Chapter 3
Elicited Expert Knowledge

Jane M. Booker

Every decision and problem solution involves the use of knowledge gained from the experiences and thought processes of humans. Even for data-rich problems, humans influence how data are gathered, interpreted, modeled, and analyzed. For data-poor problems, such as those assessing risks of never-seen, rare, or one-of-a-kind events, knowledge from experts may be the sole available source of information. Assessing the risk of nuclear deterrence failure is an ill-posed problem that falls into the data-poor category. As a result, experts are needed (1) to supply the information and knowledge for the risk assessment and (2) to define and structure the deterrence problem. These two uses of elicited expert knowledge are discussed. For both, formal elicitation methods for bias minimization are recommended and briefly described. Formal elicitation also involves planning and the use of methods for obtaining the best-quality information from the experts' thinking and problem solving. This formalism includes the characterization of uncertainties, which are prevalent in the deterrence problem, and the analysis of the elicited information, which is necessary for assessing the likelihood and consequence constituents of risk.

Every decision and problem solution involves the use of knowledge gained from the experiences and thought processes of humans. For considering many problems of scientific or technical natures, observations, experiments, and tests provide useful data and insight into the physical world. For example, in meteorology, large amounts of data are continuously available for modeling and forecasting. In contrast, the problem of assessing the risk of the failure of nuclear deterrence is a data-poor problem. Historically, only two incidents of nuclear weapons use have occurred, both during World War II. There have also been other events relating to deterrence failure, such as the close call of the Cuban missile crisis in the 1960s. However, the limited historical data that exist on both actual use and close calls are subject to different interpretations. Theory or fundamental principles about the behavior of nations and groups of people are inadequate and lack sufficient validation to augment the sparse historical record with authoritative information. For such data-poor problems, analysts rely heavily on knowledge from experts.

While everyone can have an opinion, not everyone is an expert. Experts are recognized by their peers as knowledgeable in a subject-matter field and qualified to solve problems and to answer questions related to the subject matter.[1] Some use the terms *subject-matter expert* and *source expert*. The term *knowledge* is used in this chapter to distinguish the expertise formally elicited from peer-recognized experts from opinions that are asked of nonexperts or asked in an ad hoc manner. Examples of the latter would be a reporter asking a person on the street for their opinion about a current event or quoting a person's internet posting. In contrast, formal elicitation of knowledge involves careful planning and preparation of the subject matter, the selection of experts, the question formulation, the response format, the elicitation environment, the elicitation techniques to be used, and the analysis methods used to obtain results. A few scholars have published on these formal elicitation techniques,[2] with Meyer and Booker being the first.[3]

The primary goal of formal elicitation is to gather the best-quality knowledge, in as pristine a form as possible, from experts. This goal imposes a general tenet and approach: to design, implement, and analyze an elicitation that is expert oriented by using the terminology, practices, and cognition of the experts. Formal elicitation draws from many fields, including cognitive psychology, decision analysis, statistics, mathematics, anthropology, and knowledge acquisition. The elicitation and analysis methods are designed to detect, counter, or minimize biases arising from human cognition and behavior and to add rigor, defensibility, and ability to update ever-changing knowledge.[4]

Because knowledge is constantly changing, it is important to understand that an elicitation captures the current state of knowledge, no matter how poor or uncertain it may be. In rare-event subject areas such as nuclear deterrence failure, expert knowledge carries a heavy burden, perhaps being the sole source of information for long periods of time. Such reliance on expertise in these cases makes the goals of formal elicitation even more important.

Formal elicitation serves two purposes in considering the problem of assessing the risk of nuclear deterrence failure:

1. Elicited knowledge is necessary to provide information for any of the techniques and methodologies used for assessing the risk of nuclear deterrence failure.
2. Elicited knowledge can also prove useful in structuring the problem and selecting methods for assessing risk.

