in the predictions, and interpret results.

1.3 Characteristics of a Formalized Expert-Judgment Process

A formal expert-judgment process has a predetermined structure for the collection, processing, and documentation of experts' knowledge. As discussed in Chapter 3, this includes professionally designed procedures to select problem areas and experts and to train experts for the elicitation of their judgments. The actual elicitations of judgments should involve the expert and a professionally trained person to assist the expert in expressing judgments. The elicited judgments and their rationales should be carefully documented.

There are advantages and drawbacks in using such a process. The advantages include the following:

Improved Accuracy of Expert Judgments. The methods in a formal expert elicitation process improve the accuracy and reliability of the resulting information over less structured methods (Lichtenstein, Fischhoff, and Phillips, 1977 and 1982; Lichtenstein and Fischhoff, 1980; Fischhoff, 1982). This is so because psychological biases are openly dealt with, problems are defined and communication is improved (Merkhofer, 1987), issues are systematically analyzed, and rationales and results are documented. The level of expertise may also be improved over less structured methods since a formal process encourages a broadening of the range of expertise. Experts are carefully selected in a formal process rather than in a haphazard manner for reasons of convenience.

Well-Thought-Through Design for Elicitation. The procedures that will be used in a formal expert-judgment process are designed specially for the problem being faced. The design relies on the knowledge concerning expert opinion, previous studies that have used formal expert judgment, and knowledge of the problem domain to be studied. Careful planning of the process can substantially reduce the likelihood of critical mistakes that will render information suspect or biased. Mistakes such as including experts with motivational biases, failing to document rationales, inadvertently influencing the experts' responses, failing to check for consistency, and allowing individuals to dominate group interactions can be avoided.

Consistency of Procedures. A formal expert-judgment process enhances consistency and comparability of procedures throughout a study and across related studies because participants follow the same procedures. On the other hand, informal processes are often subject to the whims and desires of participants.

Scrutability. A formal process requires the establishment and dissemination of rules and procedures for elicitation and use of expert judgment. A normal part of a formal expert-judgment process is the documentation of procedures and assessments, which helps to ensure that various reviewers and users of the findings can understand and evaluate the methods and insights of the study. Since the methodology and its implementation are transparent, there is accountability.

Communication. Establishing a formal process helps to provide for reference documents useful in communication and external review. A formal process also encourages communication and understanding among experts and analysts about the problems studied and the values assessed.

Less Delay. Projects have been delayed because critical judgments were not carefully obtained or documented, and a formal expert-judgment process had to be designed and conducted before the project moved forward (DOE, 1986). A well-executed formal process would have avoided costly delays.

There are also drawbacks to the formal expert-judgment process:

Resources. There are costs in designing and implementing a formal process. Documentation is often more extensive with a formal process, and more resources are thus required.

Time. The time to establish and implement a formal process may be significantly greater than that required for an informal process. Scheduling of participants from external organizations adds a layer to the effort that is not present in an internal, informal process.

Reduced Flexibility. Formalization of the process may reduce flexibility and make on-going changes to the study more difficult. If it is necessary to redo part of a study, reenacting the expert-judgment process may be cumbersome and expensive.

Vulnerability to Criticism. The transparency of a formal process and the documentation of procedures and findings open it to inspection and criticism. Expert judgment is an area in which misunderstanding of the methods and aims still exists, but a carefully designed and implemented process may thwart such criticisms.

While a formal process often requires more resources and time than an informal process initially requires, a faulty process that fails to withstand criticism or must be redone because of inappropriate design or improper execution may end up failing to satisfy the project's objectives and cost more in both time and resources. The potential for further costs in an informal study should be considered when evaluating the need for a formal process.

Formalizing the elicitation of expert judgments can clearly be expensive and time consuming. For this reason, the areas in which the process should be used should be carefully selected. It is neither practical nor reasonable to formalize the use of expert judgments in all aspects of HLW repository performance assessment.

1.4 Previous Formal Uses of Expert Judgments in HLW Program

Several studies involved the formal elicitation and use of expert judgment on important problems facing the HLW program. Recent studies relevant to performance assessment analysis of HLW repositories are outlined here. In Chapter 2, five areas in need of formal expert judgments in HLW disposal are described: scenario development and screening, model development, parameter estimation, information gathering (e.g., data collection and experiments), and strategic repository decisions. Collectively, the analyses outlined here address problems in all five areas.

The Draft Environmental Assessment for the Hanford site in Washington State (DOE, 1984), reports an analysis that screened candidate horizons and identified a preferred horizon. A multidisciplinary team developed a set of eight measures to rank the horizons. These measures involved repository performance, construction ease, and costs. Deterministic and probabilistic descriptions of the candidate horizons were developed using the eight measures. The probabilistic descriptions were probability distributions based on analytical models, available scientific data, and explicit assessment of expert judgments. Because none of the candidate horizons dominated the others, a utility function was also assessed, using value judgments of the interdisciplinary team to combine the measures. The horizon descriptions were then evaluated using the utility function to rank the candidate horizons.

At the Hanford site, the formal elicitation and quantification of expert judgment helped in designing an underground test facility (Golder and Associates, 1986). To estimate groundwater and methane gas flow into the proposed test facility, estimates of site-specific geologic, hydrologic, and dissolved gas parameters were obtained. Specifically, probability distributions were assessed for 41 parameters pertaining to flow path length, timing of encounters with geologic features, and transmissivity and storativity of the geologic surroundings near the test facility. The entire elicitation exercise included developing an influence diagram to help identify parameters to be assessed, identifying a panel of experts to be assessed, and conducting training sessions on probability elicitation for the panel of experts before the elicitation sessions.

Formal elicitation of expert judgment was extensively used in a multiattribute decision analysis comparing horizontal and vertical emplacement modes for casks of spent nuclear fuel in a salt repository (Fluor Technology, Inc., 1988). First, 10 attributes covering health and safety, cost, and environmental concerns were selected. An influence diagram related several variables to these attributes. Expert judgment was elicited to provide probability distributions for both emplacement modes for some of the variables. Deterministic estimates were obtained for others. These estimates were input into a simulation model to describe the emplacement modes in terms of the attributes. A utility function was then assessed using the value judgments of a Fluor employee to evaluate alternatives.

The Department of Energy, following a recommendation of the Board on Radioactive Waste Management of the National Academy of Sciences, chose multiattribute utility analysis (MUA) as the methodology to rank five potential host sites for an HLW repository in the United States. The analysis (DOE, 1986) provided part of the information to reduce the number of possible host sites to three. In the MUA, two different types of experts were used. One type was senior managers of DOE who provided value judgments about risk attitudes and value tradeoffs among the objectives of the study. The second type were specialists in one or more of the technical areas needed to assess repository performance. These technical experts were divided into six panels addressing economic costs, environmental impacts, social impacts, transportation of waste, repository construction, and postclosure considerations. The technical experts were asked to develop measures of repository performance for both the preclosure and postclosure phases of HLW disposal; formulate scenarios for the postclosure phase; screen the scenarios to eliminate those that did not apply to particular sites; quantify the likelihood of each scenario occurring during the first 10,000 years after repository closure; estimate radionuclide discharge to the accessible environment in 10,000 years for each scenario; and finally, decide on the performance of each potential site for each of the performance measures (Merkhofer and Keeney, 1987).

The Board on Radioactive Waste Management reviewed the methods used in the multiattribute utility analysis of potential repository sites. As part of its review, the Board stated (Appendix H, DOE, 1986):

> While recognizing that there is no single, generally accepted procedure for integrating technical, economic, environmental, socioeconomic, and health and safety issues for ranking sites, the Board believes that the multiattribute utility method used by DOE is a satisfactory and appropriate decision-aiding tool. The multiattribute utility method is a useful approach for stating clearly and systematically the assumptions, judgments, preferences, and tradeoffs that must go into a siting decision.

In addition, the expert judgments and methods in this report were publicly scrutinized by peer review (Gregory and Lichtenstein, 1987).

A subsequent analysis was based on the same expert judgments elicited for the multiattribute utility siting study. Because the Nuclear Waste Policy Act of 1982 stated that three sites should be characterized, Keeney (1987) analyzed portfolios of three sites for simultaneous characterization and strategies for sequential characterization. Based on 1986 characterization costs estimated to be $1 billion per site, sequential characterization strategies were identified that could save $1.7 to $2.0 billion compared with simultaneous characterization of the three sites chosen by the DOE. This portfolio analysis and the multiattribute utility siting analysis provided insights used by Congress in designing the Nuclear Waste Policy Act Amendments Act of 1987 that eliminated the simultaneous characterization of three sites and chose Yucca Mountain, Nevada, as the planned repository site.

Merkhofer and Runchal (1989) summarized a study to quantify judgmental uncertainty in values of hydrologic parameters at a repository site. Specifically, experts obtained cumulative density functions (cdfs) for the values of (1) effective porosity, (2) average effective porosity, and (3) anisotropy ratio at the Hanford site. Two different groups of technical experts were used in the study. One group was five well-known hydrologists not directly involved with the site investigations at Hanford but, nevertheless, familiar with waste-disposal issues. The second group was three hydrologists involved in the characterization of the site. The probability elicitation process utilized structured interviews between a trained interviewer and each of the experts. The interviews consisted of five phases: motivating, structuring, conditioning, encoding, and verifying (Stael von Holstein and Matheson, 1979). To reduce the differences in judgments between the experts, all the results of the original assessments were anonymously exchanged, as suggested by the original Delphi method (Dalkey and Helmer, 1963). The revised probabilities showed at most only minor revisions; even though there was a considerable diversity of opinion. The experts indicated that any substantial changes would occur only after the exchange of logic and data by the experts.

HLW repository operation requires the transport of waste from nuclear power plants to the repository. A study by Westinghouse Electric Corporation developed a set of objectives for evaluating spent nuclear fuel transport explicitly using the judgments of experts (Westinghouse Electric Corporation, 1986). To establish a comprehensive set of objectives, three panels with individuals in the nuclear industry, state governments, and public interest organizations were guided through sessions to create and structure objectives. Structured objectives of the three panels were combined into one hierarchy for review. These objectives concerned health and safety and economic, environmental, political, social, and equity considerations as well as scheduling and flexibility. The results were a basis for further analysis and communication among interested parties. The process of eliciting the objectives and the results is found in Keeney (1988b).

These studies clearly indicate that experts have been and will be used in a variety of ways to address critical issues relevant to the long-term disposal of HLW in repositories mined in deep geologic formations. In some cases, the experts provided quantitative assessments (e.g., quantification of the uncertainty about a parameter, or the likelihood of a scenario occurring); in other cases, they addressed qualitative identification and screening problems (e.g., selection of appropriate measures of repository performance, formulation and screening of postclosure scenarios); and in still other cases, they provided value judgments (e.g., attitudes toward risk and value tradeoffs). The fundamental concepts in the formal elicitation and use of expert judgment are generic and independent of the type of issue the experts address. However, the choice of specific techniques during the elicitation process and the way the judgments are used to address a problem should be issue-specific.

1.5 When to Use Expert Judgment

Formal methods should be used whenever the benefits are greater than the costs. Indicators of when the elicitation of expert judgments should be formalized are as follows:

Lack of Data. When extensive, noncontroversial data directly relevant to a problem is lacking, existing data must be supplemented with expert judgments, and it may be worthwhile to obtain them using a formal elicitation process.

Importance of the Issues. Formal methods are most appropriate when the expert judgments will have a major impact on the study and improvements in the quality of the judgments are then most worthwhile. Important issues also draw the most scrutiny. A formal methodology promotes documentation and communication and should be employed when the issue studied is apt to receive extensive review and criticism or when the findings will be widely disseminated.

Complexity of the Issues. When a problem is complex, or when several experts are employed either redundantly or as a team, formal methods are appropriate. These methods can provide the structure so that all participants understand the methods used and apply procedures consistently.

Level of Documentation Required. Formal methods are a vehicle to obtain complete and consistent documentation of the methods and the findings. Informal methods often produce documentation that is incomplete with regard to the assumptions and procedures used. The critical reviews that the study will undergo, the variety and types of users, and the uses of the information may also suggest whether a formal process should be instituted. In some studies, the expert judgments may be important findings and, perhaps, used in subsequent studies, so formal methods are needed.

Extent of the Use of Expert Opinion. When expert judgments are used extensively in a study, formalization of the collection and processing of that information is apt to be done most accurately, consistently, and efficiently using formal methods. Costs that are fixed regardless of the size of the effort, such as creation of forms, training, etc., may be spread over many assessments. Also, when similar assessments are to be made by various experts, formalization of the procedures is necessary for consistency.

1.6 Relationship of Formal Use of Expert Judgment to Informal Use, Modeling, and Data Collection

As stated in the introduction, expert judgment enters performance assessments in many places. The question is therefore not whether to use expert judgment, but whether to use it formally or informally, and how to use it with other sources of information like basic physical principles, models, and data.

Informal use of expert judgment means implicit and undocumented use. Given the cost of formal expert judgment, this may be reasonable in many instances in performance assessment. In some cases, "semi-formal" uses may be advocated, such as brainstorming and/or taped group discussions about the issues. In such cases, it is important to identify carefully the objectives of the use of expert judgment and to be sure that its benefits outweigh its costs. Documentation is still important in semi-formal uses of expert judgments, because complex interactions may be involved.

Peer review should not be confused with the formal use of expert judgment. Both peer review and formal expert judgment are explicit and documented processes to increase the likelihood that a resolution of an issue is of highest quality. However, the formal use of expert judgment attempts to bring out the available information that bears on the problem as part of its solution, while peer review evaluates and criticizes a given approach and solution to a problem. It should be noted that formal use of expert judgment can, and often should, be subject to peer review. Thus, these processes are compatible.

When formal expert judgment is used, a question arises about how it relates to other activities such as collecting data or modeling phenomena and processes. A simple answer is that any of these means of obtaining and quantifying information should be used in a cost-effective mix that solves the particular problem. In addition, formal expert judgment can often be beneficial in integrating diverse sets of data and modeling activities and results. Thus, expert judgment and data collection and modeling activities should never be seen as substitutes, but as complements.

To contrast formal expert judgment to data collection and modeling, consider its favorable and unfavorable properties (Einhorn, 1972). Expert judgment is a flexible and general source of information. A formal expert-judgment process is unique in that it can readily incorporate many disparate pieces of information into a coherent evaluation. Formal expert judgment, though, does not possess some properties of well-behaved experimental/statistical data. For example, increasing the number of experts whose judgments are collected does not ensure that the "average" judgment will somehow converge to the true value. Nor can the usual assumption of independence and the assumption of convenient underlying distributions be called forth for use in expert-judgment processes as they often are in the analysis of experimental data. It should be noted, however, that in most complex problems experimental/statistical data are not well behaved in this respect, either.

Formal expert judgments will not be as precise and clear as computer or mathematical models. However, these models build on expert judgment and may also suffer from the same limitations. Models that do not account for unforeseen factors or ignore potentially important variables fail in the same way that expert judgment fails when an expert or group of experts do not properly recognize or account for all important factors.

DRAFT

2. AREAS IN NEED OF FORMAL EXPERT JUDGMENT IN HLW DISPOSAL

Expert judgment has been used and will be used in many aspects of performance assessment as well as in other analyses, evaluations, and decisions related to HLW disposal. However, it may not be useful to formalize all expert judgments. As discussed in Section 1.3, there are m[illegible] advantages and drawbacks to formal expert judgment, and consequently, the decision of when to use it has to carefully consider benefits against costs.

In this chapter five areas of performance assessment in HWL repositories are discussed for which the benefits of formal expert judgment may outweigh its costs. These five areas are (1) scenario development and screening, (2) model development, (3) parameter estimation, (4) data collection and experimentation, and (5) strategic repository decisions. This chapter does not describe these areas in a comprehensive manner, but rather highlights those aspects in which expert judgment is likely to be formalized. It should be noted that some of these areas are described in detail elsewhere (Cranwell et al., 1989; Bonano and Cranwell, 1988).

2.1 Scenario Development and Screening

To carry out a comprehensive performance assessment of the possible releases of radionuclides to the environment and to obtain probabilistic assessments of these releases and the resulting health effects, an analysis should consider the possible future states of the repository as influenced, for example, by climatic, geologic, and hydrologic changes in the natural repository environment as well as by changes in the physical and chemical characteristics of the man-made repository system. Recognizing this need to consider the repository system and its changes comprehensively, both the NRC (1983) and the EPA (1985) require that all physically plausible events and processes be considered in a performance assessment. In this context, events are discrete changes in the evolving states of the repository system, while processes are continuous and coherently linked changes.

Cranwell et al. (1989) describe a methodology developed by Sandia National Laboratories (SNL) for the selection and screening of scenarios. This methodology was developed for the NRC and is currently used by a number of countries in their nuclear-waste disposal programs. (Scenario Working Group, Nuclear Energy Agency, Organization for Economic Cooperation and Development, Paris, France.) Although other approaches with a slightly different focus are being developed (Thompson et al., 1988), DOE is also expected to use the scenario approach in performance assessment analysis of an HLW repository at Yucca Mountain. Scenario selection and screening involves (1) initial identification of plausible events and processes, (2) classification of events and processes, (3) initial screening out of unimportant events and processes, (4) combining of important events and processes into scenarios, and (5) screening of scenarios to arrive at a final set for consequence analysis. Both for screening and for subsequent analysis, each scenario is assigned a probability of occurrence during the regulatory period (i.e., 10,000 years). Expert judgment is used in all steps of scenario selection and screening and in the estimation of probability of occurrence of scenarios as summarized below.

2.1.1 Identification of Events and Processes

The initial listing of physically plausible events and processes is a creative task that depends almost exclusively on expert judgment. There is no widely accepted method for arriving at this list, and there is no method for ensuring that all potentially significant events and processes are included in the initial list (except by defining a category like "none of the above" and thereby ensuring completeness). Formalizing expert judgment is one means of decreasing the likelihood that important events and processes have been omitted. Formalized expert judgment is likely to be more useful than ad hoc methods because it draws on a variety of experts, and because it is documented it can be scrutinized by many individuals and groups interested in including events and processes that they consider signific[illegible]

2.1.2 Classification of Events and Processes

For completeness and organizational purposes, events and processes are often classified as naturally occurring, human induced, and repository induced. Often, the events and processes are classified as affecting either the release of radionuclides from the repository to the geosphere or affecting the migration of radionuclides through the geosphere. Expert judgment combined with principles of groundwater flow and transport phenomena is used to classify events and processes.

2.1.3 Screening of Events and Processes

The initial list of events and processes is often generic. Thus, the list should, in principle, be shortened on a site-specific basis. That is, events and processes must be screened for each site. The NRC (NRC, 1983, 1988) suggests to classify the events and processes into

- Anticipated Events and Proc[illegible] Natural geological events and processes presently occurring or known [illegible] occurred during the Quaternary Period (1.8 million years ago to the prese[illegible] addition, one may want to consider natural events and processes that are not presently taking place but may be anticipated sometime in the future.

- Unanticipated Events and Processes - Natural and human-induced events and processes that are not likely during the 10,000-year regulatory period but are sufficiently credible that they cannot be ignored.

- Not Credible Events and Processes - Events and processes outside the other two categories.

Anticipated events and processes and unanticipated events and processes, according to the NRC (NRC, 1988), must be considered in the development of scenarios for a performance assessment to demonstrate compliance with the containment requirement of EPA Standard 40 CFR Part 191.13 (EPA, 1985) and the NRC Rule 10 CFR Part 60.113 (NRC, 1983). Events and processes that are not credible can be eliminated from further consideration. Classifying events and processes into these categories depends on the experts' interpretation of historical records, site-characterization information, and conceptualizations of the future of the repository and even of human behavior. This interpretation will, in turn, depend on a given expert's technical background and may depend on the information base and approach to the problem. Some aspects of the classification can be highly speculative because the meaning and interpretation of information depend on how an expert visualized the evolution of the system. In addition, the screening process depends on the expert's definition of "credible."

2.1.4 Formulation of Scenarios

Scenarios are formulated from all possible combinations of events and processes remaining after screening. Typically, an event tree is used to generate all possible combinations of events and processes. The procedure is straightforward if the initial list of events and processes is fairly complete and potentially significant events and processes have not been screened out. While this can, in principle, be done mechanically, expert judgment is needed to prune first-cut event trees and to check their consistency and completeness. The formulation of scenarios can also be done using fault trees by working backwards from potentially important future state(s) of the disposal system and relating these outcomes to possible causes. Expert judgment is needed in identifying the states and in deriving common causes of sets of events. In most cases, both event trees and fault trees should be used.

2.1.5 Screening of Scenarios

An initial screening of scenarios is based on (1) physical reasonableness, which eliminates physically impossible or implausible combinations of events and processes, (2) the consequence of scenarios, which eliminates those with little or no impact on repository performance, and (3) likelihood of occurrence. In this manner, the number of scenarios can be reduced. Expert judgments play an important role in this preliminary screening by developing criteria for screening and applying them.

2.1.6 Probability of Occurrence

Probabilities need to be assigned to scenarios for two reasons: to disregard from further consideration scenarios less likely than the screening criterion and to quantify the likelihoods of remaining scenarios to estimate cumulative radionuclide releases and health effects.

Expert judgment plays a significant role in estimating probabilities of occurrence for scenarios. Ideally, some historical data exist for a given site on climatic changes, seismic activity, volcanic activity, human intrusion, etc., that can be used to formulate models and provide input used to predict the evolution of the site (a similar approach to the global modeling advocated by Thompson et al., 1988). Expert judgment is used to interpret the data, estimate the numerical values of model parameters, and, finally, to interpret the results of simulations and arrive at probability estimates. More realistically, data are likely to be scarce. Data for some phenomena (e.g., human intrusion) may not exist or models may be nonexistent or inadequate. Expert judgment is then the main basis for estimating probability.

The probability of occurrence of the scenario is a combination of the probabilities of its individual events and processes. Expert judgment plays a major role not only in determining the probability of the events and processes, but also in the way these probabilities are combined to arrive at the probability of the scenario. For example, experts are likely to be used to decide whether a scenario's events and processes occur in a sequence and, if this is so, to determine the sequence.

2.2 Model Development

In a performance assessment, assumptions and simplifications are made about the behavior of the repository system that can be incorporated into a "conceptual model" for mathematical simulation of system behavior.

Conceptual modeling of an HLW disposal site is based on a combination of the application of physical principles and data interpretation. Once the models have been developed using whatever information or data are available (e.g., from small-scale, short-term experiments), confidence must be built that the models are adequate to predict the behavior of the system over much larger spatial and temporal scales. Both the development of conceptual models and confidence building are creative and interpretative activities that are largely founded on expert judgment.

2.2.1 Data Selection and Interpretation

Model development is based on limited, site-specific information about the system geometry, past and active processes, and potential disrupting processes and events. Little or no data will be available to determine all of these factors at the proposed repository location. Therefore, experts select and interpret data from similar sites and relate them to the repository site. Interpretations of scant geologic data are used to define the system geometry. Experts must infer such things as the geologic continuity between drill holes, the extent and thickness of units, and the extent and character of geologic discontinuities such as faults. The geometry defined by these experts is based not only on interpolation and extrapolation of the site-specific data, but on data from similar geologic environments. Many processes are active in the geosphere (i.e., water flow, vapor flow, heat flow, etc.). Experts select and interpret data to decide which processes to consider in assessing the performance of a repository system. Not only do the experts have to decide the current dominant processes, but they must predict future processes that could adversely affect the repository system. This later assessment requires the experts to identify and interpret data from similar systems (i.e., analogs to the future states of the repository). Direct measurements of system performance (i.e., integrated discharge over 10,000 years) will never be available, so inferences about the possible system behavior and the accuracy of system models are from indirect site measurements and from information about similar systems.

2.2.2 Development of Conceptual Models

Data cannot be collected over the temporal and spatial scales of interest in performance assessments of HLW repositories, so considerable data interpretation is required to formulate conceptual models. Because the conceptual model is the foundation of the mathematical models, computer codes, and data collection supporting performance assessment and because its development relies so heavily on expert judgment, formalized expert judgment could be most beneficial in modeling. A conceptual model includes simplifications and assumptions about (1) the geometry of the system, (2) the current or future physiochemical processes, (3) the boundary and initial conditions, and (4) the parameters governing these processes.