The first section of this chapter outlines topics related to and methods for conducting formal elicitation and analyzing elicited knowledge for use in the first purpose. For the second purpose, the second section of this chapter describes how formal elicitation can be used as a methodology for structuring a problem with an unknown structure, such as the ill-posed, data-sparse, multifaceted assessment of the risk of nuclear deterrence failure.

Formal Elicitation and Analysis of Expert Knowledge

The use of expert knowledge is central to all approaches to assessing the risk of deterrence failure because of the shortage of data, information, and knowledge. Regardless of the approach, methods, or models used to structure and represent this problem, data, information, or knowledge is required to characterize its features, issues, components, and conditions. A primer on formal elicitation provides guidelines designed for data-poor problems, such as this one, and covers the highlights of bias minimization and analysis methods.[5] More detail about planning, designing, implementing, and analyzing the elicitation is available in Meyer and Booker's book.[6]

Elicitation Topics

Some of the topics in eliciting expert knowledge most important to the problem of assessing the risk of nuclear deterrence failure are briefly described here. Additional topics relevant to analyzing elicited knowledge are discussed in the "Analysis Topics" subsection.

Biases

Biases are a slanting, adjusting, or filtering of an expert's thinking and original knowledge due to their needs (motivation) and through cognitive processing. Biases degrade the quality of elicited knowledge through distortion. To counter these deleterious effects, formal elicitation includes bias minimization methods for monitoring and/or controlling common biases.

Table 3.1 lists names and descriptions of common biases. While names of biases may vary in different subject areas, their descriptions and effects are common across problems. For example, near-miss bias can be described as a combination of overconfidence and availability biases.[7]

Nuclear war and deterrence are highly emotional topics, and factions exist on multiple sides of associated issues. Experts tend to place undue importance on the few facts available to them, be wishful about outcomes that support their views and agendas, and anchor to their own experiences. Availability bias is strong because experts may not have been alive when nuclear weapons were used in Hiroshima and Nagasaki, and many were children during the Cuban missile crisis. Other events may never have been widely publicized (e.g., the Norwegian meteorological rocket launch in 1995 and the Russian reaction to the NATO Able Archer 83 exercise in 1983).[8]

Wishful thinking bias manifests itself in experts with strong personal or emotion-based agendas that filter or change their expertise to fit a desired result about the success or failure of nuclear deterrence. Waltz and Sagan exhibit this bias; each uses the same historical record as evidence for his own case.[9] Waltz assesses that deterrence has been and will be a successful policy and interprets history to fit that assessment. Likewise, Sagan assesses that deterrence is prone to failure and interprets the same history to fit his view. As another example of wishful thinking bias, experts may exaggerate the risk of deterrence failure to support their favorable view of missile defenses.

Experts often anchor to their initial assumptions, conditions, or responses even when presented with opposing or new, indisputable information. Anchoring bias is detectable, and experts can be made aware of this bias. However, it is easier for anchoring to go undetected or unchallenged when there is uncertainty about whether the new information is valid; therefore, anchoring bias is difficult to detect and to overcome for the deterrence failure problem.

Another anchoring bias is that humans inherently assume that others think and behave in the same way they think and behave. The close call in the NATO Able Archer 83 exercise is one such example. From the Soviet perspective, and consistent with its military doctrine, a nuclear exercise was a useful pretext for a nuclear surprise attack. Soviet leaders, assuming that US leaders think like they do, surmised that a US surprise attack could be the true purpose of Able Archer.[10]

These biases require monitoring and understanding through formal techniques such as probing the experts for explanations, clarifications, and thought processes. Likewise, these techniques aid in distinguishing bias effects from expertise and experience.