DRAFT

The most common approach to conceptual modeling begins with a rough sketch of the model and continues to refine that sketch based on whatever experimental data and other information are available until an adequate first-cut model is produced. Typically, this is done by using one expert's judgment and interpretations of experimental data and other information. To make conceptual modeling more comprehensive and to encourage considerations of alternative models as well as scrutability of the experts' reasoning, Bonano and Cranwell (1988) suggest an approach for formalizing the use of expert judgment with multiple experts well versed on the groundwater flow and transport models. The approach forces the experts to articulate all assumptions, and to look for interpretations that challenge their conventional wisdom and are consistent with available data. The second point could lead to alternative conceptual models. Finally, the approach could include procedures for allowing the experts to identify bounding analyses and experimental investigations aimed at distinguishing between alternate conceptualizations and eventually reducing their number.

2.2.3 Confidence Building

After conceptual models for the disposal system have been assembled, appropriate mathematical models and computer codes must be developed to simulate the behavior of the system over the spatial and temporal scales prescribed by the regulations (5 km and 10,000 years).

Experts are an integral part of limited-scope activities to build confidence in models and codes. For example, international groups have been formed such as INTRACOIN, HYDROCOIN, and INTRAVAL to select problems of common interest to the radioactive waste-management community. These are simulated by interested parties, and the results are compared. These groups attempt to find discrepancies among the results from different experts and their causes. One important result is that the group may implicitly or explicitly agree that, given the current state of the art, existing models and codes are as good as they can be. To date, these groups (specifically, INTRACOIN and HYDROCOIN) have focused on benchmarking activities that are an aspect of "code verification."* The recently started INTRAVAL program goes one step further in that it aims at "validating" conceptual models, mathematical models, and computer codes.**

Validation means comparing the predictions of the models to experimental results. Because the models' predictive capabilities cannot be fully tested, "true" validation can never be achieved. The alternative is to build confidence in the models and codes through a synthesis of experiments and calculations. Experiments are likely to include laboratory and controlled-field investigations as well as natural analogs. Calculations could consist of bounding analyses and preliminary overall-system

---

*Verification is defined by the NRC as the "process of obtaining assurance that a given computer code implements the solution of a given mathematical model."

**NRC defines validation as the "process by which assurance is obtained that a model as embodied in a computer code is an accurate representation of the process or system for which the model is intended."

performance assessments. In any case, experts (1) design experiments and calculations, (2) establish the validity and limitations of these experiments and calculations, (3) define appropriate measures to ascertain the predictive capabilities of the models and codes, (4) ascertain the validity of important couplings in the models that cannot be tested, (5) interpret the results of model runs against existing and new data, and (6) judge the ability of the models to extrapolate to large temporal and spatial scales.

2.3 Parameter Estimation

Performance-assessment predictions depend on the numerical values of the parameters used by their models and codes. Selecting appropriate numerical values for parameters and quantifying the uncertainty about them is a difficult but important aspect of performance assessment. First, important parameters must be identified, and then uncertainty in their values quantified. Expert judgment is important in both of these aspects, as discussed below.

It might be worthwhile to define the terms "parameter" and "data." Parameters are coefficients or constants of models and processes that describe or control the behavior of a model. Coefficients refer to the proportionality constants such as hydraulic conductivity and diffusivity needed in rate equations such as Darcy's law and Fick's law, respectively, and to the mean and standard deviation of a probability distribution. Data are values taken from experiments, observations of physical processes, or other sources, as well as functions (parameters such as the mean or variance) calculated from them.

2.3.1 Identification of Important Parameters

Conceptual models enhance the quality of a performance assessment (e.g., improving the description of uncertainties about cumulative radionuclide releases and their effects on humans). Therefore, parameters should be identified to enhance the likelihood that their quantification leads to improved performance assessment. Initially, the identification and selection of important parameters requires substantial judgment by the experts who decide how a given parameter may affect the descriptions of uncertainty for repository performance.

Once parameters are identified, their relative importance can often be ascertained by sensitivity analyses (i.e., by varying the value of the parameter and determining the overall variation in the probability distribution of radionuclide emissions or some other intermediate performance measures) (Cranwell et al., 1987; Bonano et al., 1989). For example, Bonano et al. (1989), in their analysis of a hypothetical HLW repository in basalt formations, show that the hydraulic conductivities of some geologic layers were important, while those of other layers did not influence the total radionuclide discharge in 10,000 years. These results indicate that to reduce uncertainty about the containment requirement (40 CFR Part 191.13), research should focus on reducing the uncertainty in the value of the hydraulic conductivity for the important layers and not the others. Intuitively, one could have stated a priori that hydraulic conductivity in general is a relatively important parameter. However, for stratified repository sites, it is important to distinguish among the different strata and identify the most important, which can be achieved only with a preliminary performance assessment.

There are various approaches for sensitivity analysis, but unfortunately, there can be large inconsistencies in the results from different approaches (Iman and Helton, 1985). Iman and Helton also show that different interpretations of the results from a given sensitivity-analysis approach can lead to a different ranking of important variables. The problem is further complicated because not all sensitivity-analysis approaches are appropriate under all circumstances.

Thus, expert judgment clearly plays an important role in the identification of parameters, in the selection of sensitivity analyses, and in the assessment of the importance of parameters.

2.3.2 Quantification of Uncertainty in Parameters

To assess the uncertainty in performance predictions for HLW disposal systems, it is necessary to quantify the uncertainty in the input parameters of the models and codes used. The uncertainty in parameters can be expressed in a variety of ways. One way is to estimate a mean value and the variance about the mean. Another way is to determine the range of possible parametric values and to assess a probability density function (pdf) covering that range. The latter method is conventionally used in performance-assessment analyses for HLW repositories (Cranwell et al., 1987; Bonano et al., 1989) because it provides a complete description of uncertainty and facilitates the generation of multiple samples of the values of input parameters for carrying out Monte Carlo simulations. For these reasons, the examples below focus on the assessment of pdfs for input parameters.

In principle, estimation of the possible range of values and pdfs of input parameters should rely on a very large sample of field data. However, such a large sample is not likely to be collected at a candidate repository site. Expert judgments are required to determine what samples to take and how to interpret the results and to assess a probability distribution on the basis of the sample. Using Bayes' theorem, expert judgments can also be combined with data to arrive at a revised pdf for a parameter. Techniques for the elicitation and use of expert judgment can also be applied to quantify expert knowledge on a given parameter (e.g., hydraulic conductivity) to form a "prior" pdf for that parameter. If n observations are obtained during site characterization, a joint distribution of the n observations can be constructed. This joint distribution from collected data is used to modify the prior pdf to arrive at a "posterior" pdf.

Given that experts have to decide on what to sample and given that financial and other practical considerations are likely to prevent the collection of large amounts of data, it is imperative that expert judgments supplement sampling with documented and traceable procedures. The study described by Merkhofer and Runchal (1989) in Section 1.4 is an example of the use of expert judgments to quantify the uncertainty in the value of key parameters.

Another area in which expert judgment may play a considerable role is in the quantification of the spatial variability of hydrologic parameters. Although geostatistical techniques (such as kriging) exist for these purposes, they require input information, such as the mathematical form of the covariance function (describing spatial correlation), which is likely to be determined using expert judgment (see Bonano and Cranwell, 1988).

2.4 Information Gathering

Expert judgments are used with other sources of information to improve behavior predictions for the repository system. The current state of knowledge serves as a basis to decide what type of information should be collected and how it should be collected to predict the future behavior of the repository with less uncertainty. Additional information can be gathered in a variety of ways: collection of site-specific data, collection of related off-site data, laboratory experiments, and analysis with model systems. Expert judgment is important in selecting among the alternatives to obtain more information.

The activities to obtain new information are likely to depend heavily on expert judgments. If field data are to be collected at a proposed disposal site, experts must address issues such as the test to be conducted; the number, location, and depth of drilled boreholes; and interpretation of collected data; etc. In laboratory experiments, experts deal with issues such as how representative the experiments are of field conditions; under what conditions the experiments are likely to be invalid; how the laboratory data are to be used with field data; etc. Finally, if analyses use existing models to supplement experimental information, experts need to address issues such as how the adequacy of the models was established; what key assumptions are in the models that cannot be tested; and how to select the parameter values in the model(s) so that they represent the current state of knowledge about the disposal system, etc.

When contemplating any of these questions, one should consider the prior knowledge about the repository and its performance, the possible changes that could be produced by new information, the likelihood of these changes, and the cost of the information against its benefits. Clearly, any of these considerations requires a substantial amount of expert judgment, both about uncertainties (e.g., the prior uncertainty about a parameter) and about values (e.g., whether a million dollar experiment to decrease the uncertainty about a parameter is worth the cost).

2.5 Strategic Repository Decisions

The four areas of performance assessment discussed in Sections 2.1 to 2.4 pertain to the need for formal expert judgment within a repository designed, constructed, and operated according to a given set of specifications. Hence, the performance assessment largely depends on decisions about the design, construction, and operation of a repository, which will affect the postclosure behavior of the repository. For example, repository-induced events and processes must be considered in the development of scenarios (Section 2.1). All these decisions must rely heavily on expert judgment.

Many design decisions are critical. For example, the exact depth and size of the repository needs to be determined. The angle of the shaft to deliver the canisters to the repository needs to be decided. There are important decisions concerning the exact placement of the canisters. Should they be placed vertically or horizontally or at some other angle? And how near to each other should they be? These decisions could impact postclosure regulatory requirements such as canister lifetime and release rate from the engineered-barrier system, which, in turn, could affect radionuclide transport through the geosphere and release to the biosphere. Clearly, these decisions require both factual judgments (e.g., the lifetime of a canister), and value judgments (e.g., the worth of adding engineered barrier systems) from experts.

For each of the design decisions, there are complementary construction decisions. There may be different alternatives to sink and enlarge the shaft to reach the repository. Different alternatives may be useful for excavating the repository, both in terms of the techniques used and the timing of the activity. Different materials may be used to insulate the shafts, and different engineering solutions may be found for constructing the repository floors and walls. All these decisions affect the repository performance and involve crucial expert judgments that weigh performance against the costs and preclosure benefits.

Repository operation during the preclosure period also influences postclosure performance. For example, the management of the placement of canisters affects the degree of compliance with the design concepts of engineered barriers. Some decision problems may be necessitated by design or construction errors. Others will necessarily need to account for the possibility of such errors. In a similar vein, decisions about removing slightly damaged canisters or leaving them in the repository will affect long-term repository performance. Any of these decisions requires both factual and value-laden expert judgments.

The general point here is that one cannot examine expert judgment in (postclosure) performance assessment in isolation from the preclosure decisions and the numerous expert judgments involved in them. Simply put, postclosure expert judgments are only as good as the preclosure assumptions and judgments on which they are based.

3. ELICITATION, USE, AND COMMUNICATION OF EXPERT JUDGMENTS

In Chapter 2, five critical areas in need of expert judgment in performance assessment of HLW repositories were identified. This chapter describes the available formal approaches to elicit, use, analyze, and communicate expert judgment.

Section 3.1, defines the main terms used in formal expert-judgment processes. While the specific problems and the applicable techniques for eliciting expert judgments vary from situation to situation, the overall process is generic. It consists of identifying the elicitation issues, selecting the experts, training the experts and carrying out the elicitation sessions (Section 3.2). Within this process several techniques are useful, depending on the specific task at hand. These include identification techniques (e.g., generating scenarios or conceptual models), screening techniques (e.g., selecting scenarios), quantification techniques for probabilities (e.g., quantifing uncertainties about a parameter), and quantification techniques for values (e.g., evaluating alternative conceptual models). Many variants of these techniques are described in Section 3.3. Once individual expert judgments are elicited, they can be analyzed and used in a variety of ways. Section 3.4 describes the issues and procedures for combining expert judgments. There are several approaches to communicating expert judgments. These include the specific form of documenting expert judgments and of presenting the results of expert elicitations. These approaches are described in Section 3.5. Finally, Section 3.6 discusses the interpretation use, and misuse of expert judgments.

3.1 Definitions

This section defines some technical terms used in this report such as *issue*, *judgment*, *expert*, and *probability*, and *factual*, *value*, *quantitative*, *explicit*, and *formal judgments*.

A repository *issue* is a question about the present state of a repository, its future state, or events and processes that may lead it from one state to another. Issues may concern assumptions about the repository and the related natural and human systems. Issues may also concern the method of analysis for performance assessment. Issues are questions that should be addressed to carry out a performance assessment.

A *judgment* is an inference or an evaluation based on an assessment of data, assumptions, criteria, and models. There are two basic types of judgments: judgments about facts and judgments about values. Judgments about facts are usually called *beliefs* or *opinions*. People express their beliefs or opinions regarding propositions about facts or events whose truth or falsity can, at least in principle, be proven. For example, a person may believe that a nuclear waste repository will cost in excess of $20 billion in 1988 currency. Or a person can have the opinion that there will be no radionuclide discharges to the accessible environment from a nuclear waste repository within the first thousand years following closure. Although it would take 1000 years to determine the truth about whether such discharge occurred, this is in principle possible.

DRAFT

Judgments involving the use of criteria, priorities, and tradeoffs are usually called *value judgments*. There is no possibility of proving a value judgment true or false as can be done with factual judgments. For example, when comparing the value of the health benefits for workers with the health benefits for members of the public, some people might conclude that a worker fatality avoided is as important as a public fatality averted. Other people might conclude that a public fatality averted is more important because workers take the risks voluntarily. Such differences in value judgments are quite legitimate expressions of different social philosophies or priorities.

Many judgments mix factual and value elements. For example, beliefs about the costs of a nuclear waste repository, coupled with a value judgment about the socially desirable tradeoff between costs and benefits of the repository, could lead to the conclusion that the repository is "too expensive." Similarly, beliefs about the predictive ability of a model, coupled with a value judgment about the relative importance of predictive ability vs. simplicity, could lead to the conclusion that the model is "adequate."

An *expert* has or is alleged to have superior knowledge about data, models, and rules in a specific area or field. Expertise is characterized by easy access to relevant information and by the ability to process that information and to use it effectively. Shanteau (1987) observed other characteristics that define experts: the ability to simplify complex problems and to identify and react to exceptions; a strong sense of responsibility; confidence in their own judgment; and adaptability related to their knowledge domain. The domain of an expert can be a factual domain (e.g., a scientific data base) or a value domain (e.g., the area of policy tradeoffs). Factual and value domains are often mixed, however, and one of the characteristics of expertise is the ability to separate factual and value components of judgments. For example, experts decide what data are relevant, what models should be used, how to interpret data to make recommendations, etc. Any of these decisions involve both value and factual judgments.

Expert judgments can be *implicit* or *explicit*. An explicit expert judgment is stated and documented for others to appraise. For example, when a particular conceptual model for a repository is chosen, the reasoning behind that choice can be made explicit in writing. Or when a numerical estimate of a parameter value is chosen, supporting evidence can justify that choice. In contrast, implicit expert judgments are not available for appraisal and need to be inferred from actions and statements that are available for appraisal. For example, when screening scenarios, certain screening criteria may have been applied, but these criteria or their rationale may not have been explicit.

An explicit expert judgment can be *quantitative* or *qualitative*. A quantitative judgment expresses opinions or evaluations in numerical terms. Examples are the estimation of a parameter or the judgment of a probability of an event. Another example is the statement that public fatalities are four times as important as worker fatalities when evaluating health impacts from the repository. Explicit *qualitative* judgments are often expressed as verbal statements like "acceptable," "high chance," or "virtually impossible." The decision that "reasonable assurance" has been provided that all regulatory requirements will be met is an explicit qualitative judgment. Many qualitative judgments enter scenario screening and conceptual-model selection and may be used to make the judgments explicit.

Quantitative expert judgments about facts can be expressed as *probabilities*. Probability is a degree of belief in an unverified proposition (DeFinetti, 1937; Ramsey, 1931; Savage, 1954). Probabilities record the state of knowledge that an expert has about a specific proposition. These propositions can be about uncertain events (e.g., "there will be an earthquake of magnitude 7 or higher on the San Andreas fault within the next 30 years") or about uncertain quantities (e.g., "the average travel time of radionuclides in medium A"). Uncertain quantities are also called *random variables*. Probabilities are numbers between 0 and 1 (inclusively), and they obey the laws of probability theory. Nonprobabilistic quantitative judgments include ranges of parameters or point estimates such as the "best guess" of a parameter value.

Quantitative judgments about values can be expressed as *utilities*. Utilities express the tradeoffs among attributes of the alternatives to which the value judgments are relevant (Keeney and Raiffa, 1976). For example, in selecting experiments for testing a given performance assessment model, a tradeoff is made between the information to be gained and the cost of the alternative experiments. Possible tradeoffs may be between the costs and benefits of laboratory experiments vs. field tests.

Decision analysis is a systematic procedure to assist experts and decision-makers in making judgments and choices in the presence of uncertainties, risks, and multiple conflicting objectives. Decision analysis comprises a philosophy for problem solving, formal axioms and models for inference, evaluation, and decision making, and a set of techniques for their implementation. Decision analysis includes techniques for decomposing issues and problems, quantifying expert opinions and value judgments, analyzing and using these judgments, and recombining the decomposed problem.

3.2 <u>The Process of Eliciting Expert Judgments</u>

3.2.1 Identification of Issues and Information Needs

In the previous section, issues were defined as questions about the present state of a repository, its future state, and events and processes that may lead it from one state to another. Resolution of issues improves the quality of decisions about the repository and, as a special part of such decisions, the quality of performance assessments.

Issues range from general to fairly specific and from extremely complex to simple. For example, a general, complex question may be, "Which conceptual model provides an adequate description of the past, present, and future states of the repository?" A fairly specific and somewhat simpler question may be, "Within a given conceptual model, what is the appropriate numerical value of a parameter describing hydraulic conductivity?" Issue identification may involve identification of the geologic and hydrologic features of the repository, identification of all major failure modes and pathways to the accessible environment, and identification of possible conceptual models and scenarios for analyzing failures.

Early in issue identification, emphasis should be on broadening the range of issues rather than narrowing it. It is often useful to invite persons outside the analysis staff to participate in this early stage. For example, public interest groups may be asked to express their concerns, objectives, and potential scenarios regarding failure modes in the repository. External review can aid in achieving completeness of the analysis and curtail criticism for failing to examine some issues. Examining and discarding an issue will be more acceptable than justifying, after the fact, why the issue was not considered at all.

Once a complete list of candidate issues has been created, it should be screened to identify those most relevant to repository performance. Relevance includes both judgments of the likelihood that an issue influences the overall probability of a failure at a repository as well as the extent of the possible consequences of failures. Screening should employ both criteria.

After reducing the set of issues, information needs should be identified. In making decisions about the acquisition of information, consideration should be given to the relative accuracy, cost, and availability of alternative sources of information. The result, again, is not a final list, since the issue under consideration will be further analyzed and reviewed as issue descriptions are formulated and decomposed into subissues.

Clearly laying out the issues for the experts is crucial. If five experts are asked to write down their understanding of an issue, one is apt to get five somewhat different descriptions. Critical differences can arise in the assumptions that experts make. The understanding of the initial conditions may vary greatly. If these assumptions or initial conditions are not explicitly defined, there can be an ensuing confusion during subsequent elicitations regarding the issue.

3.2.2 Selection of Experts

Performance assessment for HLW repositories requires several types of experts: generalists, specialists, and normative experts. The generalists should be knowledgeable about various overall aspects of the repository performance assessment. They typically have substantive knowledge in one discipline (e.g., geology, hydrology, transport phenomena) and a general understanding of the technical aspects of the problem. However, they are not necessarily at the forefront of any specialty within their main discipline. The specialists, on the other hand, are at the forefront of one specialty relevant to the performance of the repository, but they often do not have the generalist's knowledge about how their expertise contributes to the overall performance assessment. Normative experts typically have training in probability theory, psychology, and decision analysis. They assist generalists and specialists with substantive knowledge in articulating their professional judgments and thought processes so that they can be meaningfully used in the performance assessment. A high-quality performance assessment requires the teamwork of all three types of experts.

Each expert to be used in a performance assessment should be carefully selected to achieve a high-quality performance assessment. Operationally, this means that the performance-assessment team should address all the complex technical aspects of the problem and do this in a logically sound, practical manner that is open to evaluation and peer review. The assessment should be politically acceptable, compatible with existing scientific and governmental institutions, and conducive to learning (Fischhoff et al., 1981).

3.2.2.1 Selection of Generalists

Generalists oversee completion of the performance assessment and provide quality control for the performance-assessment models and resulting analyses. Hence, generalists are usually selected from among the professionals within the organization responsible for the performance assessment. In selecting these generalists, project management should consider technical skills, organizational skills, and personal interaction skills. The generalists must have an understanding of the technical aspects of the overall performance assessment at a level where they can substantively communicate with specialists and normative experts. They should have organizational skills to schedule appropriately the gathering of information for the performance assessment. Generalists also need personal interaction skills to interact effectively with the numerous project personnel, specialists, and normative experts involved in the performance assessment.

3.2.2.2 Selection of Specialists

There are three alternatives to consider in selecting specialists: (1) a single specialist to provide the set of judgments required, (2) a panel of more than one specialist in which each provides the set of judgments required, and (3) an expert team of specialists with the synergistic knowledge to provide a single set of judgments in situations requiring broader substantive knowledge than is typically possessed by an individual. The following addresses the identification and selection of individual specialists, panels of specialists, and expert teams.

The process of selecting specialists must be considered reasonable. Whether selecting individuals, panels, or teams, the first step is to identify specialists whose judgments might be appropriate for the performance assessment. The performance-assessment staff may have a number of suggestions for possible specialists. Others may come to mind from reviews of the published scientific literature addressing specific topics of interest. Parties interested in HLW disposal, such as utility companies and environmental groups, may have suggestions for appropriate specialists. Indeed, an open solicitation of nominations for specialists, including self-nominations, is one way to instill public confidence in the process. On important problems like HLW disposal, a formal solicitation of experts in the form of a request for expertise (much like a request for proposal) could be very useful to identify the full range of expertise available and to ensure that an adequate search for expertise has occurred. Once a list of candidate specialists for use on a specific aspect of performance assessment is identified, a selection process must occur.

In the selecting a specialist, there are a number of important considerations. Foremost, it is critical to ensure that the specialist has the expertise necessary. This should be verified by reviewing the individual's vita, by discussion with peers in the field of specialty, and, most importantly, by discussions directly with that expert. It is also important that selected specialists be perceived as having that expertise by peers and others in related fields. If these criteria are met, then the potential specialists need to be both willing and available to participate. Another key consideration is whether they are willing to have their name attached to their expert judgments in the project documentation (Section 3.5.1). Naming experts may enhance the quality of the expressed judgments, but more significantly, it increases the ability to evaluate the process and raises its credibility. The criteria used for selection should be explicit and well documented.

It is very important to avoid any potential conflict of interest between the specialists and the results of the performance assessment. A common issue is whether the prospective specialists derive their employment or any income from organizations charged with conducting the overall performance assessment or with constructing the repository. Those available specialists with no conflicts should be chosen based on their expertise.

Individuals with a perceived or real conflict of interest may not allow this conflict to influence their professional judgments. Furthermore, we would not like to exclude crucial information from the performance assessment simply because a knowledgeable individual had a potential conflict of interest. Therefore, it is important to design the explicit elicitation and use of expert judgment such that the knowledge and reasoning of experts with potential conflicts can be made known to selected specialists in a timely manner. This communication process may include distribution of written publications and analyses, as well as oral presentations.

When a panel of specialists is to be selected, each specialist should, of course, have a high professional stature. However, additional issues are important. One of these is how many specialists are appropriate. Evidence suggests that three to five experts are usually sufficient to tap most of the expertise (Clemen and Winkler, 1985). It is desirable to have the full range of legitimate opinions on a particular scientific topic available on any panel of specialists and this implies that the specialists on a panel should be as independent as possible. Diversity is achieved when the specialists' sources of information and their reasoning processes are different, and their approaches (e.g., theoretical models vs. experimentation) and professional training are different. Of course, to some degree, all experts would likely be at least somewhat familiar with the work of other experts in their fields. In addition, they would base their judgments on common scientific and engineering principles and knowledge. Thus, specialists cannot be completely independent, but this goal is important because it provides a more complete picture of the state of scientific knowledge as well as lending credibility to the performance assessment by representing a broader viewpoint.

A quality performance assessment requires the expert judgments based on knowledge and experience in many disciplines. These expert judgments will need to be logically integrated, along with all other relevant information and data, into models. No expert teams are necessary if the results of expert judgments from individuals or panels are naturally packaged to integrate into the analysis. However, at other times the natural package of information based on experts' judgments can only be acquired from an expert team comprised of specialists in related but synergistic disciplines. An example is a study involving seismicity on the east coast of the United States. Each expert team was comprised of at least one seismologist, one geologist, and one geophysicist (see Electric Power Research Institute, 1986).

Each specialist on an expert team should meet all of the qualifications of individual experts stated above. The disciplines whose knowledge is essential to the scientific problem under investigation must be represented as part of each expert team. The performance-assessment staff and then the expert team itself must ensure that all relevant disciplines are included. The performance-assessment staff originally selects the specialists for the expert team based on project needs and the required scientific judgments. The expert team and performance-assessment staff should initially review the task and outline procedures to combine logically the judgments of various team members to provide the required overall judgments. If specific expertise is identified as lacking from the team at this stage, the team should be augmented with additional specialists possessing the required knowledge.

3.2.2.3 Selection of Normative Experts

The criteria for selecting normative experts are essentially the same as those that guide selection of individual specialists. Both the process of selection and its results are important because both influence the quality and the perceived quality of the ensuing elicitations of expert judgments. Normative experts require a sound theoretical and conceptual knowledge of probability and techniques for eliciting judgments, and they need to be knowledgeable about the psychological processes occurring in the specialists' minds as they are processing information to produce requested results. Normative experts should also have significant skill and experience in working with technical professionals to make them feel comfortable in expressing their judgments and in explaining their reasoning. Finally, normative experts should possess the communication skills necessary to interact substantively with project generalists and specialists and to document thoroughly the results of expert elicitations.

As with specialists, the qualifications of normative experts can be verified by appraising the individual's vita, discussion with peers experienced in elicitation and with specialists whose knowledge has been elicited by the individual in question, and by discussion with the individual. Unlike the case with specialists, prospective normative experts can be asked to demonstrate their skills in actual elicitations using individuals on the performance-assessment staff as specialists.

3.2.3 Training

The professional literature on expert judgment clearly stresses the importance of training experts in various aspects of the task facing them (Spetzler and Stael von Holstein, 1975; Merkhofer, 1987; von Winterfeldt and Edwards, 1986; Mosleh, Bier, and Apostolakis 1988). Training consists of the following tasks:

- familiarizing experts with the expert-judgment process and motivating them to provide formal judgments,
- giving experts practice in expressing their judgments formally,
- educating the experts about the possible biases in expert judgment and applying debiasing techniques.

To accomplish these tasks, it is desirable to convene the experts individually or as a group before the actual elicitation for at least a day. The training session should be led by a normative expert with an in-depth knowledge and experience in the art and science of formal expert-judgment processes.

The remainder of this section provides some general guidelines and ideas about how to accomplish these three tasks.

Familiarizing the experts with the judgment process and motivating them to provide formal judgments. In most expert elicitations, the experts are specialists with substantial knowledge in a fairly restricted domain who have developed their own styles of communication and expectations about types of questions they can or cannot answer. They are usually very cautious regarding conclusions and judgments that may appear to be beyond the direct implications of data and experimental findings, scientific reasoning, or models.

Providing formal expert judgments is usually unfamiliar to experts, and sometimes it may even be threatening. They may feel that they will be asked unreasonable questions. In particular, they may worry that they will be asked to provide more precise answers than their current knowledge justifies. In addition, they may not understand why they should express their judgment at all, or if so, why in terms of numerical judgments such as probabilities or utilities. Furthermore, they may consider the expression of judgment based on incomplete knowledge to be inferior to the scientific work that would improve their knowledge base. Finally, they may worry that their judgments may be misused or misrepresented.

It is therefore important that the training session address these concerns explicitly. First, the normative expert, with technical input from generalists, should provide an overview of the performance assessment and indicate where the specific expert judgments will be used. The normative expert should point out that the experts were chosen to accomplish an important task and explain why they are among the more suitable for this task. Second, the need for formal expert judgment should be stressed. In performance assessment for HLW repositories, this need clearly arises because there are large uncertainties about scenarios, models, and parameters, and data are scarce. In addition, many decisions involve tradeoffs, as in between development cost and predictive accuracy in a conceptual model. Third, the normative expert should stress that there are no right or wrong answers to questions about expert judgments and that the purpose of the elicitations is to assess both what the experts know and what they do not know. Fourth, the normative expert should clearly explain that the process of eliciting expert judgments is not a substitute for further work in the expert's fields, but is, rather, a tool to summarize their current information. Formal elicitation of expert judgment often identifies very clearly where sufficient knowledge exists, and where more research is needed. Finally, the way in which judgments will be used should be explained carefully. If, for example, judgments are averaged across experts, this should be explicitly stated and discussed.

The normative expert should present a number of examples to illustrate various forms of expert judgments. These include implicit and explicit judgments, qualitative and quantitative judgments, and probability and utility judgments. The examples should preferably be drawn from the substantive knowledge domain of the specialists, such as geology or hydrology.

Most experts know that they use judgment in their work all the time, but the specific forms of judgments in expert elicitations, especially probability and utility judgments, are likely to be unfamiliar to them. It is therefore useful to explain the basic concepts as well as the main properties of probabilities and utilities. Experts should be shown many examples of probability distributions and utility functions from within and outside of their field.

An important issue in any expert elicitation is the definition of the variable or event for which the judgment is to be expressed. The normative expert should present many examples of well-defined and ill-defined events and variables and illustrate them with the pitfalls of poor definitions: misunderstandings, miscommunication, and inappropriate assumptions.

Even after a thorough training session, some apprehension and concern may remain. Most of these remaining concerns can be addressed only in performance of the tasks and it is therefore more useful to give the experts some practice in elicitation of expert judgments rather than discussing the issues abstractly.

Giving experts practice in expressing their judgments explicitly. There are several aspects of expert judgments that require practice:

- making implicit judgments explicit,
- decomposing problems, and
- providing numerical judgments, especially probabilities and utilities.

To show how implicit judgments can be made explicit, the normative expert should present the experts with several simple tasks involving judgments and afterwards point out that the answers require judgment and many answers include implicit assumptions. For example, when asked whether a canister in a repository will leak within the first thousand years, an expert may say that this is extremely unlikely. Implicit in this judgment are assumptions about the repository condition and canister corrosion. The normative expert should elicit these assumptions and point out their role in the judgments made.

Most expert judgments can be aided by decomposing the problem. For example, when estimating groundwater travel time through a layered medium, an expert may decompose his judgments by defining several layers and estimating groundwater travel time separately for each layer. Judgment of the relative contribution of each layer can then be combined with the conditional estimates of groundwater travel time to arrive at an expected groundwater travel time.

There are several modes of decomposition. For factual judgments, event trees, fault trees, and functional decompositions are helpful (McCormick, 1981; Raiffa, 1968), and for value judgments, value trees and objectives hierarchies are used (Keeney and Raiffa, 1976). Since any of these may be useful for representing and decomposing expert knowledge in a specific problem, it is useful to provide experts with some training in each mode.

The third area of practice is the actual elicitation of numerical values, especially probabilities and utilities. This can be done by carrying out some example elicitations interactively with the group. The literature on cognitive illusions and probability biases (Hogarth, 1980; Kahnemann, Slovic, and Tversky, 1982; von Winterfeldt and Edwards, 1986) has many useful examples.

All tasks that are likely to occur in the elicitation sessions should be practiced. At a minimum, the experts should learn to respond to questions both outside their field and within their field, to factual and value problems, to questions about discrete events and continuous uncertain variables, and to difficult and easy questions. It is best to begin with easy questions on discrete events outside the experts' field and to end with difficult questions on continuous uncertain variables in their field. This sequence allows the experts to develop a degree of comfort with answering questions before the challenging and presumably more uncomfortable questions are posed.

Educating experts about biases and applying debiasing techniques. Cognitive psychologists have identified many biases in expert judgments (Hogarth, 1980; Kahnemann, Slovic, and Tversky, 1982) Two general classes are *motivational* biases and *cognitive* biases. Motivational biases can occur because the expert has a stake in the issue considered that may lead to conscious or unconscious distortions of his judgments. For example, a bridge engineer is motivated to claim that a bridge that he just helped to build is absolutely safe (i.e., the probability of it collapsing is zero). Cognitive biases occur when experts fail to process, aggregate, or integrate appropriately the available data and information. Most experimental research is on cognitive, rather than motivational biases, yet it is important in the training sessions to discuss and elaborate on both.

Research on cognitive biases has concentrated on probability cognitive biases, and this section focuses on them. However, cognitive biases occur in utility judgments as well. Some recent experiments (Weber et al., 1988) indicate, for example, that objectives presented in more detail tend to be weighted more heavily. Furthermore, cognitive biases can occur when structuring and framing the task at hand. Two common structural biases are incomplete specification of alternatives and incomplete statement of the assumptions underlying judgments. Fischhoff et al. (1978), for example, showed that car mechanics and other subjects often fail to recognize all possible failure modes of a car defect (e.g., failure to start). Experts often make estimates based on "normal" conditions or assumptions, but fail to make these conditions or assumptions explicit.

Most cognitive biases related to probability judgments include

Overconfidence	Giving probability judgments that express less uncertainty than the experts' knowledge would justify (i.e., too tight or too steep probability distributions);
Anchoring	Adjusting judgments insufficiently after anchoring on an initial estimate (e.g., a mean or median);
Availability	Overestimating probabilities of events that are easily imaginable or recalled;

DRAFT

Ignoring base rates	Focusing on concrete evidence and data as a main source of probability judgments and ignoring more abstract information like base rates and prior probabilities;
Nonregressive prediction	Ignoring the unreliability of the relationship between variables and therefore making predictions as if the relationship were reliable.

Training should focus on the more likely biases in the particular performance assessment. In scenario construction and selection, for example, likely biases are incomplete events and assumptions, availability, and overconfidence. In the identification, appraisal, and selection of conceptual models, anchoring and availability, are most likely. In the assessment of uncertainty for parameters of models, overconfidence, anchoring, and nonregressive prediction are likely.

Debiasing techniques have only recently been developed (Kahnemann and Tversky, 1979; Fischhoff, 1982). For motivational biases, awareness of motivational factors both by the expert and by the elicitor is important. Sometimes it helps to present the question in the form of a hypothetical gamble (Section 3.3.4) to counteract motivational biases. For example, an expert may state that it is absolutely impossible that a nuclear reactor containment fails at pressures below 120 psig. In that case, one might ask him, if he is willing to accept a bet awarding him $10 in the event that no U.S. reactor containment will fail below 120 psig in the next 10 years vs. the loss of all of his possessions if one such accident occurs. Experts should be trained in such questions and be made aware in the training that the elicitator might attempt to debias them this way when they suspect motivational biases.

For cognitive biases, familiarity with the task, awareness of the bias, feedback, and personal experience with the bias help to reduce it. A useful training exercise is to provide experts with a catalogue of probability questions that are similar to those used in the bias experiments and to let them experience the bias themselves. While this does not assure self-correction, it at least alerts them to the problem in a more vivid way. Since overconfidence, anchoring, availability, and nonregressiveness seem to be the main problems that might influence a performance assessment, a questionnaire that induces these four biases would make excellent training material.

3.2.4 Conducting Elicitation Sessions

The elicitation of expert judgments should be based on a well-defined set of issues (Section 3.2.1). However, since the issues are identified before the selection of the experts, the experts may have suggestions for redefining details of the issue they are supposed to address. Before beginning the elicitation, it is therefore important to discuss the issues, the possible problem decompositions, the events and variables, and the questions that will be asked. In the elicitation of probability judgments, it is especially important that the events and variables are well defined. In the elicitation of utilities, it is important that the objectives and scales for measuring them are well defined. For qualitatively described events this means, among other things, that the events are mutually exclusive and collectively exhaustive and that all conditioning events are defined. For quantitative variables, this means, among other things, that the meaning, dimension, and unit of the variable are well defined. If events or

DRAFT

variables are ill defined, various implicit judgments may enter the elicitation to fill the "definition gap." Different experts may make different assumptions, and the elicitators and analysts may apply other assumptions in analyzing the responses, leading to confusion, miscommunication, and poor performance analyses.

If expert judgments provide specific inputs into a performance assessment, it is important that they match the requirements of the overall analysis. Thus, there also should be preelicitation discussion of the nature and amount of expert judgment required by the overall performance assessment.

Alternative problem decompositions should be discussed, but some discretion should be left to the experts in matching the individual decomposition to their thought processes. In addition, there often are alternative means of expressing the elicitation events or variables through probabilistically related events or through functionally related variables. Again, each expert should feel free to choose among the alternatives that best accommodate his or her thinking, as long as the resulting responses can be related functionally or probabilistically to the elicitation events or variables.

It helps for the staff involved in the elicitation and one or two generalists or specialists to think through the whole elicitation process and practice it. Guidelines for the elicitation should be drawn up, and materials (forms, graphs, etc.) should be designed for the actual elicitation.

An elicitation is an interaction between at least two people: the specialist and the normative expert. The specialist provides judgments, for example, in the form of probabilities or utilities, as well as all relevant technical reasoning concerning judgments and conclusions. In addition to verbal statements, the specialist should provide written materials documenting the reasoning as well as any background material used in preparing for the elicitation.

The normative expert is knowledgeable in the art and practice of expert elicitation, with special knowledge in probability and utility elicitation. The normative expert asks the specialist to provide specific answers to questions regarding the events or variables considered, assists the specialists in explicating their reasoning, ensures that the required information is obtained, checks the consistency of the specialist's judgments especially with the laws of probability and documents the numerical results for later processing.

In some elicitations, it is useful to request the participation of a generalist for expertise in the requirements of the overall project and expertise in the specialist's area. The generalist ensures the technical validity and consistency of the specialist's judgment, clarifies technical issues, documents the specialist's technical reasoning, and provides technical data and assumptions when needed.

3.2.4.1 Basic Elicitation Arrangements

The elicitation should take place in an undisturbed environment, preferably a separate room without telephone interruptions, visitors, or disturbing noise. The desk arrangement should be comfortable, encourage interaction among the individuals involved in the elicitation, and have work space and sufficient space for documentation materials, forms, and recording devices.

There are several ways of documenting an ongoing elicitation: tape recording, written notes by the normative expert, written notes by the generalist, and notes or documents that the specialist brings into the session. Tape recordings provide a complete voice record. During taped elicitation sessions, it is important to refer explicitly to the materials and documents, figures, and tables used in the discussion to facilitate transcription and cross-referencing in the written documentation. While tape recordings may provide more detail than necessary, they can be important for accountability, and for verification and clarification during written documentation.

Notes taken during the elicitation session by the normative expert and the generalist have different focuses. The normative expert focuses on writing down judgments and making lists, tables, and figures summarizing and relating these judgments for communication and feedback. In case of probability elicitation, for example, the elicitator should write down the probabilities as tables, distributions, or functions that allow quick consistency checks and calculations for feedback. While most documentation of the normative expert is numerical, it is useful to note on the tables and plots the the specialist's rationale for certain judgments. The generalist should record the specialist's reasoning in support of the judgments as well as cross-referencing it to the specialist's own documentation. It is important that the documentation schemes of the normative expert and the generalist are similar so that they can be cross-referenced when documentation is consolidated.

3.2.4.2 Structure of a Standard Elicitation Session

A standard elicitation session begins with easing the specialist into the situation and mapping out the task. The normative expert should ask the specialist to provide a brief overview of his or her approach to the problem and, in particular, the problem structure and decomposition used. After this exchange, the normative expert should define a road map for the remainder of the elicitation to determine the amount of work ahead.

Next the definition of the events or variables to be elicited should be reconfirmed. The normative expert should define the events and variables carefully, check the various meanings with the specialist and the generalist and write down the dimensions and units on the forms prepared for the elicitation. Assumptions, especially about conditioning events, should be discussed and documented.

In the case of a decomposed event or variable, the normative expert should first map out a rough decomposition to clearly describe the logic used and simplify the judgmental tasks. Next the normative expert uses any combination of specific techniques (Section 3.3) to elicit expert judgment. These techniques range from largely qualitative for identifying scenarios, models, or events to mixed qualitative-quantitative for screening, to largely quantitative for probability and utility judgments.

Consistency checks by the normative expert are important to assure the internal logic of the expert judgments and to assist in identifying sources of inconsistencies and resolving them. Consistency checks should be used to stimulate the specialist's thought processes. In probability elicitation, for example, it is useful to ask the same question by eliciting the desired probabilities directly or by eliciting probabilities for related variables or events. At a minimum, decomposed judgments should be reaggregated to arrive at a calculated judgment about the elicitated event or variable, and this calculated judgment should be compared with the specialist's intuition.

3.2.4.3 Post-Elicitation Activities

The specialists should be given quick feedback on the results of the elicitation. In particular, they should be shown the numerical information in the form of tables and distributions. Changes required by the specialist upon such feedback should be adopted and reasons for them should be carefully documented.

In some cases, it is desirable to organize a group meeting of specialists, generalists, and normative experts after the individual sessions to discuss agreements and disagreements and whether it is possible or desirable to reach consensus. There are several ways to organize such an interaction (See Section 3.4.4 and Seaver, 1978). In some instances, it may even be desirable to reelicit some individuals after this group session.

Sometimes specialists may want to change their elicitations after a significant time has passed. Such change requests should be probed carefully but accommodated if feasible within the framework of the overall project. Reelicitation may be necessary, and the documentation should reflect the revisions and the reasons for them.

The basic design also requires eliciting one specialist at a time. It is conceivable to elicit several specialists simultaneously, for example, in groups or classroom sessions. While this method is preferable to a pure questionnaire format, it suffers from some of the same drawbacks. In particular, classroom settings require more conformity on case structure and decompositions, allow less flexibility in individual responses, and may suppress expressions of alternative views.

There are, of course, many variants to the postelicitation activities. An important issue is whether the elicitation to achieve group consensus, to aggregate different judgments, or simply to report the results from different specialists (Section 3.4.).

3.3 Techniques for Expert Judgment Elicitation

An expert engages in three fundamental cognitive processes when making judgments: (1) *identification* of options or events to be judged; (2) *screening* of the options and events; and (3) *quantification* of comparative judgments about the options and events. *Identification* consists of recall, search, and creation. Recall identifies easily available alternatives, search systematically lists existing alternatives, and creation generates previously unknown or inaccessible alternatives. *Screening* consists of selecting screening attributes, setting screening constraints, and selecting alternatives based on the attributes and constraints. *Quantification* consists of assigning numbers to factual or value judgments about alternatives. Factual judgments about events or random variables are usually quantified by probability distributions. Value judgments (e.g., about the advantages or disadvantages of alternative conceptual models) are usually quantified by utility and tradeoff judgments.

The literature on identification techniques is fairly small. There are a few techniques for creative option and event generation (Pearl, 1978; Pitz, Sachs, and Heerbroth, 1980; Gettys, Fisher, and Mehle, 1978; Keeney, 1988a). Most screening techniques consist of setting numerical cutoffs on selected screening attributes and searching for the subset of "survivors." Keeney (1980) describes the basic idea for screening in a value judgment context, and several reports discuss the use of "cutoff probabilities" for screening undesirable events (Department of Energy, 1986; Okrent, 1980; Wilson, 1984).

In contrast to the small literature on identification and screening techniques, there is a rich literature on quantification techniques that draws mainly on psychophysics (Poulton, 1979; Ekman and Sjoberg, 1965; Zinnes, 1969) and decision analysis (Raiffa, 1968; Brown, Kahr, and Peterson, 1974; Keeney and Raiffa, 1976; von Winterfeldt and Edwards, 1986). The decision analysis literature typically emphasizes quantification of probabilities (Spetzler and von Holstein, 1975; Selvidge, 1975; Seaver, 1978; Keeney, 1980; Stillwell, Seaner, Schwartz, 1981; Wallsten and Budescu, 1983; Merkhofer, 1987) and utilities (Keeney and Raiffa, 1976; Keeney, 1980; Edwards and Newman, 1982).

The following three sections summarize this literature and make recommendations about techniques for identification, screening, and quantification.

3.3.1 Identification Techniques

Identification techniques primarily assist experts in identifying scenarios and conceptual models for performance assessment. In scenario identification, the emphasis is on stretching the experts' imagination and on creative processes of event generation. Conceptual model identification, emphasizes generating desirable model alternatives.

3.3.1.1 Techniques for Event and Scenario Identification

Recall and search are fairly trivial tasks in event and scenario identification. In the recall mode, one simply asks the experts to list all the events and scenarios that they recall that are relevant for the normal performance of the repository or for scenarios that could adversely impact that performance. In the search mode, experts survey the literature for relevant events or scenarios. It helps to enrich the set of events and scenarios by asking nonexperts and those with a stake in the decision (e.g., environmental groups, residents living near the repository). The emphasis at this stage should be on completeness and comprehensiveness, not on logic, reasonableness, or likelihood of occurrence.

Event and scenario creation is the most interesting and innovative aspect of this task. There are three cognitive techniques to creative scenario generation:

- forward and backward induction;
- value-driven event and scenario generation; and
- analogy- or antinomy-driven event and scenario generation.

Forward and backward induction builds on the notion that scenarios are logical sequences of events linked through processes. It begins with listing all possible and conceivable events that could occur related to a repository. In the forward induction mode, events are linked to create an event tree that fans out from initiating events to events that may occur in thousands of years. Provided that the events and processes are defined sequentially, this event tree can, in principle, be constructed mechanically, typically leading to a very large tree representing with thousands of scenarios. This tree should be pruned to eliminate branches that are impossible, extremely unlikely, or redundant. In the backward induction mode, the final states of the repository are the starting point of the process. A possible final state may be defined as "major releases to the accessible environment occur in the year 3000." Backward induction defines the possible causes of this final event and thus works back to the initial conditions, events, and processes that make it possible.

Forward induction typically creates too many scenarios, while backward induction may create too few. By applying both processes and reconciling the results, it should be possible to identify a subset of scenarios that spans the range of scenarios relevant to the performance of the repository.

The second technique begins with the question: What are the performance objectives for a repository and how can they be achieved? Presumably the main objective is to protect public health and safety, but other objectives like cost and long-term environmental protection may be important as well. After identifying a set of objectives, events and scenarios are developed that would lead to extremely poor, average, and extremely good performance on each objective (Keeney, 1988a; Edwards et al., 1987). For example, in the case of health and safety, an "undisturbed-performance" or "base-case" scenario without major geological events or human intrusions would presumably lead to average performance. Adding favorable assumptions about the behavior of the canister materials and the rock medium may lead to extremely good performance. Combining major magnetic and seismological events with poor geology and excessive corrosion may lead to very poor performance. While this technique tends to look at the worst case in terms of health and safety, it is very instructive to look at other cases and other objectives as well.

Event and scenario creation by analogy or antinomy attempts to stimulate the thought processes of the experts (Jungermann and Thuering, 1987). In an analogy, one would take the events and scenarios out of the context of an HLW repository and ask experts to instead think of the repository, for example, as a coal mine containing lethal gases. The question would be: What could go wrong in this coal mine? The follow-up question would be: Do any of these coal mine events and scenarios apply to the real repository case? In an antinomy one could ask experts to think of the repository, for example, as containing the most precious human possession that required protection from attempted theft. The question might be: How can thieves enter the repository, and how can theft be prevented? Again, the answers would be checked for their relevance to repository performance.

Any of these three techniques can be combined with various forms of interactions among experts. These include Delphi-type techniques (Linstone and Turoff, 1975; Dalkey, 1969), the Nominal Group Technique (Delbecq et al., 1975), and several forms of brainstorming. Furthermore, they can be substantially enhanced by involving individuals with very different perspectives regarding the repository (e.g., local residents, environmentalists, and nuclear engineers). Since the purpose at this point is to assure comprehensiveness, any inputs that are novel and creative should be appreciated. Peer review is another useful mechanism to identify events and scenarios that have been overlooked.

It is very important that the activities during event and scenario identification and the results are carefully documented. In particular, reasons for eliminating certain events and scenarios should be carefully recorded.

3.3.1.2 Identification of Conceptual Models

As in scenario identification, recall and search are fairly straightforward activities to identify conceptual models. The main technique for the innovative creation of conceptual models is similar to the value-driven technique described above (Pitz and Sachs, 1984; Pearl, 1978; Keeney, 1988a). The technique begins with a listing of the desired properties or objectives for a conceptual model. Next the experts develop features of conceptual models that would serve one objective well. After completing this task with the first objective, it is repeated for the second, the third, and so on. Features developed from subsets of objectives are combined to characterize one possible conceptual model. Repeating this process suggests many different conceptual models.