Table 3.1. Common biases

	Name	Definition
Cognitive Biases	Anchoring	An expert's failure to sufficiently adjust from their first, long-held, or unchallenged impression in solving a problem—the expert anchors to first, long-held, or unchallenged impression. Sometimes this bias is explained in terms of Bayes' theorem as the failure to adjust knowledge in light of new information as much as it should be adjusted using Bayes' mathematical formula.
	Availability	A bias that results from how easily an expert can retrieve particular events from memory. This affects how accurately frequencies (and probabilities) are estimated. Because memory by its nature is selective, a strong agenda will affect retrieval.
	Inconsistency	Inability to maintain the same problem-solving heuristic, definitions, or assumptions through time because of the limited information-processing capacity of the human mind.
	Overconfidence	The tendency to underestimate the true amount of uncertainty in giving an answer. For example, experts are frequently asked to estimate ranges around their answers to reflect their uncertainty. If experts are requested to put a range around their answers such that they are 90 percent sure that the range encompasses the correct answer, they will tend to underestimate the uncertainty by providing a range that is as much as two or even three times too narrow.

Table 3.1—*continued*

	Name	Definition
Motivational Biases	Group think	The tendency to modify knowledge and/or information so that it agrees with that of the group or of the group leader. Individuals are generally unaware that they have modified their thinking and responses to be in agreement. This bias stems from the human need to be accepted and respected by others. Individuals are more prone to group think if they have a strong desire to remain a member, if they are satisfied with the group, if the group is cohesive, and if they are not a natural leader in the group.
	Impression management	Resulting from social pressure, this bias occurs when the expert responds to the reactions of those not physically present. For example, the expert answers survey questions in a way that maximizes approbation either from society in the abstract or from the administrator of the elicitation in particular.
	Misinterpretation of the expert	The altering of the expert's thoughts as a result of the methods of elicitation and documentation.
	Social pressure	An effect that induces individuals to slant their responses or to silently acquiesce to the views that they believe the interviewer; their group, supervisors, organization, or peers; or society in general will accept. This altering of an individual's thoughts can take place consciously or unconsciously. The social pressure can come from those physically present or from the expert's internal evaluation of how others would interpret their responses. People's need to be loved, respected, and recognized induces them to behave in a manner that will bring affirmation. Political correctness is an example.
	Training bias	The tendency of the data gatherer, analyst, or both to misinterpret data/information from others for their own purposes (for example, choosing quotations, references, or events that suit the interviewer's purposes).
	Wishful thinking or conflict of interest	A tendency that occurs when individuals' hopes influence their thinking and responses. For example, people typically overestimate what they can produce in a given amount of time. In general, the greater the experts' involvement and the more they stand to gain from their answers, the greater this bias.

Elicitation Setting

The quality of elicited knowledge depends on the interviewer's ability to question experts about the assumptions they use, the heuristics and cues involved in their thinking, and their problem-solving processes. These details are best elicited in face-to-face elicitation sessions, making a personal interview the preferred setting for eliciting knowledge on the deterrence failure problem. Modern teleconferencing may provide a convenient alternative. Ideally, two interviewers conduct the elicitation: one who has subject-matter expertise and one who has elicitation expertise. As necessary, each should train the other before the interview.

Because of the multidisciplinary nature of this problem, some group elicitations may be necessary for different experts to interact. Group elicitation sessions suffer from biases different from those typical of individual interviews. For example, experts may be prone to agree with an influential member in the group. Group-related biases can also be minimized with proper use of elicitation methods. The setting that provides the least opportunity for the interviewer to understand the expert's thinking and the most opportunity for biases is the mail-in questionnaire.

A common language and terminology may not exist among the different subject areas involved in assessing the risk of nuclear deterrence failure. Thus, for group elicitations and the subsequent analysis, the interviewer must provide experts with background, assumptions, and definitions of terminology from different subject areas. Even with a single expert, the interviewer may have to remind the expert of changes in terminology.