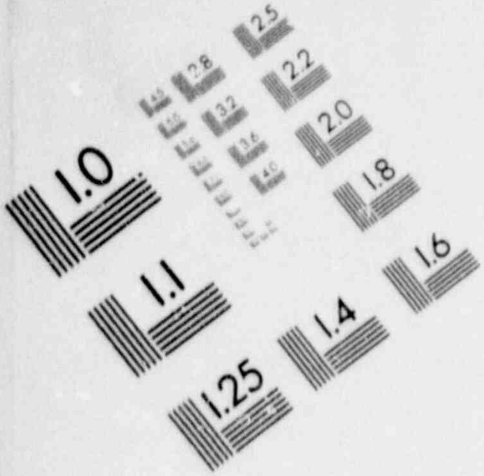
Having generated a large number of conceptual models, the next task is to narrow this set down to a reasonable size. This task includes examining all conceptual models on all objectives simultaneously and eliminating those that are clearly unacceptable on one or more objectives. Since this task involves screening, many of the techniques discussed in the next section will be applicable.

3.3.2 Screening Techniques

The first step in screening scenarios or conceptual models is to identify the attributes with which to screen alternatives. This step is followed by setting target levels or constraints on the attributes. Alternatives are then screened out that do not meet the target levels and constraints. Typically, this process is iterative: when too many alternatives survive, more stringent target levels or constraints should be applied. When too few survive, target levels or constraints should be relaxed.

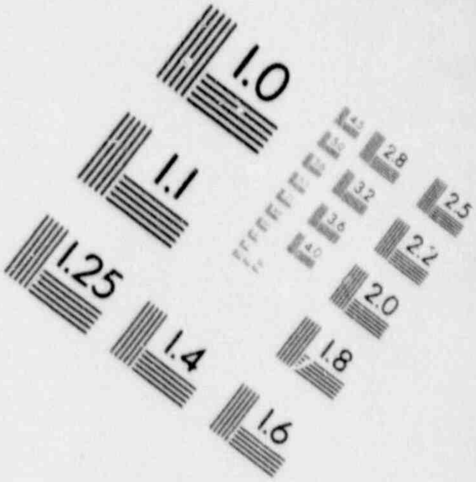
Identification of Attributes. Scenarios should be physically consistent sequences of events. It is therefore important to screen out those that are logically flawed. For example, if one event is the coming of another ice age combined with the migration of the earth's population to the southern hemisphere, it is logically inconsistent to couple this event with large numbers of human exposures because of radioactive leakage. Given another ice age, it is improbable, although not logically inconsistent, that there would be exploratory drilling for minerals other than the radioactive materials themselves.

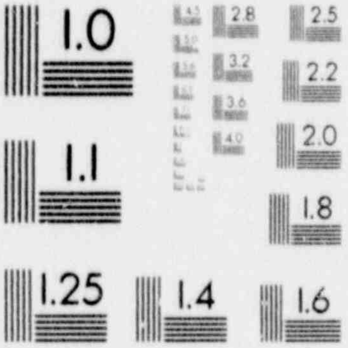
Before eliminating a particular scenario because of a physically illogical sequence of events, it is instructive to ask several experts to explain the presumably illogical sequence. In the above example, some experts may find the combination of icing and exploratory drilling illogical. But others may speculate that the exploratory drilling for some yet unvalued mineral would go on all over the world even in unfriendly climates, just as it is going on in the polar regions today.

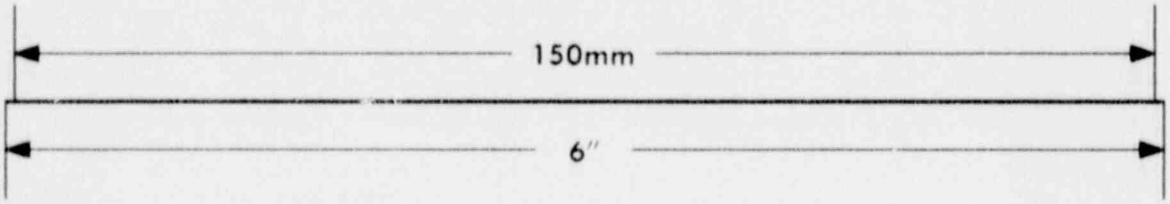


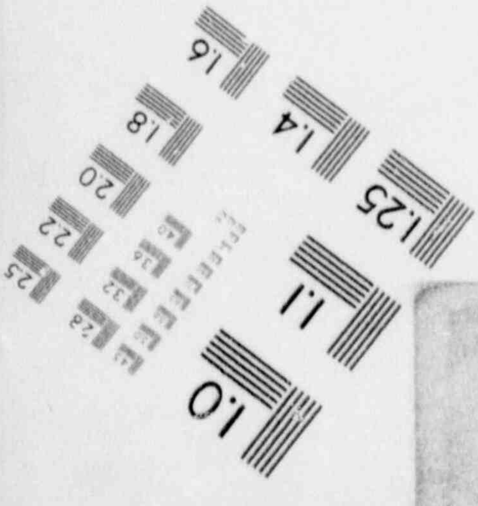
1

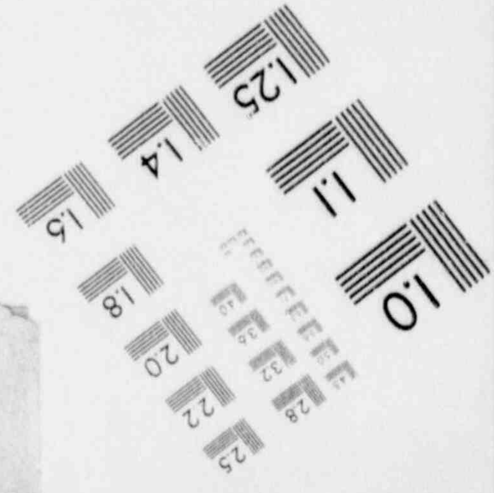
IMAGE EVALUATION
TEST TARGET (MT-3)





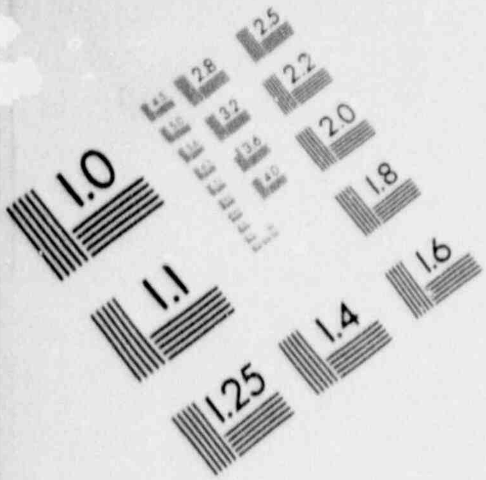


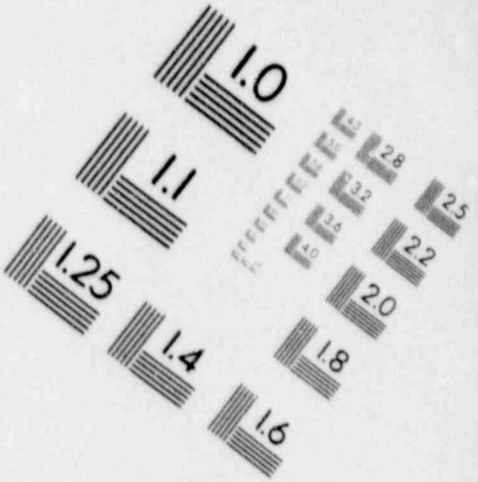


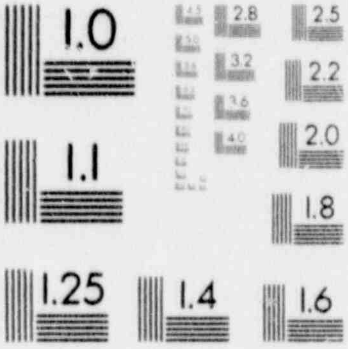


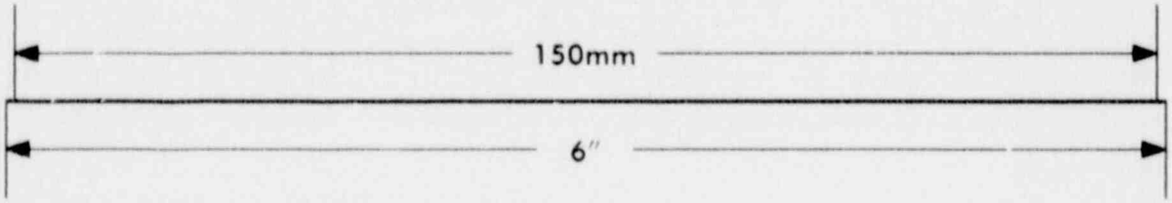
1

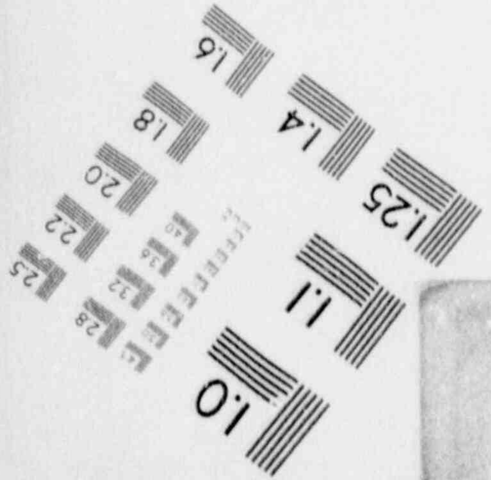
IMAGE EVALUATION
TEST TARGET (MT-3)

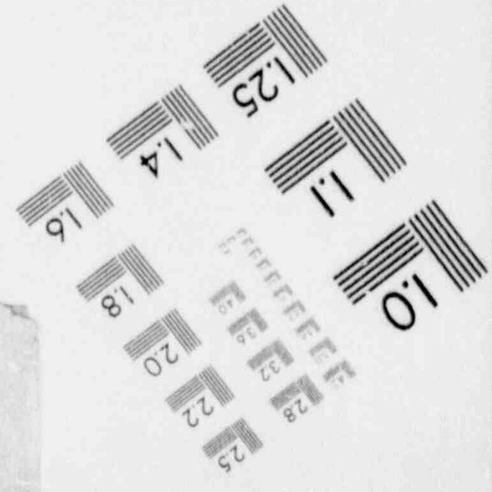






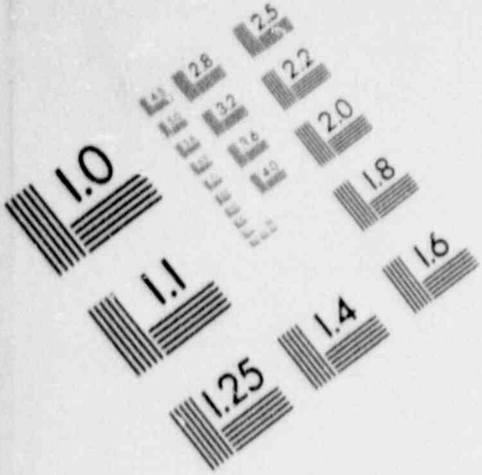


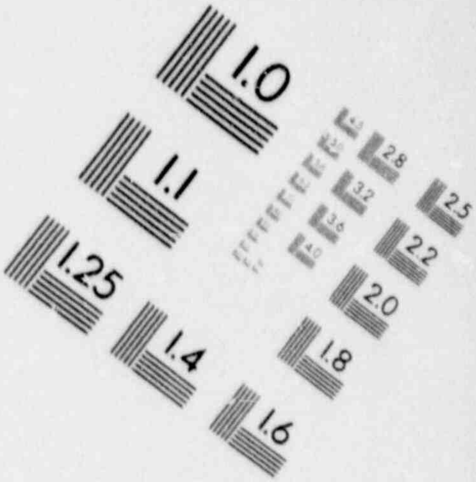


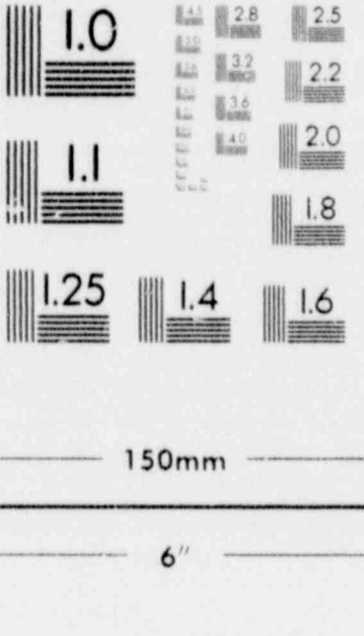


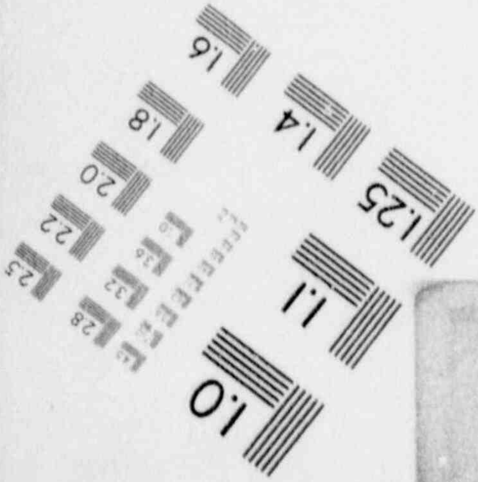
1

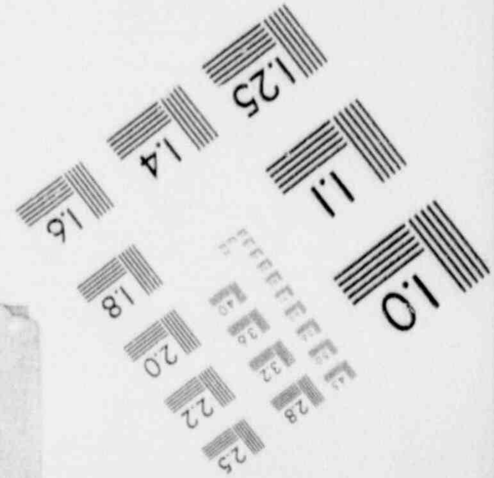
IMAGE EVALUATION
TEST TARGET (MT-3)











Scenarios can also be screened on potential consequences, eliminating scenarios with relatively insignificant impacts, and probability. Probability criteria can be defined on the whole scenario, on individual events, and on part of the sequences of events. In addition, probability criteria can be set differently, depending on the consequences of a scenario. It is useful to spell out different sets of probability criteria and investigate their use before fixing target levels and constraints.

Attributes for screening conceptual models can be very diverse. Examples include scientific acceptance, predictive ability, ability to estimate the parameters, simplicity, and cost. Techniques for identifying and structuring such attributes are described in Section 3.3.3.

3.3.2.1 Setting Target Levels or Constraints

In scenario screening, a main issue is the selection of probabilities to screen out extremely low probability events or scenarios, eliminating those that most people would consider "incredible," "implausible," "virtually impossible," or even "unbelievable" or "inconceivable." These target probabilities can pertain to an event in a scenario or to the total scenario. These probabilities are linked, as the probability of any event in a scenario must be larger than the probability of the scenario. In other words, if a single event in a scenario has probability p, then the scenario has to have a smaller probability pq, where q is the conditional probability of all the other event elements of the scenario given the event under consideration.

When setting event or scenario screening probabilities, one should consider the possible consequences. A common technique is to define smaller screening probabilities on overall scenarios if the possible consequences are more significant. For nuclear power plant accidents, for example, a screening probability for a core meltdown may be 10^{-6}, but the screening probability for a core melt with containment failure may be set as low as 10^{-9}. A more explicit approach is to set a target level on the probability distribution or, alternatively, on the complementary cumulative density function (NRC, 1975). Yet another approach is to combine target levels with potential benefits as described in Wilson (1984).

Screening conceptual models is more complicated, since there are more attributes to consider. Keeney (1980) discusses this issue in the context of screening alternative sites for energy facilities. He points out that screening is a simplified selection process and as such requires value tradeoffs among the screening attributes.

To illustrate this point, consider two screening attributes of conceptual models: cost of computer run time and empirical validity. One could set target levels on both attributes. For example, one could say that to be selected, a model run should not cost more than $10,000 and the expected error in predicting radionuclide travel time should be less than 100 years.

Alternatively, one could set the target levels at a model run not costing more than $10,000 but an expected error in predicting radionuclide travel time of less than 50 years.

Notice that the second set of target levels is more restrictive on the empirical validity attribute. Thus, in effect, by using the second set of target levels, we assign more weight to the attribute of predictive validity. This is a general feature of setting target levels and constraints: setting these levels by itself involves crucial value tradeoffs among attributes.

Multiattribute utility analysis (Keeney and Raiffa, 1976) makes these tradeoffs explicit and could be used to set constraints and target levels. While a full-fledged multiattribute utility analysis may be too costly for the purpose of screening, it is important to be cognizant of the tradeoffs made when setting target levels and constraints. As a practical rule, it helps to set target levels and constraints interactively, starting with very lenient levels and examining the set of surviving conceptual models after each setting of target levels.

3.3.2.2 Selection

Once attributes and target levels or constraints are defined, the selection is essentially mechanical. It is useful, however, to reiterate and go through a number of changes in setting target levels and constraints to investigate their implications for the selected subset. It is also useful to explain the logic of the process to a broad range of interested parties and to let them critique both the process and the result.

3.3.3 Decomposition Techniques

Problem decomposition is widely used in scientific study to simplify a complex problem into components that are more manageable and more easily solved. Problem decomposition has also been recognized as an important tool in expert judgment elicitation (Raiffa, 1968; Brown, Kahr, and Peterson, 1974; Armstrong, Denniston, and Gordon, 1975).

Problem decomposition in elicitation refers to breaking down issues to provide for easier and less complex assessments that can be recombined into a probability distribution or utility function for the quantity of interest. The recombination is usually accomplished through a mathematical model that expresses the quantity of interest as a mathematical function of component quantities. The techniques decomposition depend on whether the problem is a factual or value problem. Event trees, fault trees, and functional decompositions are used for factual issues, and objectives hierarchies are used for value issues.

3.3.3.1 Decomposition of Factual Problems

Several types of decompositions facilitate expert judgment about facts and probabilities. A familiar type of decomposition is the fault tree (McCormick, 1981), which focuses on a possible failure of a system and traces back the possible component causes of this failure. Fault trees are commonly represented as circuit diagrams that display the relations among system components and the failure of a system. In fault tree analysis, the components are assigned probabilities of failure, from which overall failure probability of the system can be found. Usually failures of various components are treated as independent events, although sometimes common causes lead to related component failures. Fault trees serve as a vehicle for the decomposition of expert judgments when the component events are dichotomous (0 to 1), independent, and the overall failure event is logically related to the component events. However, when decomposing, care must be taken to ensure that completeness is not lost. When finer detail about the causes of failure of some event in a fault tree is sought, experience suggests that incompleteness can easily occur (Fischhoff, Sloric, and Lichtenstein, 1978).

While fault trees end in a single failure event and trace its possible causes, event trees begin with an initiating event and draw out its possible consequences. The event tree lays out the sequence such that the probabilities of successive events are conditional on their predecessors. The branching in an event tree leads to a proliferation of paths, each path having a terminus associated with a system state or consequence. Event trees are a natural means of representation when phenomena have discrete outcomes. When the outcomes are continuous, however, the use of event trees requires that the continuous outcomes be approximated by a discrete categorization of ranges of the outcome variables.

A related type of decomposition uses the conditioning of possible events on known or hypothesized events (Bunn, 1984). The events can be laid out as an event tree where predecessor events are the conditions for the event in question. For instance, the probability of event A may be conditioned on the hypothetical events B and C. The assessment task then requires the probabilities of A, given various combinations of B and C and their complements. Further, the probabilities of B and its complement, given C and its complement, must be assessed as well as the probabilities of C and the complement of C. Denoting the complement of an event E by E , the probability of the event A becomes

$$\begin{aligned} P(A) = {} & P(A|B,C)P(B|C)P(C) + P(A|B',C)P(B'|C)P(C) + \\ & P(A|B,C')P(B|C')P(C') + P(A|B',C')P(B'|C')P(C') \;. \end{aligned}$$

Barclay et al. (1977) demonstrate the use of this style of decomposition to ascertain the likelihood that a nation will have the capability of producing nuclear weapons within a given time frame. An analysis and discussion of theoretical aspects of the probability decomposition are provided by Ravinder, Kleinmuntz, and Dyer (1988).

A tree structure related to the event tree is the decision tree (Raiffa, 1968; Holloway, 1979). In addition to possible events, decision trees incorporate choices or decisions that partially determine the path followed. Decision trees are particularly valuable in the evaluation of alternatives. Decision trees should be helpful in the analysis of information gathering activities associated with the potential repository and in evaluating design and construction options for the repository.

Decomposition may also use physical models of the phenomena being analyzed. The physical relationship between the quantity of interest and several constituent or determined quantities is expressed through a mathematical function such as $T = f(X,Y,Z)$. This type of decomposition is called algorithmic decomposition by MacGregor, Lichtenstein, and Slovic (1988). Rather than assessing a single probability distribution for T, the principle of decomposition leads to the assessment of probability distributions for X, Y, and Z that are combined to form a probability distribution for T. If the expert is better able to express knowledge about the constituent quantities than about the original quantity, the issue is a good candidate for decomposition. This strategy has been used in the reactor risk reference document (Wheeler, Hora, and Cramond, 1989), and the EPRI study of seismicity (Electric Power Research Institute, 1986).

If the expert possesses knowledge about X, Y, and Z and, further, knows the functional relationship f, then the expert should be able to give equivalent assessments either in terms of T or in terms of X, Y, and Z. However, the combination of X, Y, and Z is likely to be too complex for the human mind to do without substantial assistance. Decomposition, then, can serve as an aid to human thought processes in that the mind is relieved of tasks that it is ill-equipped to perform (Einhorn, 1975).

3.3.3.2 Decomposition of Value Problems

The best-known technique for decomposing value problems is structuring so-called objectives hierarchies. Objectives hierarchies structure the expert's general value concerns, intermediate objectives, and specific value-relevant attributes in a tree-like hierarchy in which the lower levels define what is meant by the upper levels (Keeney and Raiffa, 1976; von Winterfeldt and Edwards, 1986). Objectives hierarchies are structured by either the top-down or the bottom-up approach. Both approaches are implemented in interviews with experts knowledgeable about the value domain considered. They are illustrated below with an example of evaluating alternative conceptual models.

The top-down approach begins with general value concerns like *costs, scientific validity*, etc., and subsequently specifies the meaning of these general terms at increasing levels of detail. For example, *scientific validity* could be broken down into *face validity, empirical validity*, and *axiomatic validity*. *Empirical validity* could be further broken down into *experimental validation* at the *repository* and *empirical validition at other sites*. When considering a hierarchy of concerns, objectives, and attributes, it is important to pursue and to eliminate means objectives.

The bottom-up approach begins with listing the features that differentiate the options. From this list, features are eliminated that are not relevant for comparative evaluation. Among conceptual models, for example, *average run time* is value relevant because of cost and delay of feedback. On the other hand, *place of development* may not be value relevant. Having screened for value relevance, the next step is to eliminate means and pursue ends. Finally, the remaining features are clustered and organized into a logical hierarchy.

The results of the top-down and bottom-up approaches should be similar hierarchies with general value concerns at the top and specific attributes at the bottom. Once a first-cut hierarchy is built, the following checks can be used to examine and revise it:

- Are any concerns, objectives, or attributes redundant?
- Is the set of concerns, objectives, and attributes exhaustive?
- Are the concerns, objectives, and attributes independent?
- Is the tree manageable for further analysis?
- Are the lowest level attributes operational; that is, can one measure and compare, for example, conceptual models on them?

DRAFT

Checking and revising often involves returning the initial hierarchies to the experts for reexamination.

The previously described decompositions of factual and value problems are fairly formal in that they express the results as trees or functions.

3.3.3.3 Variants of Decomposition

Decomposition can also be used less formally. The goal of a less formal procedure might be to promote deeper insight into the rationale for judgments and to enhance the interchange of beliefs and assumptions about the likely causes of studied events without formally encoding the decomposition. The decompositions might be in terms of casual or mitigating factors that are loosely related to the event or quantity of interest. In this form, decomposition enhances the experts' introspection and communication.

A key aspect of decomposition relates to the source of the model or models used as a decomposing framework. The models can be imposed upon the experts from an external source, or they can be generated by the experts. Individual experts may be allowed to choose their own decompositions, or a consensus decomposition may be used.

Using a single decomposition has several advantages. First, the costs of recombining the judgments may be substantially reduced. Experience with NUREG-1150 indicated that the effort to process elicitations from multiple experts who used unique decompositions was much greater than expected (Wheeler, Hora, and Cramond, 1989).

Another potential advantage of using a single decomposition is that comparisons can be made among elicitations for component quantities and events. Combining assessments at the component level and then recomposing is also feasible when a single model is employed. Neither comparison at the component level nor aggregation at a subissue level is feasible with multiple decompositions.

A single decomposition by multiple experts also has important drawbacks. First, there needs to be significant discussion to ensure that all experts understand and accept the chosen decompositions, which is often difficult to achieve. Second, the influence that a decomposition has on the ultimate result is considerable. Requiring experts to abide by a single model may force their judgments to appear to be in agreement and thus understate their underlying differences as to the appropriate processes and assumptions. And if the decomposition itself is somewhat faulty, the results can be misleading. It is important to recognize that the decomposition itself embodies much information.

The advantage of multiple decompositions is that a wider variety of approaches to the problem are permitted. Single decompositions may understate the true uncertainty about an issue because the experts are forced to conform to a single view. Research has shown that the method of analysis, or decomposition, is important in forming judgments (Booker and Meyer, 19__). Multiple decompositions also provide a vehicle for discussion and documentation of alternative viewpoints--an important by-product of the expert-judgment process.

When an issue requires the expertise of several experts, decompositions are particularly useful. Teams of experts who collectively possess the requisite knowledge may be formed to address the issue. Each team member must embrace his or her portion of a collectively acceptable model so that the team's judgments are coherent and based upon the same conditions and assumptions. In such a setting, the decomposition separates the issue into components that can be addressed by members of the team having the relevant expertise. The decomposition also is the basis for integrating the assessments of the team members. A team format where teams had the flexibility to modify their models was used in a seismicity study of the Eastern United States (Electric Power Research Institute, 1986).

3.3.3.4 Benefits and Costs of Decompositions

Decomposition beyond a point may detract from the quality of the information obtained. Decomposition should be done until a balance exists between the difficulty of the assessments, the complexity of the decomposition, and the inherent number of assessments that must be made. In some instances, no decomposition may be desirable.

Problem decomposition is beneficial in two ways. One is that the expert judgments obtained through decomposition may better represent the true state of knowledge about the problem. This is because simpler assessments can be made more accurately by the experts because their answers will be better calibrated. Psychological biases such as overconfidence and the base-rate phenomena are thought to be less pronounced for easy tasks than more difficult tasks, so decomposing into easier tasks may lessen the impact of these biases (Merkhofer, 1987; Lichtenstein and Fischhoff, 1980). Mathematical recomposition of assessments relieves the expert of a difficult integration or aggregation task.

The second type of benefit from decomposition is the stimulation of alternative views and the documentation of reasoning that follows naturally from a decomposition. The use of multiple decompositions also helps explain why experts differ in their rationales.

Cost may be relevant when considering decomposition. The number of assessments may increase substantially because many questions may be required for a single issue. Beyond this expense, an additional requirement is that computer programs or other methods be constructed to perform the recomposition. The diversity of potential decompositions often precludes the use of existing software. Significant analyst effort is usually required to recompose an issue. Decomposition may also produce the false impression of objectivity and sometimes may introduce bias by systematically omitting an important component.

3.3.4 Techniques for Quantifying Probability Judgments

Probability elicitation techniques are described in several references (e.g., Spetzler and von Holstein, 1975; Selvidge, 1975; Seaver, 1978; Keeney, 1980; Stillwell, Seaver, and Schwartz, 1981; Wallsten and Budescu, 1983; von Winterfeldt and Edwards, 1986; Merkhofer, 1987). In addition, several reviews of experimental validation of these techniques exist (Peterson and Beach, 1967; Goodman, 1972; Lichtenstein, Fischhoff, and Phillips, 1977, 1982; Slovic, Fischhoff and Lichtenstein, 1977; Pitz and Sachs, 1984). Drawing on this literature, there appear to be four distinct classes of procedures, depending on the nature of the uncertain quantity (discrete events vs. continuous random variables) and the nature of the questions asked (magnitude judgments about events vs. indifference judgments about gambles). The resulting taxonomy is shown in Table 3.1.

The eight techniques listed in this taxonomy are the most commonly used ones in the quantification of probability judgments. Before describing these techniques in detail, it is useful to spell out some general guidelines for probability elicitation that are applicable to all eight techniques.

Table 3.1

Taxonomy of Probability Elicitation Techniques

Variable	Judgment	
	Magnitude judgments about events	Indifference judgments about gambles
Discrete Events	Direct probability Direct odds	Reference gambles (discrete) Certainty equivalent (discrete)
Continuous Quantities	Fractile technique Interval technique	Reference gambles (continuous) Certainty equivalent(continuous)

First, it is important to begin with easy questions. For example, when comparing the probabilities of two rare events, an expert may initially have no feeling for the absolute magnitude of probabilities, but it may be fairly easy to establish a rank order of the relative likelihood of the events. Second, it is preferable to select observable quantities for eliciting probabilities. As an specific case, one observes failures of equipment rather than failure rates. Assessing the cumulative probability for the number of failures with 100 units originally operating for a fixed time period in extreme conditions may be easier than assessing the probability for the likelihood (i.e., a parameter) that an individual unit will fail in that time period with those conditions. Third, it is useful to ask the same question in different ways and to use the results for consistency checks. These consistency checks should not be presented as a challenge to the expert, but rather as a means to stimulate thought and to improve judgments. Fourth, it helps to have computer support for decompositions, reaggregation, consistency checks, and displays.

3.3.4.1 Magnitude Judgments about Discrete Events

The techniques described in this subsection involve two or any finite number of mutually exhaustive and exclusive events to which probabilities have to be assigned by making direct numerical magnitude judgments. These probabilities should add to one by virtue of the addition law of probability. For two events, one need elicit only one of the probabilities, but it is good practice to check on the other one as well. For multiple events (e.g., 10 or more), it is usually worthwhile to reconsider the event space, either by clustering events or by identifying the continuous quantity that corresponds to the events. Frequently, with a continuous quantity, it is easier to construct probabilities for many events, since one can exploit monotonicity, single peakedness, and other properties of the probability distribution.

Direct Probability. This is perhaps the simplest technique. The elicitator asks the expert, "What do you think the probability is that this event occurs and why?" Often it is useful first to obtain a rank order of the probabilities of the events considered. In the case of two events, the first question may be which is more likely and why, followed by a judgment of the magnitude of the probability for the more likely event, and finished by the judgment of the probability of the less likely event. Assuming that the two events are mutually exclusive and collectively exhaustive, these two probability judgments would, of course, have to add to one.

For more than two events, there are two variants of this procedure: one can either ask the expert to assign probabilities to each event separately without the constraint of adding to 1.0 or to do so with that constraint. When time permits, it may be desirable to ask the questions without constraints and check the sum. This sum will often be larger than 1.0, since experts tend to overestimate probabilities, especially when they are small. Adjustments will then be necessary so that the revised sum is 1.0.

Direct Odds. Sometimes the probabilities of events are hard to judge abstractly, but easier to judge in comparison. In this case, the normative expert can ask the substantive expert to state the relative odds of one event in favor of the other for selected pairs of events. If there are two mutually exclusive and exhaustive events A and B, the expert would need only to state the odds O(A), in favor of A over B. From O(A) the probability of A can be calculated as

$$p(A) = O(A)/\{1 + O(A)\} ,$$

from which the probability of B follows. Similarly, for n events, the expert needs to assign n-1 odds, and the resulting probabilities can be calculated. However, as in the direct probability procedure, it might be useful to elicit n or more odds, point out the inconsistencies, discuss them and resolve them.

3.3.4.2 Magnitude Judgments about Continuous Uncertain Quantities

The uncertain variable in this category is a continuous numerical quantity. The techniques described in this subsection also apply if the variable is dense and has interval quality. The two magnitude judgment techniques are mirror images of each other. In the *fractile* technique the normative expert provides the substantive expert with a probability and asks for a magnitude of the uncertain quantity such that the

DRAFT

probability of the true value falling below it is equal to that probability. In the *fixed point* technique, the normative expert provides the substantive expert with a set of fixed points of the uncertain quantity and asks for the probability corresponding to these fixed points or for intervals in between them.

3.3.4.3 Fractile Technique

The fractile technique is the most widely used probability elicitation technique for continuous uncertain quantities. It is used to construct the cumulative density function of the uncertain quantity that describes the expert's current state of knowledge. A z-fractile is that magnitude x_z of the uncertain quantity x such that there is a probability of z that the true magnitude falls below x_z and a 1-z probability that it falls above it. The lower bound therefore should be the 0.0-fractile and the upper bound should be the 1.0-fractile. The cumulative density function simply plots the fractiles against the probabilities that the actual magnitude falls below it.

After carefully defining the uncertain quantity, the substantive expert is asked to state its upper and lower bounds. In other words, he or she should define two magnitudes such that there is absolute certainty that the true magnitude would fall in between these extremes. In practice, because a continous variable may have no obvious lower or upper bound, assessments may focus on the 0.01 and 0.99 and/or on the 0.05 and 0.95 fractiles as relative extremes. After the initial extremes are defined, it is often useful to ask probing questions. The substantive expert is asked to consider a hypothetical event in which the actual magnitude of the variable considered was found to lie outside the range of extremes. Can this event be explained? Clearly, if any credible explanation exists, the extremes were not 0.0- and 1.0-fractiles. Credible explanations also provide a basis for estimating the probabilities of being outside the extremes. Such considerations can lead to revisions of the initial extremes.

After having obtained the extremes, the normative expert typically moves to the middle range of the uncertain quantity and attempts to identify the magnitude of the uncertain quantity such that the substantive expert thinks the chances are about 50-50 that the actual magnitude would fall above or below that value. This point is called the median or the 0.5-fractile of the cumulative density function. The answer should be probed, especially if it falls exactly in the middle of the range between the extremes (since this suggests arithmetic averaging) or if it is very close to one extreme (since this suggests poor definition of extremes or a poor selection of the scale and unit of measurement).

Having obtained three points of the cumulative density function (the extremes and the 0.5 fractile), the remaining tasks are to elicit between two and four additional fractiles. If they have not been determined in setting extremes, it is often useful to elicit the 0.05 and the 0.95 or the 0.01 and 0.99-fractiles next. To obtain the 0.05-fractile, the normative expert asks the substantive expert to state that magnitude of the uncertain quantity such that the probability of the true magnitude falling below it is 0.05. Finally, the 0.25 and 0.75 fractiles are commonly assessed.

Usually knowing the extremes and five fractiles is sufficient to sketch a cumulative density function. The normative expert should smooth a graph of this function and discuss its shape with the expert. In addition, it is very helpful to show the plot of the corresponding probability density function, which shows the symmetry or asymmetries of the cumulative density function more clearly.

3.3.4.4 Interval Technique

In the interval technique the normative expert preselects points of the uncertain quantity and asks the substantive expert to assign them probabilities. There are two versions of this method. In the open interval version, the substantive expert assigns probabilities that the actual magnitude falls into the open intervals below and above each selected point. In the closed interval version, the substantive expert states the probabilities that the true magnitude falls between the preselected points.

Both versions of the interval technique begin with extremes, preferably bounds or the 0.01 and 0.99 fractiles, just as in the fractile technique. In the open interval version, the normative expert then chooses three to seven points between and asks, for each point, what the probability is that the actual magnitude of the uncertain quantity is above or below that point. Having obtained these probability judgments, the normative expert can then smooth a cumulative density function and proceed as with the fractile procedure.

In the closed interval version, the normative expert again lays out three to seven points, possibly equally spaced, but this time asks the substantive expert to assign probabilities that the true magnitude falls in each of the intervals. The result can be plotted both as a cumulative density function or as a probability distribution. It is useful to begin by rank ordering the probabilities of the intervals before assigning actual probabilities.

Both versions can be used in consistency checks. In addition, the fractile method can be mixed with the fixed-point method. It is quite easy, for example, to infer fractile-type questions from interval elicitations and to construct interval-type questions from fractile-type results. For example, after constructing the 0.25, 0.5, and 0.75 fractile, the substantive expert should consider the intervals below the 0.25 fractile, between the 0.25 and the 0.5 fractile, between the 0.5 and 0.75 fractile, and above the 0.75 fractile to be equally likely.

3.3.4.5 Indifference Judgments Between Gambles with Discrete Events

The techniques discussed in this subsection derive probabilities from comparisons among gambles with discrete events and (usually hypothetical) monetary outcomes.

Reference Gamble Technique. To illustrate the reference gamble technique, the expert is asked to select one of two gambles. The first gamble involves the event "It will rain tomorrow" with unknown probability. If it rains, the expert will receive a stated prize; if it does not, he will receive nothing. Alternatively, he can choose the gamble in which he receives the prize with known probability p or otherwise nothing with probability 1-p. If the expert bets on rain, the probability p is reduced until the expert is indifferent between the two gambles. If indifference occurs when the probability is p_I, this probability is assigned to the likelihood of the event because the expert should be indifferent when there are equal chances of winning the prize with both gambles.

Certainty Equivalent Technique. The certainty equivalent technique is somewhat simpler in that it asks only for comparisons between one gamble and one sure amount rather than between two gambles. However, in order to use it, one must verify (or assume) that the substantive is an expected value maximizer. To illustrate the technique, consider again the gamble for \$10 if it rains vs. nothing if it does not. The normative expert asks the substantive expert to state a certain amount of money at which he would be indifferent between playing the gamble or taking less as a gift. To facilitate thinking about this question, the normative expert could begin by asking whether the substantive expert would prefer a certain amount of \$1 over playing the gamble. If the substantive expert emphatically says that he would prefer to play the gamble, the normative expert could change the certain amount to, say, \$9. At this point the substantive expert may consider the certain amount to be much more attractive. The normative expert then continues to vary the certain amount until the substantive expert is indifferent between the choices. At this point, the certain amount is said to be the certainty equivalent of the gamble.

Assume, for example, that the certainty equivalent in this case is \$7. Then, by the assumption of the expected value principle,

$$\$7 = p(\text{Rain})\$10 + p(\text{No Rain})\$0$$

or

$$p(\text{Rain}) = .70 \ .$$

Similar schemes can be devised with multiple event gambles.

3.3.4.6 Indifference Judgments among Gambles with Continuous Uncertain Quantities

This report will not describe indifference techniques for continuous variables as they are direct extensions of the techniques for discrete events. The main idea in applying these techniques to continuous quantities is to discretize these variables using ranges of values and to apply the indifference techniques to the discretized events (Matheson and Winkler, 1976).

3.3.5 Techniques for Quantifying Value Judgments

Many expert judgments related to the performance of an HLW repository will include value judgments, especially in screening scenarios and selecting conceptual models. It is always important to make these value judgments explicit and document them carefully. In some cases, it also may be important to quantify value judgments with multiattribute utility elicitation techniques (Keeney and Raiffa, 1976; von Winterfeldt and Edwards, 1986). These techniques range from simple rating techniques to sophisticated indifference techniques to multiattribute utility functions. This section describes two techniques with different degrees of technical sophistication that are applicable to the task of evaluating conceptual models: the simple multiattribute rating technique (Edwards, 1977) and an indifference technique to elicit a measurable multiattribute value function (Dyer and Sarin, 1979). These techniques are fairly similar in the basic task structure, but differ in the procedure of the elicitation.

DRAFT

There are seven steps in an evaluation:

1. Define the objectives for evaluation.
2. Develop attributes and scales for measuring the objectives.
3. Estimate the performance of the alternatives with respect to each attribute.
4. Develop single attribute value functions.
5. Develop weights for the attributes.
6. Convert the performance estimates of step 3 into single attribute values using step 4.
7. Calculate an overall value for the alternative, typically by a weighted average using the weights in step 5.

The simple multiattribute rating technique and the measurable multiattribute value function technique differ primarily in steps 4 and 5. In the rating technique, both single attribute value functions and weights are elicited using direct numerical rating judgments. In the indifference technique, both elements are elicited using tradeoffs and indifference judgments. Before detailing these techniques, we will briefly discuss steps 1 to 3.

The objectives hierarchy provides a logical structure of the objectives for evaluating the alternatives (i.e., conceptual models). We discussed some principles for constructing an objectives hierarchy in Section 3.3.3 on decomposition techniques for value problems.

Developing attributes and scales that measure the objectives in the objectives hierarchy is still an art. There are two types of attribute scales: natural and constructed. Natural attribute scales are numerical scales commonly used. For example, run time of a conceptual model may be defined in terms of seconds of CPU time. A constructed scale is needed when no natural scale is available or convenient. An example is scientific acceptability of a conceptual model. In this case a scale can be constructed that defines qualitatively (perhaps a paragraph or more) several distinct achievement levels. For example, the worst level could be defined as "a conceptual model that has virtually no scientific acceptability, only a few supporters, and very little published evidence supporting it." The best level could be defined as "a conceptual model that has very high scientific acceptability, many supporters of high scientific status, and significant published support." Similarly, intermediate levels could be defined.

The next step (step 3) estimates the performance or achievement of each alternative on each of the attributes. This is a nonprobabilistic version of an expert elicitation. In the assessment of conceptual models, a group of experts may be convened who estimate attributes such as run time, scientific acceptability, cost, etc. If the uncertainty about these estimates is significant and if it is important to quantify this uncertainty, complete probability distributions should be elicited using the techniques in Section 3.3.3. With uncertainty, a multiattribute utility function, rather than a value function, will be necessary to compare alternatives.

3.3.5.1 Simple Multiattribute Rating Technique

To construct single attribute value functions with this technique, the worst and the best levels of the attribute scale are identified and arbitrarily assigned a value of 0 and 100, respectively. For natural scales, several values between the worst and the best level are then selected and rated on the 0 to 100 scale. The resulting points are plotted, and a single attribute value curve is fitted. For constructed scales, each constructed level is rated on the 0 to 100 scale. The same process is followed for all attributes.

To obtain weights for the attributes, two hypothetical alternatives are constructed, one representing all the worst attribute scale levels, one representing all the best. The expert is then asked to imagine being stuck with the worst alternative. Which attribute would he or she like to change most from its worst to its best level? Which is second, etc.? This ranks the value differences for attribute ranges between worst and best levels of the attributes.

Next, the attribute range that was ranked highest (i.e., which the expert would like to change the most) is assigned 100 importance points and an attribute range (not necessarily in the list) that is utterly unimportant is assigned 0. All other attribute ranges are rated between, according to their relative importance. The resulting raw range weights are normalized to add to one.

3.3.5.2 Indifference Technique for Measurable Value Functions

To obtain single attribute value functions, an indifference technique called bisection is used. The expert is again presented with the worst and the best levels of an attribute. Next, he or she is asked to identify a mid-level of the attribute (not necessarily the numerical mid-point) such that the increase in value obtained by stepping from the worst level to the mid-level is equal to the increase in the value obtained by stepping from the mid-level to the best level. This mid-level is the value midpoint. By arbitrarily assigning a value of 0 to the worst level and a value of 100 to the best level, the value midpoint has a calculated value of 50. By further bisecting the range between the worst level and the value midpoint, the value midpoint and the best level, etc., a value function can be defined to any reasonably achievable detail. For attributes with natural scales, the results can be plotted as a value function. This process is repeated for all attributes.

To elicit the weights, the expert is presented with two hypothetical alternatives that vary only on two attributes, while all other attributes are held constant at some level. The first alternative has the worst level of attribute A and the best of attribute B. The second alternative has the best level of attribute A and the worst of attribute B. The expert is asked to state a preference for one of the alternatives. If the preference is for the first alternative, he or she is asked to worsen the level of attribute B in the first alternative until both alternatives are indifferent. If the preference is for the second alternative, the expert worsens the level of attribute A in the second alternative until both alternatives are indifferent. In either case, the elicitator assists the expert by providing easy comparisons along the way to indifference.

Once the indifference is established, the *relative* weights for attribute A vs. attribute B can be calculated assuming an additive value model. Let $(a_o,b^*,c,d,...)$ be the first alternative with the worst level of attribute A and the best level of attribute B, and let $(a^*,b_o,c,d,...)$ be the second alternative with the best level of attribute A and the worst of attribute B. Both have identical levels c, d, etc., of attributes C, D, etc. If the first alternative is preferred, then attribute B should be worsened to, say, level b' to achieve indifference. The indifference means that the overall values, denoted by v, of the alternatives are now equal so

$$v(a_o, b\ , c, d, \ldots) = v(a^*, b_o, c, d, \ldots) \ .$$

Using the additivity assumption, we can write

$$w_A v_A(a_o) + w_B v_B(b\ ) + w_C v_C(c) + w_D v_D(d) + \ldots =$$

$$w_A v_A(a^*) + w_B v_B(b_o) + w_C v_C(c) + w_D v_D(d) + \ldots ,$$

and since, by definition,

$$v_A(a_o) = v_B(b_o) = 0 \text{ and } v_A(a^*) = v_B(b^*) = 100 \ ,$$

$$w_A/w_B = v_B(b\ )/100 \ .$$

Obtaining n-1 such equations and using the convention that the weights should add to one provides the solution for the weights in this procedure.

3.3.5.3 Aggregation Steps

Step 6 is identical for both techniques. It consists of a mechanical conversion of the performance measures obtained in step 3 into single attribute values using the results of either the rating or indifference technique. Step 7, also identical for both techniques aggregates single attribute values and weights to a weighted sum. Having completed a full cycle using these techniques for making value judgments, it is good practice to compare the calculated results with the experts' intuition and to iterate.

3.4 Combining Expert Judgments

When using a panel of experts, there are three basic reasons to combine the judgments of individual experts. The first is to provide a base case, or more than one base case, for analysis and sensitivity analysis in the performance assessment. The second is to gain insights from the analysis for decision making. The third is to simplify analyses and, therefore, to save time and effort in acquiring these insights.

Depending on the types of judgments, combining expert judgment takes somewhat different forms. In the qualitative expert-judgment tasks (identification and screening), the combination consists of generating a joint list of things such as initial events and processes or screened scenarios. In probability judgment, individual probabilities or probability distributions are combined. In value judgments, individual functions or weights are combined.

3.4.1 Combining Lists

The simplest approach to create a joint list is to take the union of the individual lists. Often, the creation of the joint list involves some restructuring and some relabeling. Such changes should be communicated to the experts that created the individual lists, and care should be taken to assure that their individual concerns are reflected in the joint list. Beyond these suggestions, however, there is little technical advice about how to combine qualitative information.

3.4.2 Combining Probability Judgments

A key issue in combining probability judgments concerns what should be combined. The answer in almost all cases is that the overall probability judgments of the individual experts or expert teams should be combined. These overall judgments are typically a joint probability distribution function over the set of technical variables. Combining at this level recognizes that the fundamental unit in expert assessment is the state of knowledge of the expert. By combining across the complete representation of experts' knowledge, different experts can use different models, logic, data, and processes to develop and represent their overall judgment. Combining experts' judgments at component levels in the process (e.g., combining marginal probability distributions) would put severe restrictions on the assessments of the individual expert. Each of the experts would essentially have to go through the same reasoning processes and provide the same intermediate representations of knowledge. In addition, if experts are in disagreement on their judgments and if the judgments are combined at component levels, you can develop situations in which the overall judgments of each expert would lead to a preference of an alternative A to an alternative B, but where alternative B would be preferable using the combined judgments (Raiffa, 1968).

3.4.3 Combining Value Judgments

As with probability judgments, the appropriate level of aggregation is at the level of overall utility functions, not at the level of single-attribute utilities or value tradeoffs. There are, however, additional problems with aggregating utilities (Arrow, 1951; Keeney and Raiffa, 1976). These problems are a result of the difficulty of making impersonal comparisons of utility. As a practical solution to this comparability problem, Keeney and Raffia (1976) propose the concept of a supra decision maker that is to incorporate the value judgments of each individual decision maker. Using the supra decision maker model and making certain regularity assumptions, it is reasonable to aggregate individual (overall) utilities as a weighted average.

With value judgments, a fair amount of agreement usually exists about the general nature of the single attribute utility functions (see Section 3.3.4). In particular, agreement is likely to be found about the direction and the monotonicity of the utility function. If the utility functions have very different shapes, the underlying attribute may not have been clearly defined. On the other hand, weights are very personal expressions of value judgments and value tradeoffs. It is impossible to speak of "better" or "correct" weights. Experience has shown that in many controversial problems, the differences in value judgments appear as legitimate differences in weights (Edwards and von Winterfeldt, 1987).

3.4.4 Behavioral vs. Analytical Combination

The two general approaches to combining expert judgments are referred to as the behavioral approach and the analytical approach. With the behavioral approach, the experts on a panel are brought together to discuss and combine their judgments. In this process, the thinking, logic, and information of the different experts are exchanged. This may bring about some reconciliation of differences and result in a single representation of the state of knowledge, or it may minimize the differences among experts. The behavioral approach seems particularly useful when the experts have basic differences in fundamental assumptions upon which their judgments are based. In this situation, the interaction among experts promotes deep thinking about the problem that can lead to more thorough understanding and documentation. A possible serious disadvantage is that some experts may be dominated or "forced" to suppress their ideas to maintain harmony on the expert panel.