Question Phrasing and Response Mode

One of the most important bias minimization techniques is proper phrasing of questions. Avoiding "loaded" questions such as "When did you stop beating your wife?" requires only minimal effort. However, asking unbiased questions is difficult, especially when the subject is sensitive or emotional, such as might be the case when discussing nuclear weapons use or war. It is helpful to use terminology consistent with the expert's common practice and to repeat the expert's own words back to them. For guidelines on question phrasing, Payne's book is the classic reference[11] and its guidance is used in conjunction with formal elicitation methods.[12]

Response mode refers to the format the interviewer choses for the answers to questions posed to the experts. Examples of response modes include multiple choice, open-ended essay, continuous numerical scale, odds ratio, range of values, comparison, ranking, and likelihood. Some of these are described in the next section on structuring. Likelihood may be a concept consistent with the way many experts think, and it is general enough to encompass definitions used by specific communities of practice. In contrast, probability is only rarely appropriate to very specific communities.

Uncertainty

All knowledge, data, and information have uncertainty associated with them. Uncertainty can be defined as that which is not precisely known. Examples particular to the issue of nuclear deterrence include uncertainty in the number and nature of nuclear close calls, uncertainty about whether a state leader's statements in a speech are true, uncertainty about how a potential adversary views the use of nuclear weapons, and uncertainty about whether a group can construct a nuclear weapon.

More often than not, uncertainties are ignored or assumed negligible because it is difficult to recognize and treat them. When addressed, uncertainty is often measured quantitatively, such as by using a range of values or a probability. However, uncertainty can also be expressed qualitatively when knowledge and information are also qualitative.

The qualitative nature of the knowledge and information associated with the question of nuclear deterrence failure is conducive to qualitative uncertainty representation. The deterrence literature is filled with phrases such as *not impossible*, *possible but not probable*, *plausible*, and *belief*. These words express a degree or measure of uncertainty regarding the subject under consideration. For example, an expert stating that something "is possible but not probable" implies that possible is less likely than probable. The words themselves have an uncertainty inherent in their interpretation. For example, how unlikely is "possible"? Experts expressing qualitative uncertainties should be asked to provide definitions or examples to illustrate the meanings behind their words. This clarification aids in comparing uncertainties between issues and between experts.

General information theories can be used to quantify uncertainties, and they provide standards or yardsticks by which uncertainties can be compared.[13] One general information theory differs from another based on the types of uncertainties it characterizes and the properties (axioms) it follows. For example, Zadeh fuzzy sets and logic have properties designed to turn qualitative linguistic information into quantitative uncertainties.[14]

However, mathematical theory is lacking for combining uncertainties characterized using different general information theories, making it difficult to mix the use of different theories within a problem. This is one reason why probability theory is often chosen for a problem even though it characterizes only one type of uncertainty: the uncertainty of the outcome or result of an indeterminate event.[15] Once that event has occurred, and its outcome determined, there is no uncertainty and the probability of that event is either 1.0 if the event occurred or 0.0 if not. This basic meaning of probability is not readily practiced even by scientific and technical experts.

Quantitative, experimentally derived data are subject to uncertainties from measurement, experimental conditions, initial conditions, environmental or system controls (or lack thereof), and unexplained random variations. Most scientists are taught to characterize these uncertainties by using probability theory. Probability has a mathematical definition based on measure theory and crisp sets. Unfortunately, the reasons for using probability get lost in its common usage—one reason why probability is commonly viewed as the exclusive method for characterizing uncertainty.

Despite the common usage of probability for uncertainty, there are three difficulties in using probabilistic uncertainties for the risk of nuclear deterrence failure. First, not all uncertainties inherent in the deterrence failure problem fit into the probabilistic definition. Uncertainties relating to linguistic information or resulting from conflicting information, misclassification, lack of knowledge or theory, or lack of specific detail or its reverse—generalization—are not well characterized or quantified by probability. Some of the other general information theories are designed to characterize these uncertainties. Regarding linguistic uncertainty, previous attempts have been made to equate or transform words to numbers. One of the most common is the Sherman Kent scale.[16] Weiss developed another scale based on legal standards of proof.[17] The disadvantage of using predefined scales is that an expert's definition of words such as *likely* may not match the definition in the scale. Ideally, each expert would define such a translation based on how they think about the term.