Analytical combination procedures are comprised of a logic and formulas consistent with that logic developed by the analysts (e.g., the normative experts) for combining individual judgments (Fischer, 1981; Genest and Zidek, 1986). The complete set of analytical combinations of expert judgments that seem reasonable for consideration is the convex combination of the individual expert judgments. In other words, it is the set of additive weightings of the various expert's judgments such that the sum of the weights is one. One of these combinations is the average of the various experts' judgments. Other combinations, in which the weight on one expert is one and weights on all the others are zero, are simply an expression of the state of knowledge of the individual rated one. The obvious advantages to analytical combination procedures is that they are easy to use, it is easy to do extensive sensitivity analyses around any base case combination, and individual experts have no influence on the judgments of other experts after the elicitation.

The most common analytical combination procedure is the average, in which all experts receive an equal weighting. A substantial amount of evidence suggests that this average weighting often produces a reasonable base case for analysis (Seaver, 1978; von Winterfeldt and Edwards, 1986). However, some experience suggests that differential weighting techniques to account for the relative expertise of individual experts result in a better combined representation of knowledge (Ashton and Ashton, 1985). One useful property of weighting techniques that positively weighs all individual assessments is that the full range of the variable under consideration is included in the combined representation. In other words, the weighting does not eliminate the range of diversity among different experts (Merkhofer, 1987). This property of combining judgments is of particular concern in risk analysis.

A combination of behavioral and analytical procedures can be used for combining individual experts' judgments. In this case, behavioral methods are first used. Here, the individuals exchange all their reasoning and data and assumptions upon which their judgments are based. If this process results in any changes of judgments by individual experts, the implications of these changes are included in updated representations of the individual expert's state of knowledge. If this process happens to lead to a commonly held representation of the state of knowledge, then that representation of each individual should also be the representation for the group. If, after behavioral aggregation approaches, there are still residual differences between the individual experts, these can be combined by an analytical procedure as outlined above.

Regardless of how expert judgments are combined, the resulting uses of the experts' judgments should recognize three important items. First, any report should include more than one possible combination. This should facilitate hard thinking about the implications of different combinations and inform readers that there is no absolutely correct way to do the combination. Second, different procedures for combinations may provide different insights from the analysis. For instance, if the combination is chosen that takes the "most conservative" estimate on any variable, the result should be a theoretical bound on the "most conservative" possible overall judgment based on the individual expert's judgments. If the analysis indicates, for instance, acceptable implications with these conservative (i.e., high) probabilities of failure, then perhaps no further analysis is necessary. Third, in all situations, the reported results should not be only combinations of the individual judgments. It is essential that the individual expert's judgments are also thoroughly reported and documented as discussed in Section 3.5.1.

3.5 Communicating Expert Judgments

3.5.1 Documentation

The reasons for documenting the use of expert judgment on technical problems are specified by the following objectives: (1) to improve decision making, (2) to enhance communication, (3) to facilitate peer review and appraisal, (4) to recognize and avoid biases in expert judgments, (5) to indicate unambiguously the current state of knowledge about important technical and scientific matters, and (6) to provide a basis for updating that knowledge.

Complete documentation of the use of expert judgment would include both the interaction with the experts and the results (i.e., expert judgments) of that interaction. Thus, documentation would describe the selection of experts, the decision on whether to have expert teams, and whether to have panels of specialists. Documentation would include the selection of the specific issues to be addressed by the specialists and how these were chosen. It would include the normative training about the methods used to elicit expert judgments from the specialists and the preparation process to provide any necessary or requested substantive information to the specialists. Finally, documentation would certainly include the results (e.g., probability distributions) from any elicitation of expert judgment, as well as the reasoning to support them.

The fundamental unit of information of explicit expert judgments is the information provided by each expert. Hence, in any documentation, it is crucial to clearly distinguish between the information provided directly by each expert and any processing of that information, such as smoothing, interpolation, extrapolation, combining of the judgments of different specialists, or drawing of inferences from the judgments of experts. Maintaining, as part of the documentation, the individual expert judgments, potentially provides more information for decision making than if the information were aggregated (Clemen, 1987).

The documentation of an individual's expert judgments should indicate what was done, why it was done, how it was done, who the individuals involved were and what their roles were, what the resulting judgments were, and what reasoning was used to support these judgments. The documentation should begin with a clear definition of the specific issue being addressed and should contain unambiguous definitions of all the specific terms used in the elicitation. All assumptions about conditions that prevailed or would prevail that relate to the expert judgment should be stated. For instance, if one is assessing judgments about ground-water travel times, assumptions about the particular rock types, the amount of fracturing in the rock, and the tortuousity of the rock might be assumed by a given expert. If so, these assumptions should be stated. The judgments as they are stated by the expert should be provided in the documentation. To support these judgments, the logic and data on which they are based should be completely specified. Any calculations that the expert considered important in determining his judgments or models used should be indicated. All literature, whether public or restricted, should be specified.

It is also important to document the approach by which the expert judgments were elicited. Some of this documentation may appear as a general section ahead of many elicitations since the procedure used for many expert assessments would be similar. However, the documentation would include both a description of the procedures and an explanation of why they were used, as well as examples of their use. In some specific problems, it is important to document what was not done. If some professionals are likely to question the process because of what was not explicitly done, clarification about why this was so may contribute to many objectives of documentation stated above.

The documentation should also indicate the types of consistency checks performed in the assessment of an individual's expert judgments. Invariably with complex expert assessments, such inconsistencies occur and are identified by these consistency checks. That is, in fact, one reason for going through a careful process to elicit expert judgments. Identification of the inconsistencies allows experts to understand their source and to adjust appropriately their judgments to account for this increased understanding. The final, consistent set of expert judgments are those utilized in the performance assessment and this set requires the documentation just described.

When a panel of experts is used for a problem, additional documentation is necessary. It is important to document how individual expert judgments are combined. The discussion in Section 3.4 indicates many guidelines for selecting a combination procedure. It is important to document the individual expert judgments in a common format and in the same format as the combination of expert judgments. The documentation should clearly indicate agreements and disagreements among the experts and the reasoning for any disagreements.

Documentation can take significant time and effort. Hence, it is very important to begin with a system for documentation and a standard form to be used in documenting all experts. Because the specific issues addressed by different experts may vary, this form must be general enough to handle a wide range of specific problems. The responsibility falls upon a normative expert to document the results of any elicitation of expert judgment and upon the generalists and specialists to document the technical and scientific reasoning that led to those results. However, once the documentation of an individual specialist's judgments is completed, it is important that the specialist review, making any necessary adjustments and then approving it as accurate.

Many factors need to be considered when selecting a documentation approach. Part of the documentation can include audio taping or video taping the elicitation sessions. With either, it is essential to provide written documentation in addition. In situations where there are many separate individual elicitations, it would probably be better to have the documentation of some elicitations more complete and polished than others. For example, with 100 elicitation sessions, each involving a specialist, a generalist, and a normative expert, it might be appropriate to have five of them carefully documented with a quality of writing appropriate for publication in peer-reviewed technical journals. The other expert elicitations should be documented with the same quality of logic, but not necessarily with the same thoroughness and style in writing appropriate for journal publication. This would save a great deal of time in documentation, and yet provide the essential information for achieving the objectives of documentation stated above.

The final issue about documentation concerns whether the experts should be anonymously treated or whether their names should be clearly assigned to their expert judgments. The main argument to maintain anonymity is that some experts might feel a pressure to take the "party line" of their organization if their name were associated with their judgments. With anonymity, they presumably could state what they really think. On the other hand, with the names of experts clearly stated along with their judgments, there is an additional motivation for the expert to be clear and thorough and consistent. Naming experts greatly enhances the perceived quality of the analysis and the ability of others to appraise and utilize the expert judgments. Indeed, experts typically possess a strong sense of responsibility for their judgments and a confidence about them. In other words, experts are willing to stand behind their judgments and have these represented as such (Shanteau, 1987). In the recent elicitation of expert judgments from approximately 50 experts in numerous disciplines for the NUREG-1150 project on the safety of nuclear power plants, only one indicated that he would prefer not to have his name attached to his judgments. Because of the importance to the overall study of attaching the experts' names to their judgments, one criterion in selecting experts should be the willingness to have his or her name associated with the judgments.

3.5.2 Presentation of Results

The presentation of results of expert elicitations discusses and appraises the insights from the expert judgments and their implications for decision making. The objectives of this presentation are to inform decision makers and others about these implications and to have a constructive influence on decision making. The presentation of results of expert elicitations is distinct from the documentation of the elicitations. Documentation simply states the results of the expert elicitations, but presentation uses the judgments of the analysts to appraise the relevance of the expert judgments to the decision faced.

It is important to recognize that the presentation of results is itself a decision problem for which there are many alternatives (Keeney and von Winterfeldt, 1986). How deep the presentation is, whether illustrative examples are used to indicate insights, and whether the insights are expressed mainly in qualitative or also in quantitative fashion are alternatives for that decision problem. These alternatives involve factors such as how and how much to use cumulative distribution functions or probability density functions (Ibrekk and Morgan, 1987), tables, diagrams, and decomposed probability trees. Alternatives also concern the degree to which there is comparability among the assessments of different experts. The presentation section may also contain decision analysis about the value of obtaining additional information regarding various uncertain phenomena investigated using expert judgment. Key considerations in deciding on a presentation alternative include for whom and for what specific decision-making purposes the presentation is prepared.

For an HLW repository, the performance assessment provides insights for technical and licensing decisions and for communication to government officials and the public. Presentation of the results of the expert judgments should indicate how these judgments relate to whether the repository can be safely operated and meet legal standards. The presentation should indicate clearly which of these judgments are crucial to decisions on whether the repository can perform safely and legally. It should also indicate what changes in these judgments might lead to different implications and the bases that could lead to those changes in judgments. The presentation of results should clearly indicate which disagreements between experts are relevant to whether the repository can be safely and legally operated, and which are important. Particularly for those that are important, it would be significant to indicate how one might resolve the disagreements among experts. This resolution might be possible simply with additional interaction among the experts, with additional experts, or only through additional gathering of data and scientific experiments.

3.6 Interpretation, Use, and Misuse of Expert Judgments

Expert judgments are crucial in the performance assessment of an HLW repository. However, as is the case with all scientific work, expert judgments can be misinterpreted, misrepresented, and misused. To enhance the likelihood that this does not occur, it is important to interpret and use expert judgment in performance assessment appropriately.

The formal use of expert judgment in performance assessment is a complement rather than a substitute, for other sources of scientific and technical information, such as data collection and experimentation. Expert judgments should not be considered equivalent to technical calculations based on universally accepted scientific laws or to the availability of extensive data on precisely the quantities of interest. Expert judgments are perhaps most useful when they are made explicit for problems in which site data are lacking, since they express both what the experts know and do not know.

Expert judgments are a snapshot of the state of knowledge of the individual expert about the stated item of interest. As new data, calculations, or scientific understanding become available, these should be systematically incorporated within the existing state of knowledge. This learning process, which is a natural part of science and knowledge, will result in changes in the expert's judgments.

Since different experts may have different information or different interpretations of information, there is no logical reason why various experts should have the same state of knowledge. For new and complex problems, a diversity of opinions might be expected. If such differences exist, these would clearly be identified in expert assessments. For a problem as important as the design and construction of an HLW repository, it is useful to know the range of expert interpretations.

Numerous expensive and lengthy projects have been suggested to investigate the physical conditions at a potential HLW repository site and the phenomena that affect those conditions. With the explicit use of expert judgment, the value of the information derived from such projects can be calculated. This provides a sound basis for selecting projects that should be pursued. When one recognizes that the combined cost of proposed projects is several billion dollars, the significance of systematically appraising proposed projects becomes obvious.

The main misuses of explicit expert judgments stem from misrepresentation or over-reliance on them. Expert judgments often have significant uncertainties, and it is critical to include these in the documentation. For example, just reporting an average without a range or a probability distribution for a quantity of interest gives the illusion of too much precision and objectivity. Expert judgments are sometimes inappropriately used to avoid gathering additional management or scientific information. These judgments should complement information that should be gathered, not substitute for it. Sometimes decision makers with a predisposed desire to prove the HLW site is safe or to select a given design alternative seek experts whose views support or justify their position. This is clearly a misuse of expert judgments. However, it is worth noting that with formal expert judgments, it is easier to identify weaknesses in the reasoning behind a decision.

In conclusion, it is worthwhile to remark on circumstances that should be considered successes or failures resulting from expert assessments. Science and knowledge are constantly changing. Thus, it is natural that as the knowledge of an individual changes, his or her expert judgments will likely change. The representation of expert judgments as probabilities and utilities facilitates adjustments to account for new information. Even after the completion of a given assessment, an expert may recognize that he failed to account for some important information. The assessment process is designed to enhance the likelihood that such omissions are recognized. Then it is easy to update the overall expert judgment to account for the omission. The ability to change and the need to change expert assessments are not failures of the experts, the assessments, or the assessment process. Rather, they are natural and desired features to deal with the reality of science and knowledge for a complex problem such as an HLW repository.

After the explication of expert judgment, someone or some organization may wish to demonstrate that some of the assessments are not correct. For example, if some organization felt that the groundwater flow parameters near the repository site were incorrect, they might begin additional experimentation or search for additional information that would support their point. If this led to a process that eventually improved the overall state of knowledge, that would not be a failure of the assessment process. Rather, it would be one of the desired products of explicitly eliciting expert judgments. Because the overall intent of the expert judgment assessments and of performance assessment is a safe and legally operated repository.

The formal use of expert judgment in the performance assessment of an HLW repository contributes to understanding, learning, communicating, and decision making. In the final appraisal, the significance of the explicit use of expert judgment should be evaluated by the overall value it adds to the performance assessment. Naturally, this is the same criterion applied to any of the inputs for or aspects of a performance assessment.

4. SUGGESTIONS FOR THE USE OF EXPERT JUDGMENT IN HLW DISPOSAL

This chapter specifies how to apply the techniques for eliciting and using expert judgment discussed in Chapter 3 to the five problem areas of HLW disposal outlined in Chapter 2: scenario development and screening, model development, parameter estimation, information gathering, and strategic repository decisions. Some of the techniques apply to each of the five areas, and others are relevant only to single areas. For each of the five areas, experts must be selected and trained for the elicitation process, an appropriate elicitation process must be designed, and results must be thoroughly documented and presented.

For scenario development and screening, identification and screening techniques are directly applicable to produce the set of scenarios for which probabilities are then assessed.

For model development, the identification and screening techniques are initially most relevant to select the variables to use in the conceptual models. Techniques for quantifying values may also be relevant to evaluate alternative models. Then mathematical models are developed to quantify the conceptual models. In this process, information gathering techniques are utilized as well as parameter estimation, both of which are addressed in the descriptions of the two problem areas that follow.

The main techniques in parameter estimation are screening to select the key parameters and quantification of the uncertainties in the form of probability distributions for those parameters.

Information gathering provides better information for the other areas of scenario and model development and parameter estimation. Information gathering uses techniques for identifying and screening information-gathering strategies and for quantifying probabilities and values.

Strategic repository decision making can use all the techniques described in Section 3. First there is the task of generating alternatives for the construction and operation of the repository, which can use identification and screening techniques. Decision and event trees are next used to decompose the alternatives and events in a logical sequence. Objectives hierarchies are used to decompose the objectives that are relevant to evaluate the outcomes of decision and event sequences. Probability quantification techniques are used to assign probabilities to events in the decision tree, and utility quantification techniques are used to assign utilities to outcomes. Then decision analysis can be used to develop insights for decision making.

4.1 Scenario Development and Screening

SNLA's methodology for development and screening of scenarios that hypothesize the possible future states of the disposal system was described in Section 2.1. The methodology consists of the following: (1) identification and classification of events and processes, (2) screening of events and processes, (3) formulation of scenarios, and (4) screening of scenarios. In addition, we discussed earlier the need to estimate the likelihood of occurrence of each scenario to demonstrate compliance with the containment requirement in the EPA Standard (40 CFR Part 191.13). Below we present guidelines for the applying techniques described in Chapter 3 to each of these areas.

4.1.1 Identification and Classification of Events and Processes

The main objective of these tasks is to arrive at a comprehensive list of events and processes from which the scenarios are formulated. A secondary objective is to classify the events and processes to increase the likelihood that the list is indeed comprehensive. This classification should also be useful for organizational purposes.

The group of experts that prepares the list of events and processes needs to be interdisciplinary. The experts should be specialists that have substantive knowledge in at least the following disciplines: general geology, seismicity, volcanology, tectonics, resource exploration, climatology, hydrology, and mining and/or rock mechanics. In addition, since future human behavior (e.g., human intrusion) can strongly influence, and indeed create, future scenarios, the experts should also include historians, sociologists, and psychologists knowledgeable about issues of technological change. It should be noted that these specialists should not be required to have in-depth knowledge of nuclear waste disposal issues; the specialists should be complemented by generalists (i.e., experts with general knowledge in performance assessment). Generalists show the specialists how their judgments contribute to the performance assessment.

The experts should be sensitized to biases, primarily availability (Section 3.2.3). The bias of availability in this context refers to a possible tendency of the experts to rely too heavily on existing records that do not necessarily represent the future adequately. The experts may not allow for adjustments to the existing information and may need some training from the generalists on performance assessment and how their judgments will be used.

The particular elicitation techniques applied in the identification and classification of events and processes were described primarily in Section 3.3.1: forward and backward induction, value-driven identification, and analogy/antimony-driven identification. We believe that more than one elicitation technique should be used to enhance the likelihood that the sets of events and processes are comprehensive.

The approach should be documented so that interested individuals may clearly discern the rationale of the elicitation process and the results. Intermediate lists as well as the final list of events and processes should be presented and should also include the steps to go from one list to another if multiple lists preceded the final one. An additional advantage of distributing the sets of events and processes is that any omitted examples may be identified and then, of course, added to the list.

4.1.2 Screening of Events and Processes

The basic problem is to screen out insignificant events and processes from the list generated in the previous step. While the list of events and processes should be generated generically as well as specifically for each site, the screening out of events and processes by necessity must be site specific. To screen out events and processes, screening criteria must first be formulated and applied to arrive at a "final" list of

DRAFT

events and processes to be used in formulating scenarios. The importance of both steps cannot be overemphasized. If the screening criteria are developed poorly, then the likelihood increases of eliminating potentially significant events and processes and/or of including insignificant ones. If the criteria themselves are not applied correctly, the same consequences are possible. In either case, the purpose of screening is defeated.

The specialists selected for identifying events and processes can also be used for identifying screening criteria. They should be trained specifically to overcome biases such as "overconfidence" and "availability" (Section 3.2.3).

The elicitation techniques for screening events and processes are discussed in Section 3.3.2. The first part of the elicitation exercise should concentrate on developing the screening criteria based on physical reasonableness, potential consequences, and likelihood of occurrence. The second aspect of the elicitation exercise should focus on setting reasonable constraints for the screening criteria. For example, in dealing with the likelihood of occurrence of a given initiating event or process, what probability of occurrence is too low? The last part of the exercise should be the application of the screening criteria. Multiattribute utility analysis (Section 3.3.5) is an approach for explicitly making tradeoffs between the different criteria. It is important to point out that iterating through the target levels and constraints in the criteria is recommended as a mechanism for determining the impact that these may have on the final list of events and processes.

The documentation and presentation of results mainly explains clearly the logic of the approach used in sufficiently general terms that it can be followed and critically reviewed by a wide range of interested parties. The documentation should allow not only critique of the approach, but of the results as well. The result should be a final list of events and processes that will be combined to form scenarios.

4.1.3 Generation of Scenarios

Once unimportant events and processes have been eliminated from further consideration, the surviving ones are combined to form scenarios. This step can be conducted by generalists knowledgeable about the application of event trees. The forward and backward induction techniques described in Section 3.3.1 and techniques for combining may be useful.

4.1.4 Screening of Scenarios

The guidelines for using expert judgment in this step are identical to those described in Section 4.1.2 for the screening of events and processes. The problem is to reduce the number of scenarios for the performance assessment to a tractable and representative set. This is accomplished by aggregating scenarios and by developing and applying screening criteria as in the screening of events and processes. The screening criteria should again stress physical reasonableness, potential consequences, and likelihood of occurrence. The selection and training of experts, the elicitation techniques, and the documentation and presentation of results should be identical to that in Section 4.1.2.

DRAFT

4.1.5 Probability of Scenarios

The problem to be addressed by the experts in this step is twofold: estimating the probability of the individual events and processes comprising a scenario, and combining these probabilities to arrive at the probability of the scenario. To estimate the probability of the individual events and processes, the experts need to identify the initiating event or process and decide whether the occurrence of the other events and processes in the scenario are conditional on the occurrence of the initiating one.

This step requires a multidisciplinary team of specialists with substantive knowledge in general geology, seismicity, tectonics, volcanology, climatology, hydrology, rock mechanics and mining, etc. Generalists with knowledge of performance assessment can provide insights on what type of scenarios are likely to be more significant. Finally, normative experts with experience in probability elicitation are needed to train the other groups of experts as well as to serve as the elicitators.

The specialists should be trained in overcoming probability biases (mainly overconfidence, anchoring, and availability), decomposing, expressing judgments explicitly, probability encoding, and assessing conditional probabilities. The specific elicitation techniques applicable to this step are the probability quantification techniques described in Section 3.3.4. The techniques for estimating the probability of discrete events such as the direct probability technique or the direct odds technique may be particularly useful.

4.2 Model Development

The development of models for performance assessment includes the development of conceptual models, mathematical models, and associated computer codes. This effort involves the selection and interpretation of available data and other sources of information, the formulation of relevant assumptions, and confidence building in the models and codes developed. Each requires expert judgment.

4.2.1 Data Selection and Interpretation

This task mainly provides the basis for the formulation of conceptual model(s) of the disposal system. Experts select and interpret data and other information that will lead to the establishment of the system's geometry; boundary and initial conditions; and past, present, and future events and processes that may impact the behavior of the system (Section 2.2.1).

It is expected that specialists, generalists, and normative experts will be required to carry out this task. Specialists primarily should concentrate in the fields of geology and hydrology; however, some specialists involved in the identification and classification of events and processes in the scenario development (Section 4.1.1) should also be used here. Generalists who have participated in earlier or preliminary performance assessments of HLW disposal sites should be used in this task. Generalists should be able to provide insights regarding the relative importance of different types of data and information based on their past experiences. Normative experts should assist the specialists in searching and cataloging different sources of information.

The elicitation exercise is likely to be in three phases. In the first phase, the specialists and generalists identify both site-specific and generic sources of data and other information. For this phase, the experts should be trained to overcome "availability" bias (Section 3.2.3). The specific elicitation techniques relevant to the identification task are presented in Section 3.3.1.

In the second phase of the elicitation, the experts must screen out unimportant sources of information and select the most relevant ones. To achieve this goal, criteria must be developed to accomplish the screening step, and then these criteria need to be applied to arrive at the most relatively important sources of data and information. This phase of the elicitation is similar to that discussed in Sections 4.1.2 and 4.1.4 (Screening of Events and Processes, and Screening of Scenarios). The training and elicitation techniques are similar to those suggested in Section 4.1.2 and are presented in Sections 3.2.3 and 3.3.2.

The third phase involves the interpretation of the selected information. In this phase, the experts make inferences based on this information that will form the basis for the development of models. The experts should be trained to overcome biases associated with availability, ignoring base rates, and nonregressive predictions (Section 3.2.3). Availability refers here to the tendency to follow a conventional line of reasoning when interpreting the available information without considering evidence that may challenge this convention. Ignoring base rates as applied to data interpretation refers to ignoring soft or abstract information while focusing only on concrete evidence and data. Nonregressive prediction is the tendency to make inferences using relationships the applicability and validity of which have not been established for the system in question.

4.2.2 Development of Conceptual Models

Constructing conceptual models uses inferences based on the selection and interpretation of data to formulate assumptions for the behavior of the disposal system. These assumptions, in turn, are the cornerstone for the assembly of mathematical models and their computer codes used in the quantitative analyses. Modeling most likely will result in a multitude of alternative conceptual models because of the lack of data during the early stages of a site investigation. As more information becomes available, it could be possible to distinguish among the different conceptual models and possibly reduce their number. Finally, it would be feasible, if a number of conceptual models survive screening, to quantify a relative likelihood for each conceptual model that it adequately describes the "true" groundwater flow and transport processes, for instance.

Again, specialists, generalists, and normative experts will probably be needed. The specialists should be in the area of hydrology and should include both modelers and experimentalists as will be discussed below. Generalists should be used to assure that the specialists render judgments within the context of performance assessment. Normative experts should be used to assist the specialists in making value judgments. Some of the experts used in data selection and interpretation should be involved in this task to provide continuity. Multiple teams of experts may be appropriate.

The first phase of the elicitation is the development of meaningful criteria for the formulation of assumptions and the construction of conceptual models. These criteria include beliefs regarding the importance of model attributes such as geometry, the ability to simulate specific events and processes, groundwater flow regime, relevant parameters, complexity, etc. The selection of these criteria is likely to be based on value judgments and will require all three types of experts. While the specialists should be expected to play the biggest role in this phase, generalists should provide the basis for acceptable tradeoffs that can be made in light of regulations that need to be addressed in the performance assessment. Normative experts are likely to be elicitators. Techniques for expressing value judgments are described in Section 3.3.5.

The second phase is to develop a procedure for distinguishing among the alternative conceptual models and, if possible, screening some out. This should be accomplished by attempting to identify the salient features of each conceptual model, formulating and conducting specific analyses and experiments that could test the validity and/or importance of these features, and setting screening criteria and applying them. In this phase, both specialists in model development and experimental studies are needed because a synthesis of analyses and experiments will likely be necessary. Screening techniques described in Section 3.3.2 should be useful in this phase.

The third phase consists of an attempt to quantify the likelihood that each conceptual model that survives screening is the best of the available models. Specialists and normative experts will be needed in this phase, and probability elicitation tools such as sequential conditional probability assessment and others presented in Section 3.3.4 are applicable to this phase. Appropriate training to overcome such biases as overconfidence, anchoring, availability, and ignoring base rates, discussed in Section 3.2.3, should be conducted before the elicitation.

A portfolio of conceptual models should be chosen that, at the very least, represents extreme sets of conditions for a performance assessment and that, at the same time, can be tested during site-characterization investigations. Situations in which two or more conceptual models are very similar should be avoided. Refinement of the final portfolio of conceptual models can be done using decision analysis and, in particular, preposterior analysis (Winkler, 1972). These techniques increase the likelihood that the set of conceptual models selected is adequate for conducting a performance assessment, the results of which will allow making regulatory decisions with confidence.

4.2.3 Confidence Building

Following the development of conceptual models, mathematical models will be formulated that cast the models in terms of mathematical equations (i.e., algebraic, partial, and/or integral equations. In setting up these equations, assumptions are made, the validity of which needs to be established. Typically, because of the complexity of the equations in even the simplest models to simulate the behavior of an HLW disposal system, the solution to these equations is implemented in computer codes. Depending on the nature of the equations (linear vs. nonlinear, partial vs. algebraic, etc.) and the coupling between two or more equations, these can be solved either analytically or numerically. In any case, the implementation of neither analytical solutions nor numerical solutions is exact. For example, if an analytical solution involves an infinite series, this series needs to be truncated after a finite number of terms, or if it includes a complex integral, this integral is often evaluated numerically. Numerical solutions inherently are approximations to the "true" solution of the equation(s). In whatever form (either analytical or numerical), errors are introduced when solving the equation(s) in a mathematical model. Since the validity of these mathematical models and computer codes cannot be established over the temporal and spatial scales of interest in HLW disposal (Section 2.2.3), validation cannot be achieved in the truest sense. Nevertheless, confidence must be built to the extent that, given the present state of the art, these models and codes are deemed adequate for the job at hand: predicting the behavior of the disposal system over several kilometers and tens of thousands of years. To build confidence in the models and codes, limited-scope activities will be carried out, and expert judgment will play a major role in designing and conducting these activities, as well as in interpreting the results.

Experts are likely to be used in selecting important features in the models to be tested and the type of testing. For example, there may not be a need to test the expression for radioactive decay in the radionuclide transport equation because this is a well-established and accepted expression. On the other hand, the use of a Fickian model for diffusion to represent dispersion or the use of a linear-sorption-equilibrium based retardation factor are both models that are the subject of much criticism and should be tested. The question then becomes what tests to conduct, for example, laboratory vs. field tests. Experts will also be involved in the selection of appropriate criteria to establish the measures of goodness of the models. These are competing measures, and experts should select those criteria that are most meaningful to the regulatory requirements to be addressed. The experts must also set the limits and constraints in these criteria. Experts will also be needed to assess the ability of the models to extrapolate from the temporal and spatial scales at which they were tested to the scales of interest in HLW disposal. Finally, there are likely to be some couplings in the models that are so complex it is impractical to test their validity. In this case, expert judgment assesses the adequacy of the modeling of these couplings.

The experts required include primarily specialists and generalists; however, it may be appropriate to include normative experts, but this may not be necessary. It is suggested that multiple teams of experts be used, each team consisting of both specialists and generalists, and modelers and experimentalists.

The experts make value judgments (tradeoffs) regarding what aspects of models need to be tested, and the techniques in Section 3.3.5 should be useful. In addition, they develop criteria for establishing the validity of given models. Therefore, the techniques for setting criteria, limits, and constraints to the criteria, and the applications of the criteria in Section 3.3.2 should be employed. As the "ultimate" validation test at an HLW disposal site cannot be performed and because of the complexity of the model, perhaps one of the biggest tasks to be faced by the experts requires the decomposition (Section 3.3.3) of the overall system model into meaningful pieces. While it has been recognized that there are likely to be couplings that cannot be tested, extreme care must be taken to assure that the decomposition of the problem does not eliminate significant couplings. For example, in testing for the validity of the linear-sorption-equilibrium model as the dominant radionuclide retardation, the problem should not decompose such that a test is conducted that does not include flow-field effects because evidence exists that they have a significant impact on sorption.

4.3 Parameter Estimation

4.3.1 Identification of Parameters

As stated in Section 2.3.1, parameters are embedded in conceptual models that predict the performance of the repository in terms of radionuclide emissions and their potential health effects. Therefore, the importance of parameters is closely related to the variation in the amount of radionuclide emissions relative to variations in the parameters. The main method for identifying and selecting parameters is sensitivity analysis. In such analysis, parameters of conceptual models are systematically varied (both individually and in sets) to determine which parameter or combination of parameters has the strongest impact on radionuclide emissions.

Sensitivity analysis is currently more a craft than a science. It is therefore especially important that the expert judgments that select and interpret the sensitivity analysis for parameter identification are made explicitly.

4.3.1.1 Guidelines for Parameter Identification

At this stage of the analysis of the HLW disposal problem, the issues for parameter identification are typically fairly clear cut: Given a chosen conceptual model, what are its parameters that should be quantified for further analysis. There may be two complications with this problem definition that may require resolution before identifying important parameters. First, there may be several conceptual models, and second, there may be different ways to categorize parameters. If these complications occur, it is useful to convene an expert panel to address these issues before the actual parameter identification process. Guidelines for issue identification and selection of experts for this part of the study should be followed (Sections 3.2.1 and 3.2.2). In particular, a diverse set of experts and examination of a diverse set of conceptual models and sets and subsets of parameters should be considered.

Once a conceptual model and the possible parameters and their subsets are agreed upon, identifying "important" parameters is more technical and better defined.

Three types of experts are necessary identifying important parameters: Substantive experts with knowledge of geology and hydrology, among others; generalists with expertise in the conceptual models; and experts in sensitivity analysis. An effort should be made to obtain the best expertise in these areas, as well as to maintain some diversity of opinion. This diversity is especially important for the experts concerning the conceptual model, as they are likely to disagree *a priori* about what constitutes important parameters of the model. Less emphasis on diversity is needed in selecting experts in hydrology and geology, and even less in selecting experts in sensitivity analyses.

Training in elicitation techniques is not required in this area. However, both the substantive experts and the sensitivity analysts need to learn about the nature of the conceptual model, its assumptions, its behavior and some of its preconceptions about sensitivities. For the substantive experts, this may provide guidance for reformulating parameters (e.g., by dividing hydraulic conductivity into separate strata). This type of training alerts sensitivity analysts to possible interactions among parameters, as well as to possible problems and opportunities in carrying out sensitivity analyses. This training should consist of two parts: presentation and familiarization with the conceptual models and some of their predictions and extensive question-and-answer periods regarding the use of the conceptual models.

Because sensitivity analysis plays a key role in identifying important parameters and because the elicitation centers around a conceptual model, the elicitation session should be structured somewhat differently from the standard session described in Section 3.2.4. In particular, display and discussion of sensitivity analysis results of running parts of the complete conceptual model should be emphasized. Comparatively less time should be spent in individual elicitations, and the amount of actual numerical elicitation should be fairly small at this stage.

There are two suggestions for structuring an elicitation session in this context, depending on whether sensitivity analyses can be done on-line. If they can be done on-line, it is highly desirable to structure the elicitations as an interactive exercise in which the experts formulate hypotheses about sensitivity and importance and test them in real time. Some structure should be provided to make sure that the more prominent hypotheses are tested and that all parameters are examined. Beyond that, the experts should be able to develop their own plan for carrying out sensitivity analyses and judging their outcomes.

If sensitivity analysis cannot be done on-line, the experts should convene at least twice. The first meeting determines which sensitivity analyses should be carried out. The second meeting discusses the results of the sensitivity analyses and makes judgments about which parameters are important enough for further quantification of uncertainties. If certain parts of sensitivity analyses can be done on-line, this should be done to liven up the exercise. However, care should be taken that the on-line sensitivity analyses do not gain more prominence by making the respective parameters more available to the experts (Section 3.2.3).

In both cases (on-line vs. prepared sensitivity analyses) the experts should aim at making three judgments about the parameter:

1. Sensitivity related to selected performance measures;

2. Overall importance;

3. Need for further quantification or data collection.

4.3.2 Quantification of Parameters

A fairly large amount of research and applied work exists for quantifying expert judgments about uncertainties in parameters with probability distributions. The recommendations that follow are therefore grounded in significant amounts of experience (Section 3.3.4).

4.3.2.1 Guidelines for Quantifying Parameters

After the conceptual model and its important parameters are identified, the issue is to quantify the knowledge of substantive experts in hydrology and geology about the parameters as probability distributions. Price to any assessments, is useful to identify current or near-future data collection efforts, to put the actual expert elicitation of uncertainties before this data collection into perspective. In addition, it is very important that the parameters be unambiguously defined.

Parameter quantification addresses specific issues such as the estimation of hydraulic conductivity parameters in specific strata of the repository. Experts should be selected on a parameter-by-parameter basis. Depth of knowledge is crucial, breadth and diversity are secondary in this case. Motivational biases should be considered. For example, a hydrologist on record as stating that Yucca Mountain is an absolutely safe site for the repository might give estimates of hydraulic conductivity that are too low. It is useful to counterbalance such potential biases through expert selection.

Training should focus on constructing (usually continuous) probability density functions (pdfs) or cumulative density functions (cdfs) over parameters. The main recommendations in Section 3.2.3 apply with full force here. In particular, experts should be familiar with the probability elicitations task, and they should get ample practice using many examples of the types of elicitation that they are likely to face. Anchoring and adjustments, overconfidence, and motivational biases should be demonstrated, and debiasing procedures should be explained.

All experts must agree on the precise definition of the parameter to be elicited. For example, when hydraulic conductivity is discussed, it must be absolutely clear which strata of the repository is referred to, whether one wants to assess mean or maximum hydraulic conductivity, what maximum may mean, etc. It is useful to structure the elicitation session to involve a "generalist" knowledgeable about the conceptual model and the interpretation of the parameter within that model.

A variety of decomposition techniques may be useful, depending on the specific parameter or the expert (see Section 3.3.3). If functional decompositions are utilized, direct probability assessments should be used as consistency checks for probabilities calculated based on decomposed assessments. For example, when assessing hydraulic conductivity in four different strata and subsequently assessing average hydraulic conductivity, the results can be checked for consistency with the average hydraulic conductivity.

Parameters should usually be represented as continuous random variables. Therefore, our suggestions for applying elicitation techniques are very straightforward: use the fractile technique described in Section 3.3.4 and check it with the interval technique and perhaps a few gamble questions. Pay particular attention to the extremes and probe them carefully, possibly by considering physical impossibilities and extreme gambles. For example, when considering hydraulic conductivity, the elicitator may ask for the expected minimum and maximum areas of conductivity in the repository, for the minimum and maximum in comparable formations, and for the minimum and maximum in a variety of substances and materials. An appropriate range should then be selected. By broadening the notion of minima and maxima, the expert may be induced to consider the full range of possibilities for the case at hand as well.

Having obtained a first-cut range, the normative expert should ask the specialist to explain a set of hypothetical data that indicates events outside the range. One may also ask the expert, whether he or she would be willing to bet a large sum of money that all possible experiments would lead to the conclusion that the parameter is in the range stated. Both techniques are useful for debiasing.

4.4 Information Gathering

To better design, construct, and operate a nuclear repository, numerous important decisions must be made, many of which will affect repository performance. To improve the quality of these decisions and to improve performance assessment, numerous efforts must be carried out to gather information. Collectively, this information will be very costly. In terms of dollars, the cost will be in the billions; in terms of human resources, the cost will be in the thousands of person-years of professional time; and in terms of the environmental and social disruption of the testing to information gather, there will be significant effects. Thus, decisions about information gathering should be made carefully and thoughtfully. Information gathering cuts across the three areas discussed earlier: scenario development, model development, and parameter estimation. In all of these cases, the information is intended to improve the quality of the scenarios, the usefulness of the models, and the estimates of the parameters.

Information gathering is also different from the first three areas in that it concerns decisions. Some important decisions concerns how many, how deep, and where to drill test holes into the repository media. Another class of decisions concerns what computer codes should be developed and what conceptual models should be fleshed out into analytical models. For example, should the groundwater flow models be in two dimensions or in three dimensions, and what variables should they include? Regarding parameters in these models, how can we best estimate a variable such as porosity to reasonably balance the insight gained about the variable against the cost and effort necessary to gain that insight?

The use of the concept of expected value of sample information (Raiffa and Schlaifer, 1961) allows appraisal of the various alternatives for gathering information and selection of the one that is best given expectations about what information might be obtained from the various alternatives and about the economic cost, time required, and damage caused by that alternative. Value judgments must balance the advantages and disadvantages of gathering the information and take into account the overall goal of creating a safe, legal repository.

In the rest of this section, three special classes of problems concerning information gathering are discussed. These problems concern informational drilling, development of models, and conducting laboratory or field experiments other than drilling.

4.4.1 Informational Drilling

The informational drilling program is one of the major activities in the characterization of the repository. It should be carried out only after careful appraisal of the alternatives. To do this, there are several distinct activities that should be completed that rely partially on the use of expert judgment.

To characterize the informational drilling problem, the objectives of the drilling program and reasonable alternatives first need to be identified. Then the alternatives should be screened to specify the competitive options. For each of these competitive options, estimates are necessary for the information that will possibly be learned and for the time, cost, and damage caused. Using value judgments to balance these and the concept of the expected value of sample information, an analysis can indicate the relative desirability of the options under a wide range of assumptions. Each of these tasks are elaborated below.

The first and driving task for the informational drilling program is to specify its objectives. It is important to be very explicit about the relative desirability of different information that might be learned from the program. In this step, specialists need to be selected to assist in specifying the objectives because the objectives of the drilling program will likely be technical. It is also important that the drilling objectives be logically related to the fundamental objective of better designing, constructing, and operating a safe and legal repository. This relationship may be best specified by generalists with a broader understanding of the repository program. The techniques for structuring objectives hierarchies are useful in this task (Section 3.3.3), and careful documentation and review of the objectives hierarchy is appropriate before completing the additional tasks below.

The second task is to identify a large number of reasonable alternatives for gathering information via drilling. To develop these alternatives, specialists and generalists should again be used. At this stage, the alternatives need not be carefully refined (e.g., the exact location of each hole), but they should be specific enough to distinguish them from other alternatives.

The next task is to screen the large number of alternatives to identify those that are competitive. The relationship of the objectives of the drilling program to the fundamental objectives of the repository should be a basis for this screening. The screening criteria should at first be specified by generalists using techniques discussed in Section 3.3.2 and then be used to eliminate many noncompetitive alternatives. At a later stage in the analysis of information drilling options, when the relative desirability of alternatives that passed the screening are known, the screening criteria should be reexamined to determine whether more related screening criteria might have yielded better alternatives. The way screening criteria can be verified with information that comes later in the analysis is outlined in Keeney (1980). If the appropriateness of the screening criteria is to be verified, the original use of expert judgment to set the criteria for screening is not so significant. Expert judgment is crucial not only to screening but in setting up the relationships of objectives of the drilling program and in specifying implications of what might be learned from the various drilling options in the next task.

The fourth task is to define better the competitive options that make it through the screening. There are two aspects to this definition. The first is to specify exactly what drilling will occur, and the other is to predict the possible information learned from the drilling and its time, cost, and resulting damage. This task relies heavily on expert judgment. Some of it will be from specialists, specifically information referring to details learned about the hydrology and geology at the site. Other information will necessarily come from generalists about the time and cost of the drilling options. For each of these circumstances, experts need to be carefully selected and trained. The assessments should indicate the implications of the alternatives in terms of a probability distribution function as discussed in Section 3.3.4. An important subtask in the estimation of the impact of information is assessing the conditional probability distributions with the information that can be obtained from the alternative drilling activities and assessing the probability distribution of the information from the drilling. In particular, the probability distribution for cumulative radionuclide releases and health effects will strongly depend on the information obtained. These probability distributions are a major ingredient for carrying out a value of information analysis.

The next task is to quantify the value judgments (Section 3.3.5) necessary to integrate all the objectives of the informational drilling program. Because of the uncertainties about what will be learned by the various drilling options, a multiattribute utility function should be used to integrate these objectives (Keeney and Raiffa, 1976). Expert judgment will be necessary to specify the value judgments for the utility function. These judgments are of a policy nature because they relate to the quality of information available for key decisions regarding the repository, and they should be provided by individuals with policy positions in the repository program and stakeholders with a legitimate voice in that program. Examples of this in the repository program are discussed in Section 1.4. To assist the policy makers in quantifying their judgments, it is important to have the assistance of a normative expert with substantial experience in quantifying value judgments.

With the tasks above completed, it remains to analyze the options and identify those that provide the most information for the time and effort. At this stage, it is critical to gain the insights about why the better options are better and about why they are that much better. This interpretation is the link that provides useful information to the decision-making process from the explicit use of expert judgment in the appraisal of informational drilling options.

4.4.2 Selecting Models to Develop

With any information-gathering problem, the key is to specify the objectives to be achieved. In this case, the objectives to be achieved by developing models need to be carefully specified. Furthermore, these objectives need to be related to the fundamental objectives of designing, constructing, and operating a repository. In this regard, the fundamental objectives for model development are the same as the fundamental objectives for informational drilling. What is different in this case is the means objectives by which those fundamental objectives are achieved. To specify the relationship between the means objectives and the fundamental objectives, expert judgments of both specialists and generalists are needed. Essentially, these relationships answer the questions about how model development will contribute to better understanding and better decision making regarding the repository.

After the experts are selected, they need to be trained to distinguish between fundamental and means objectives and to understand concepts such as influence diagrams and objectives hierarchies for relating them. Then the elicitation process needs to be carefully documented. This documentation can be reviewed by a large number of peers for completeness and reasonableness, and the revised results should provide a basis for the additional tasks in selecting appropriate models for development.

The next task is to select general types of alternative models that may be worthwhile to develop. Some of these may be analytical models, and others may be simulation models represented by codes. Other factors defining the alternatives concern the number of variables in the models and exactly which variables they should be. A combination of generalists and specialists should be appropriate for defining a large number of alternative models. Identification techniques for expert elicitation discussed in Section 3.3.1 will be used extensively in this task.

The next task is to screen the alternatives to focus on those that seem most useful to provide information for the repository. In this phase, the screening models outlined in Section 3.3.2 will be utilized. The criteria for screening should be set using a combination of judgments from specialists and generalists. The exact screening criteria are not too important as their appropriateness should be verified after the models have gone through various stages of development. In general, if the models selected for development are not providing the insights expected, either because of lack of available data or field data indicates that they are inappropriate, then the models can be revised or new models selected for development.

The fourth task is essentially model development as discussed in Section 4.2 Details are found in that section, so only a brief overview is included here. The task is essentially to specify the variables appropriate for each of the models selected for development and to identify data sources to provide information about those variables. Also, using any available physical relationships, it is necessary to relate the variables to each other to provide the structure for the model. At this stage, it is essentially the judgments of specialists that are important. Normative experts should assist these experts in expressing their judgments about the relationships of the variables.

There are a number of input variables to a large model and one or more output variables of interest. Probability distributions quantify the current state of knowledge about the input variables and are used in the model to derive implications for the output variables. How this is carried out is described in Section 4.3. It relies heavily on the techniques for quantifying probability judgments discussed in Section 3.3.4.

The last task is to run the models many times and gain the insights available from them. A team of generalists and specialists will likely be most appropriate to interpret the results of the analyses. Based on these insights, it will probably be appropriate to repeat various runs of the model to gain additional insights about the sensitivity of parameter values for different variables with respect to the model's implications. At this stage, the team of experts should also verify any assumptions made in selecting models to develop. These assumptions pertain to the number of variables, the relationships between variables, and their quantification.

4.4.3 Laboratory and Field Experiments

Many laboratory and field experiments, exclusive of informational drilling, will likely be done before final design and construction of the repository. The first task in each of these situations is to specify the objectives to be achieved by the experiment proposed. As with the problems discussed above, the task is to provide information that results in a legal and environmentally sound repository through better design and construction. This task requires balancing of the impacts of the experimentation in terms of cost and effort against the value of the information learned. For each of the proposed experiments, different objectives contribute to information obtained. The kind of information expected needs to be specified using expert judgments of generalists and specialists and the assistance of a normative expert to explicate that judgment. Once these objectives are clarified, we have a basis for evaluating different alternatives for the laboratory and field experiments.

For any proposed experiment, the next task is to identify alternatives for conducting that experiment. These may vary in cost, time, or depth or breadth. They also may vary in the sophistication of testing equipment used. At this stage, the judgment of generalists with some assistance of specialists should be appropriate for characterizing the alternatives.

The next task is to screen the various alternatives to identify the types that seem more appropriate. The screening criteria should be set by the generalists using concepts described in Section 3.3.2, since the information is relevant to the overall repository program. However, at later stages in the analysis, the appropriateness of the screening criterion should be validated. If it turns out the information sought from the experiments is not being provided, the analysis should be repeated to determine which experiments should be conducted and whether they are worth the information. Experiments that at one time were thought not to be appropriate because of the expectation that certain information would become available have become appropriate when it is known that that information is not available. In simpler terms, if some field experiments are not successful, the relative desirability of others may increase.

For the alternatives that have made it through the screening, one should more carefully specify details of the experiment to be conducted. As part of this, there should be probabilistic estimates of the amount of information obtained by each of the experiments as well as estimates of their cost, time, and any damage from the experimentation. As in the task of informational drilling, two sets of quantitative estimates are especially important: the conditional probability distribution over radionuclide emission for different experimental outcomes and the probability distribution over those outcomes. The judgment of generalists will likely be necessary for some of the cost and time information, although this judgment might be augmented by some specialists, whereas the judgment of specialists will mainly be used to judge the information expected from each experiment.

The objectives in the first task above need to be integrated into an overall utility function. These value judgments should be in accordance with the techniques discussed in Section 3.3.5 and should use the judgments of generalists on the repository team. However, these value judgments should be carefully related to the policy value judgements made about the fundamental value tradeoffs of the information gathering process. In other words, since the objectives of the experiments are means to achieve the objectives of designing and constructing a repository, the specific value judgments dealing with tradeoffs among the objectives of experiments must relate to the value tradeoffs that concern the policy objectives. This relationship should be carefully documented.

The final task is to analyze the various laboratory and field experiments using the value-of-information techniques and to select those that seem appropriate. In all cases, one of the alternatives that definitely should be considered is not conducting the experiment. In some sense, one of the more useful pieces of information gathered from such an analysis is whether specific experiments, given their quality, cost, and time, are worth the effort. In some cases, it may be cheaper simply to design the repository assuming that a certain situation exists, rather than verifying it. In other situations, although the information desired might be very important, if the experiments are unlikely to provide that information, they simply might not be worth the time, effort, and cost.