Second, experts tend to violate the axioms of probability theory when providing probability estimates. For example, an expert responding with a probability of 0.05 for a particular event to occur might later respond with a probability of 0.90 for that event to not occur. More difficult-to-detect violations of the axioms of probability include a sum of multiple mutually exclusive and exhaustive probabilities that is not 1.0 and improper estimates of conditional or dependent probabilities.

Often surveys interchange the terms *probability* and *percentage*. While a probability can be transformed into a percentage, a percentage cannot always be transformed into a probability because percentages can be greater than one hundred.

Third, humans (even statisticians) are not well calibrated for estimating probabilities. As a rule, they cannot accurately express their perceived likelihood or frequency of their experiences as probabilities.[18] For extremely rare events, poor calibration of probability estimation can be magnified. For example, it is difficult to distinguish between a probability of 0.000001 and 0.0000001. This is why in some subject areas, orders of magnitude (e.g., the Richter scale for earthquakes) are used. However, if experts are not experienced in thinking in such scales, it is difficult to teach or train them. In general, it is difficult to train experts to accurately estimate probability.[19]

Unless an expert is used to dealing with and thinking in terms of probability, it is best to avoid asking for probability as a response. Other response modes and descriptions are advised, such as odds (betting odds), likelihoods, ratios, ranks, or other comparisons. The choice should be consistent with the expert's community of practice. At the very least, the interviewer should thoroughly define any unfamiliar response mode for the expert.

In those special cases in which probability is appropriate to characterize uncertainty, it should be noted that there are at least two modern interpretations of probability that are equally valid within its theory.[20] The first is what most are taught as probability—the number of event occurrences divided by the total number of outcomes. This is the frequentist or relative frequency interpretation of probability. For example, the probability of drawing a red marble from a jar containing one hundred marbles of which twenty are red is 20/100 = 0.05.

The second is the personalistic interpretation, often referred to as the Bayesian interpretation, the centerpiece of Bayesian analysis. Personalistic probability is an individual's assessed value based on their willingness to bet that they are correct.[21] For example, if an expert states that there is a 0.90 probability that the next terrorist attack on the United States will occur within three months, the expert should be willing to stake $0.90 in exchange for $1.00 if the attack occurs within three months. If the attack occurs within three months, the expert wins the $1.00, for a net gain of $0.10. If the attack does not occur, the expert loses $0.90. To prevent cheating, the expert should also be willing to make the opposite bet, where they are willing to stake $0.10 in exchange for $1.00 if the event does not occur. This two-sided bet is depicted in Table 3.2. In terms of betting odds, this example demonstrates odds of 9 to 1.[22]

Table 3.2. An example of a two-sided bet

Bet	Attack Occurs, p = 0.90	Attack Does Not Occur, p = 0.10
Expert stakes $0.90 in exchange for $1.00 if attack does occur	Expert's net gain is $0.10	Expert's net loss is $0.90
Expert stakes $0.10 in exchange for $1.00 if attack does not occur	Expert's net loss is $0.10	Expert's net gain is $0.90

An expert who believes the probability of attack is 90 percent should be willing to take either side of this bet.

Regardless of whether or not an uncertainty is probabilistic, the interviewer should elicit it along with the responses to the questions asked of experts during an elicitation. The form or format for noting uncertainties should be consistent with the way the experts think and the available knowledge.

One of the recommended forms for eliciting uncertainties is to request a range of answers after eliciting the expert's response. To avoid introducing ambiguous uncertainty in the analysis of experts' ranges, it is necessary to define what the requested range represents. For example, the range could represent absolute highest and lowest values. Unless experts are familiar with percentiles (and most are not), tying range limits to percentiles (e.g., 5th and 95th) is not recommended. To minimize anchoring bias, the expert should be encouraged to consider their range in conjunction with their response, making any necessary adjustments.