4.5 Strategic Repository Decisions

Strategic repository designs are those that directly concern the design, construction, and operation of the repository. As pointed out in Section 2.5, many of these decisions will affect the performance of a repository and therefore should be considered when developing and screening scenarios, developing model, estimating parameters, and gathering information. In a sense, any performance assessment is conditional on these strategic decisions.

For discussion it is useful to think of the analysis of those strategic decisions in terms of six components. The first two components, which identify the strategic problem, are specification of the objectives and identification of the alternatives. The degree to which the objectives are achieved by the various alternatives is quantified in the third component. The fourth component integrates the different objectives using value judgments concerning risk attitudes and the relative importance of different objectives. All the information is integrated and analyzed in component five to provide insight for decision making. Component six is documentation of the process and results.

The main techniques in these components are described in Section 3.3.3 (structuring objectives), Section 3.3.4 (probability quantification), Section 3.3.5 (value quantification), and decision analysis (Raiffa, 1968; Howard, 1968; Keeney and Raiffa, 1976; and von Winterfeldt and Edwards, 1986).

4.5.1 Specifying and Structuring Objectives

The overall objectives for constructing and operating the repository should guide the development of specific objectives for constructing or operating the repository. The techniques for constructing objectives hierarchies are useful for this step (Section 3.3.3). A group of experts representing all interested parties should be selected to specify the overall objectives for constructing and operating the repository. At this stage, it is important to have a broad diversity of opinions providing objectives for the repository, as these objectives should provide the foundation for future strategic decisions (Keeney, 1988a,b). The training for these experts need not be extensive but it should clearly indicate how the stated objectives will be used and methods that may facilitate broad thinking about their objectives. The elicitation process itself needs to be done by normative experts trained to elicit objectives in an operational manner for further analysis. The objectives should then be structured by the normative analysts, with the assistance of project members, and then carefully reviewed by peers and others interested in the repository program. Modifications are welcomed, as the intent is develop an appropriate fundamental set of objectives for the repository. Finally, these objectives should be documented.

With a given specific strategic decision, the repository objectives need to be related to specific objectives influenced by the strategic decision. That linking can likely be done by generalists with the assistance of normative experts. In essence, it is a deductive process that relates the overall objectives to a given decision problem. As always, the resulting objectives should be carefully documented after review by peers and others interested in the repository program, including all members who initially contributed to the overall objectives.

4.5.2 Identification of Alternatives

For any specific strategic decision, the alternatives need to be identified. Thus, the identification techniques Section 3.3.1 are relevant. The experts involved in specifying alternatives should have substantial knowledge about details of the specific decision to be addressed. Normative experts should assist them in defining generic alternatives (e.g., sets of alternatives that differ in terms of parameters). After a wide range of alternatives has been identified, it may be worthwhile to screen the alternatives using the screening techniques in Section 3.3.2. Appropriate screening criteria should be set by generalists to facilitate focusing on alternatives that are presumed to be better. After the analysis, the reasonableness of the screening criteria should be reexamined considering the quality of the screened alternatives. If it is likely that alternatives screened out would in fact be better than some of those retained, the analysis should be revised and repeated.

4.5.3 Impacts of Alternatives

Once the objectives and alternatives in a specific strategic decision problem are articulated, they effectively define a matrix in which objectives relate to the individual columns of the matrix and alternatives to the individual rows. To specify the impacts of the alternatives, one wants to fill in each cell in the matrix, indicating the degree to which the alternative impacts the corresponding objective. This process utilizes scientific and engineering knowledge and necessarily relies on models, data, and expert judgments. For this step, the techniques and procedures outlined for scenario development and screening, model development, and parameter estimation are repeatedly used. Since these are detailed in Sections 4.1 through 4.3, there is no need to elaborate on them here. It is simply worth noting that expertise from a variety of fields that includes the behavioral sciences, economics, and medical sciences will likely be required. Most impacts will be uncertain. In those cases, the techniques for probability quantification (Section 3.3.4) will be useful.

4.5.4 Value Judgments

At this stage, it is critical to aggregate the various component impacts for each of the alternatives. Because of the uncertainties regarding those impacts, some of these value judgments must address risk attitudes concerned with those uncertainties, and because there are multiple objectives, some of these value judgments concern critical value tradeoffs among objectives addressing environmental, social, economic, and health and safety impacts. The value judgments should be made as follows.

First, the original group who specified the overall objectives to the repository should specify quantitative value judgments regarding risk attitudes and value tradeoffs among those objectives using the value quantification techniques described in Section 3.3.5. Each of the individuals in that group should provide individual value judgments, and each of these sets of values should be carefully appraised for consistency. Also, individuals should be allowed to hear the logic of other people's points of view regarding the values and reiterate their judgments. However, it would be unlikely that everybody would have precisely the same values, so it would be unreasonable to force a consensus (Section 3.4). Each individual value should be carefully documented, and collectively they should provide a range for the values used in the problem.

4.5.5 Analysis of the Alternatives

The analysis of alternatives should integrate all the information from the preceding four components for the given strategic decision using decision analysis. Operationally, it may be reasonable to take an "average" set of the value judgments as a base case and do sensitivity analysis from this to incorporate all the different viewpoints. The intent is to identify alternatives that clearly are not competitors and identify circumstances under which each of the remaining alternatives are the best and how much better they are than the alternatives. Because of the uncertainty about quantitative parameters relating to the impacts, sensitivity analysis of some of these may also be appropriate. The experts working on this part of the problem should be analysts. It is unlikely that their use of expert judgments needs to be made explicit, but they certainly use expert judgment in deciding what sensitivity analyses to pursue. The degree of sensitivity analysis should be guided by the insights provided and the need for careful documentation.

4.5.6 Documentation of Analysis

The documentation of the analysis and its insights for decision making is essentially a collection of the documentation of each of the components of the analysis. However, it is worth recognizing that documenting the overall decision process does have some requirements different from documenting the components. This comes about because the overall process is of interest to different types of individuals, some of whom may not be concerned about details. Documentation of technical information relevant to impacts is likely of concern mainly to peers and individuals with a technical knowledge about those aspects of the repository. Documentation of the decisions made may be of concern to a large number of lay people as well as to numerous individuals concerned with or entangled by the politics of the repository problem. Documentation of the overall decision need not focus on detailed aspects of the problem that turn out not to be crucial. The documentation should very carefully explain what the alternatives are, what the objectives are for evaluating the alternatives, and the logic of why a given alternative was chosen. References can naturally be made to more detailed documentation elsewhere.

Documentation of any strategic decision should be considered itself a decision problem. One should carefully think of the objectives of the documentation and who the documentation is meant to inform because the communication alternatives have pros and cons. These need to be balanced appropriately in documenting the overall decision. The analysis of the documentation decision need not be made explicitly, but consideration of the appropriate components will likely result in better documentation.

DRAFT

REFERENCES

Armstrong, J. S., W. B. Denniston, and M. M. Gordon, "The Use of the Decomposition Principle in Making Judgments," Organizational Behavior and Human Performance 14, 257-63 (1975).

Arrow, K., Social Choice and Individual Values, (New York: John Wiley & Sons, 1951).

Ashton, A. A., and R. H. Ashton, "Aggregating Subjective Forecasts: Some Empirical Results," Management Science, 31, 1499-1508 (1985).

Barclay, S., et al., Handbook for Decision Analysis, Defense Advanced Research Projects Agency (1977).

Bonano, E. J., and R. M. Cranwell, "Treatment of Uncertainties in the Performance Assessment of Geologic High-Level Radioactive Waste Repositories," J. Mathematical Geology, 20, 543-565 (1988).

Bonano, E. J., P. A. Davis, L. R. Shipers, K. F. Brinster, W. E. Beyeler, C. D. Updegraff, E. R. Shepherd, L. M. Tilton, and K. K. Wahi, Demonstration of a Performance Assessment Methodology for High-Level Radioactive Waste Disposal in Basalt Formations, Sandia National Laboratories, Albuquerque, NM, SAND 86-2325, NUREG/CR-4759 (1989).

Brown, R. V., A. S. Kahr, and E. Peterson, Decision Analysis for the Manager (New York: Holt, Rinehart, and Winston, 1974).

Bunn, D. W., Applied Decision Analysis (New York: McGraw Hill, 1984).

Clemen, R. T., and R. L. Winkler, "Limits for the Precision and Value of Information from Dependent Sources," Operations Research, 33, 427-442 (1985).

Clemen, R. T., "Combining Overlapping Information," Management Science 13, 373-380 (1987).

Cranwell, R. M., R. V. Guzowski, J. E. Campbell, and N. R. Ortiz, Risk Methodology for Geologic Disposal of Radioactive Waste: Scenario Selection Procedure, Sandia National Laboratories, Albuquerque, NM, SAND 80-1429, NUREG/CR-1667, draft report, copy available in NRC Public Document Room.

Cranwell, R. M., J. E. Campbell, J. C. Helton, R. L. Iman, D. E. Longsine, N. R. Ortiz, G. E. Runkle, and M. J. Shortencarrier, Risk Methodology for Geologic Disposal of Radioactive Waste: Final Report, Sandia National Laboratories, Albuquerque, NM, SAND 81-2573, NUREG/CR-2452 (1987).

Dalkey, N., and O. Helmer, "An Experimental Application of the DELPHI Method to the Use of Experts," Management Science, 9, 458-467 (1963).

Dalkey, N. C., The Delphi Method: An Experimental Study of Group Opinion, The Rand Corporation, Santa Monica, CA, Report No. RM-5888-PR (1969).

de Finetti, B., "La Prevision: Ses Logique, Sources Subjectives," Annales de l'Institute Henri Poincare, 7, 1-68 (1937).

DeGroot, M. H., Optimal Statistical Decisions (New York: McGraw Hill, 1970).

Delbecq, A., A. Van de Ven, and D. Gustafson, Group Techniques for Program Planning (Glenview, IL: Scott Foresman, 1975).

Department of Energy, A Multiattribute Utility Analysis of Sites Nominated for Characterization for the First Radioactive-Waste Repository - A Decision Aiding Methodology, Office of Civilian Radioactive Waste Management, Washington, D.C., DOE/RQ-0074 (1986).

Department of Energy, Draft Environmental Assessment, Reference Repository Location, Hanford Site, Office of Civilian Radioactive Waste Management, Washington, D.C., DOE/RW-0017 (1984).

Dyer, J. S., and R. A. Sarin, "Measurable Multiattribute Value Functions," Operations Research, 22, 810-822 (1979).

Edwards, W., "How to Use Multiattribute Utility Measurement for Social Decision Making," IEEE Transactions of Systems, Man and Cybernetics, SMC-7, 326-340 (1977).

Edwards, W., and J. R. Newman, Multiattribute Evaluation, (Beverly Hills: Sage, 1982).

Edwards, W., T. Eppel, A. Hybel, D. Mefford, and D. von Winterfeldt, Design Principles for an Analyst's Work Station, Social Science Research Institute, University of Southern California, Los Angeles, Research Report No. 87-1 (1987).

Edwards, W., and D. von Winterfeldt, "Public Values in Risk Debates," Risk Analysis, 7, 141-154 (1987).

Einhorn, H., and R. Hogarth, "Unit Weighting Schemes for Decision Making," Organizational Behavior and Human Performance, 13, 171-192 (1975).

Einhorn, H. J., "Expert Measurement and Mechanical Combination," Organization Behavior and Human Performance, 7, 86-106 (1972).

Ekman, G., and L. Sjoberg, "Scaling," Annual Review of Psychology, 16, 451-474 (1965).

Electric Power Research Institute, Seismic Hazard Methodology for the Central and Eastern United States, Vol. 1, Methodology, NP-4/26 (1986).

DRAFT

EPA, 40 CFR Part 191, "Environmental Standard for the Management and Disposal of Spent Nuclear Fuel, High-Level Waste, and Transuranic Radioactive Waste; Final Rule, U.S. Environmental Protection Agency," Federal Register, Vol. 50, Washington, D.C. (1985)

Fischer, G. W., "When Oracles Fail: A Comparison of Four Procedures for Aggregating Subjective Probability Forecasts," Organizational Behavior and Human Performance, 28, 96-110 (1981).

Fischhoff, B., "Debiasing," in Judgment Under Uncertainty: Heuristics and Biases, D. Kahneman, P. Slovic, and A. Tversky (eds.) (Cambridge: Cambridge University Press, 1982), 422-444.

Fischhoff, B., S. Lichtenstein, P. Slovic, S. L. Derby, and R. L. Keeney, Acceptable Risk (Cambridge: Cambridge University Press, 1981).

Fischhoff, B., P. Slovic, and S. Lichtenstein, "Fault Trees: Sensitivity of Estimated Failure Probabilities to Problem Representation," Journal of Experimental Psychology: Human Perception and Performance, 4, 330-344 (1978).

Fluor Technology, Inc., Salt Repository Emplacement Mode Evaluation and Selection, Final Report, Advanced Technology Division, Irvine, CA (March 1988).

Genest, C., and J. V. Zidek, "Combining Probability Distributions: A Critique and an Annotated Bibliography," Statistical Science, 1, 114-148 (1986).

Gettys, C. F., S. D. Fisher, and T. Mehle, Hypothesis Generation and Plausibility Assessments, University of Oklahoma, Decision Process Laboratory, Norman, Oklahoma, Technical Report No. 15-10-78 (1978).

Golder Associates, Report to Rockwell Hanford Operations, Probability Encoding of Variables Affecting Inflow into the Exploratory Shaft Test Facility at the Basalt Waste Isolation Project, Ridgeland, Washington, Redmond, Washington (June 1986).

Goodman, B., "Action Selection and Likelihood Estimation by Individuals and Groups," Organizational Behavior and Human Performance 7, 121-141 (1972).

Gregory, R., and S. Lichtenstein, "A Review of the High-Level Nuclear Waste Repository Siting Analysis," Risk Analysis 7, No. 2, 219-223 (1987).

Hogarth, R., Judgment and Choice: The Psychology of Decisions, (Chichester, England: Wiley, 1980).

Holloway, C. A., Decision Making Under Uncertainty (Englewood Cliffs, NJ: Prentice Hall, 1979).

Howard, R. A., "The Foundations of Decision Analysis," IEEE Transactions on Systems Science and Cybernetics, SSC-4, 211-219 (1968).

Ibrekk, H., and M. G. Morgan, "Graphical Communication of Uncertain Quantities of Nontechnical People," Risk Analysis 7, 519-529 (1987).

Iman, R. L., and J. C. Helton, A Comparison of Uncertainty and Sensitivity Analysis Techniques for Computer Models, Sandia National Laboratories, Albuquerque, NM, SAND84-1461, NUREG/CR-3904 (1985).

Jungermann, H., and M. Thuering, "The Use of Casual Knowledge of Inferential Reasoning," in Expert Judgment and Expert Systems, J. L. Mumpower, L. D. Phillips, O. Renn, and V. R. R. Uppuluri (eds.) (New York: Springer, 1987) 131-146.

Kahnemann, D., and A. Tversky, "Intuitive Prediction: Biases and Corrective Procedures," Management Science 12, 313-327 (1979).

Kahnemann, D., P. Slovic, and A. Tversky, Judgment Under Uncertainty: Heuristics and Biases, (New York: Cambridge University Press, 1982).

Keeney, R. L., Siting Energy Facilities (New York: Academic Press, 1980).

Keeney, R. L., "An Analysis of the Portfolio of Sites to Characterize for Selecting a Nuclear Repository," Risk Analysis 7, No. 2, 195-218 (1987).

Keeney, R. L., "Value Focussed Thinking and the Study of Values," in Decision Analysis: Descriptive, Normative and Prescriptive Aspects, D. E. Bell, H. Raiffa, and A. Tversky (eds.) (Cambridge, England: Cambridge University Press, 1988a), 654-690.

Keeney, R. L., "Structuring Objectives for Problems of Public Interest," Operations Research, 36, No. 3, 396-405 (1988b).

Keeney, R. L., and H. Raiffa, Decisions with Multiple Objectives: Preferences and Value Tradeoffs (New York: Wiley, 1976).

Keeney, R. L., and D. von Winterfeldt, "Improving Risk Communication," Risk Analysis, 6, 417-424 (1986).

Lichtenstein, S., and B. Fischhoff, "Training for Calibration," Organizational Behavior and Human Performance, 26, 149-171 (1980).

Lichtenstein, S., B. Fischhoff, and L. D. Phillips, "Calibration of Probabilities: The State of the Art," in Decision Making and Change in Human Affairs, H. Jungermann and G. de Zeeuw (eds.), (Dordrecht, Holland: Reidell, 1977), 275-324.

Lichtenstein, S., B. Fischhoff, and L. D. Phillips, "Calibration of Probabilities: The State of the Art to 1980," in Judgment Under Uncertainty: Heuristics and Biases, D. Kahneman, P. Slovic, and A. Tversky (eds.) (New York: Cambridge University Press, 1982), 305-334.

Linstone, H. A., and M. Turoff (eds.), The Delphi Method: Techniques and Applications: (Reading, MA: Addison Weseley, 1975).

Matheson, T. E., and R. L. Winkler, "Scoping Values for Continuous Probability Distribution," Management Science, 22, 1087-1090 (1976).

McCormick, N. J., Reliability and Risk Analysis (New York: Academic Press, 1981).

MacGregor, S. Lichtenstein, and P. Slovic, "Structuring Knowledge Retrieval: An Analysis of Decomposing Quantitative Judgments," Organizational Behavior and Human Decision Processes, 42, 303-323 (1988).

Merkhofer, M. W., "Quantifying Judgmental Uncertainty: Methodology, Experience, and Insights," IEEE Transactions on Systems, Man and Cybernetics, SMC-17, 741-752 (1987).

Merkhofer, M. W., and R. L. Keeney, "A Multiattribute Utility Analysis of Alternative Sites for the Disposal of Nuclear Waste," Risk Analysis 5, 173-194 (1987).

Merkhofer, M. W., and A. K. Runchal, "Probability Encoding: Quantifying Judgmental Uncertainty over Hydrologic Parameters for Basalt," Proc. Conference on Geostatistical, Sensitivity, and Uncertainty Methods for Ground-Water Flow and Radionuclide Transport Modeling, B. E. Buxton (ed.) (Columbus, OH: Battelle Press, 1989) 629-648.

Meyer, M. A., and J. M. Booker, Sources of Correlation Between Experts: Empirical Results from Two Extremes, Los Alamos National Laboratory, Los Alamos, NM, NUREG/CR-4814 (1988).

Mosleh, A., V. M. Bier, and G. Apostolakis, "A Critique of Current Practice for the Use of Expert Opinions in Probabilistic Risk Assessment," Reliability Engineering and System Safety, 20, 63-85 (1988).

NRC, 10 CFR Part 60, "Disposal of High-Level Radioactive Wastes in Geologic Repositories, U. S. Nuclear Regulatory Commission," Federal Register, Vol. 48, Washington, D.C. (1983).

NRC, Guidance for Determination of Anticipated Processes and Events and Unanticipated Processes and Events, Draft Generic Technical Position, Division of High-Level Waste Management, Office of Nuclear Material Safety and Safeguards, U.S. Nuclear Regulatory Commission, Washington, D.C. (1988).

NRC, Reactor Safety Study: An Assessment of Accident Risks in U.S. Commercial Nuclear Power Plants, Washington, DC, Nuclear Regulatory Commission, NUREG 75/014 (1975).

Okrent, D., "Comment on Societal Risk," Science 208, 372-375 (1980).

PSAC, Level S, Specification for a Sensitivity Analysis Exercise Based on Level E Test Case, P. C. Robinson, A. Saltelli, A. Alonson, and E. J. Bonano (eds.), Probabilistic Safety Assessment Code User's Group, Nuclear Energy Agency, Organization for Economic Cooperation and Development, Paris, France (1989).

Pearl, J. A., A Goal Directed Approach to Structuring Decision Problems, University of California, Los Angeles, Report No. UCLA-ENG-7811 (1978).

Peterson, C. R., and L. R. Beach, "Man as an Intuitive Statistician," Psychological Bulletin 68, 29-46 (1967).

Pitz, G. F., "Subjective Probability Distributions for Imperfectly Known Quantities," in Knowledge and Cognition, L. W. Gregg (ed.) (New York: John Wiley & Sons, 1974).

Pitz, G. F., and N. Sachs, "Judgment and Decision: Theory and Application," Annual Review of Psychology, 35, 139-163 (1984).

Pitz, G. F., N. Sachs, and J. Heerbroth, "Procedures for Eliciting Choices in the Analysis of Individual Decisions," Organizational Behavior and Human Performance, 26, 396-408 (1980).

Poulton, E. C., "Models of Biases in Judging Sensory Magnitude," Psychological Bulletin, 86, 777-803 (1979).

Raiffa, H., Decision Analysis (Reading, MA: Addison Weseley, 1968).

Raiffa, H., and R. L. Schlaifer, Applied Statistical Decision Theory, (Cambridge, MA, Harvard University Press, 1961).

Ramsey, F. P., "Truth and Probability," in The Foundations of Mathematics and Other Logical Essays, R. B. Brainwaithe (ed.), (New York: Harcourt Brace, 1931).

Ravinder, H. V., D. N. Kleinmuntz, and J. S. Dyer, "The Reliability of Subjective Probabilities Obtained Through Decomposition," Management Science, 34, No. 2, 186-99 (February 1988).

Savage, L. J., The Foundations of Statistics (New York: Wiley, 1954).

Shachter, R. D., "Probabilistic Inference and Influence Diagrams," Operations Research, 36, 589-604 (1987).

Seaver, D. A., *Assessing Probability with Multiple Individuals: Group Interaction Versus Mechanical Aggregation*, Social Science Research Institute, University of Southern California, Los Angeles, CA, Research Report No. 78-3 (1978).

Selvidge, J., *Experimental Comparison of Different Methods for Assessing Extremes of Probability Distributions by the Fractile Method*, Technical Report, Graduate School of Business, University of Colorado, Boulder (1975).

Shanteau, J., "Psychological Characteristics of Expert Decision Makers," in *Expert Judgment and Expert Systems*, J. Mumpower et al. (eds.), (New York: Springer, 1987) 289-304.

Slovic, P., B. Fischhoff, and S. Lichtenstein, "Behavioral Decision Theory," *Annual Review of Psychology* 28, 1-39 (1977).

Spetzler, C. S., and C. A. Stael von Holstein, "Probability Encoding in Decision Analysis," *Management Science* 22, 340-352 (1975).

Stael von Holstein, C.-A., S., and J. E. Matheson, *A Manual for Encoding Probability Distributions*, SRI International, Menlo Park, CA, SRI Project 770778 (1979).

Stillwell, W. S., D. A. Seaver, and J. P. Schwartz, *Expert Estimates of Human Error Probabilities in Nuclear Power Plant Operations: A Review of Probability Assessment and Scaling*, Nuclear Regulatory Commission: Washington, D.C., NUREG CR-2255 (1981).

Thompson, B. G. J., C. J. Frizelle, D. B. Nichols, and M. C. Thorne, *A Method of Overcoming Limitations of Conventional Scenario-Based Risk Assessment by Using Monte-Carlo Simulations of Possible Future Environmental Changes*, transmitted to Scenario Working Group, Nuclear Energy Agency, Organization for Economic Cooperation and Development, Paris, France (1988).

von Winterfeldt, D., and W. Edwards, *Decision Analysis and Behavioral Research* (New York: Cambridge University Press, 1986).

Wallsten, T. S., and D. V. Budescu, "Encoding Subjective Probabilities: A Psychological and Psychometric Review," *Management Science*, 29, 151-73 (1983).

Weber, M. O., F. Eisenfuehr, and D. von Winterfeldt, "The Effects of Splitting Attributes on Weights in Multiattribute Utility Measurement," *Management Science*, 34, 431-445 (1988).

Westinghouse Electric Corporation, *Phase I Stud of Metallic Cask Systems for Spent Fuel Management from Reactor to Repository*, Westinghouse Waste Technology Services Division, Madison, PA, Report WTSD-TME-085 (1986).

Wheeler, T., S. C. Hora, and W. Cramond, Analysis of Core Damage Frequency: Expert Elicitation on Internal Events, Vol. 2, U.S. Nuclear Regulatory Commission, NUREG/CR4550 (1989).

Wheeler, T., S. C. Hora, and W. Cramond, Front-End Expert Opinion Elicitation: Draft Report, Sandia National Laboratories (1988).

Wilson, R, "Commentary: Risks and Their Acceptability," Science Technology and Human Values 9, 1-22 (1984).

Winkler, R. L., Introduction to Bayesian Inference and Decision (New York: Holt, Rinehart, and Winston, Inc., 1972).

Zinnes, P., "Scaling," Annual Review of Psychology 20, 447-488 (1969).