When eliciting uncertainty, the common bias of underestimating the real uncertainty should be monitored. Experts tend to be overly optimistic about what is known and to respond with uncertainty estimates that are too narrow relative to the state of knowledge. This is called overconfidence bias.

The word *confidence* is often used in relation to uncertainty. Too often *confidence* is used in a colloquial sense, as the dictionary definition of belief, without any technical, mathematical, or quantitative definition. To lend technical meaning to confidence, it can be defined as the complement or inverse of uncertainty.[23] For example, a commander might tell a general that they are confident the mission will be a success, using the colloquial definition. However, the general could ask for the uncertainty about the success, understanding that the larger the uncertainty, the smaller the confidence.

Neither of these definitions of confidence should be confused with statistical confidence intervals or confidence level. These terms have specific mathematical definitions in statistical inference and hypothesis testing that are not appropriate for the colloquial or technical definitions. Often decision-makers and experts confuse the colloquial definition of confidence with the statistical ones.

Decomposition Principle

Assessing the risk of nuclear deterrence failure is an extremely complex problem that cuts across multiple areas of expertise. It is unlikely that any single expert will have enough expertise to cover all aspects of the problem. Thus, experts from different and diverse subjects will have to participate in the assessment, and the problem will require decomposition into manageable parts.

Studies on human cognition have shown that experts provide more accurate knowledge when the problem is fully specified and broken down into basic constituents.[24] The more complex a problem, the more specification and decomposition is necessary. A simple example illustrates this concept: Estimate how much you spend on your home budget. Then consider all the items in the budget, and write down individual estimates for each: the groceries, utilities, rent/mortgage, clothes, education/business expenses, vacations, medical expenses, etc. The sum of these should differ from your first estimate, and the decomposed total should be more accurate.

The decomposition process includes specifying definitions, conditions, scenarios, assumptions, timelines, quantities, and parties involved. Usually, several preliminary questions that provide these specifications are asked to set the stage for the questions of interest. A structure or framework of the problem provides guidance on how to do the decomposition.

The decompositions and operating conditions of physical systems can be easily represented because of their structure. However, decompositions of complexities of human behaviors, timelines, or event sequences—all of which are applicable to assessing the risk of failure of deterrence—may not be so obvious or conducive to common structures such as fault trees. The nuclear deterrence failure problem currently lacks a systems perspective (and hence structure) or model, making decomposition difficult. Even establishing initial or boundary conditions may pose challenges because of all the facets and factors involved. It may be possible for experts to contemplate some specifically defined scenarios or special cases and begin decomposing the problem by using those.

Risk analysis has two aspects: likelihood and consequence. Risk studies usually address the likelihood first and then the consequences, even though there are interdependencies between them. Deterrence also has two aspects: capability and credibility. Both should be evaluated from the perspective of the party being deterred, and again, there are dependencies.[25] Because of the dual natures of both risk and deterrence, decomposition is a necessity for the problem of assessing the risk of nuclear deterrence failure. Other decompositions could be based on issues such as state versus non-state nuclear use; a single weapon attack versus multiple weapons; attack on US homeland versus elsewhere; unauthorized versus authorized use; and accidental versus intentional use.

Ill-Posed Problem Decomposition

The risk assessment of nuclear deterrence failure is an ill-posed problem because it is knowledge sparse, complex, and multifaceted and involves multiple subject areas and large uncertainties of various types. Thus, there is a temptation to elicit knowledge at a general level, ignoring decomposition and failing to capture specific expertise. An example of what can happen when a nonspecific question is asked of experts, consider question 5 from the Lugar survey.[26] Figure 3.1 is discussed in chapter 1 relating to biases and reprinted in this chapter for convenience. As noted in chapter 1, it shows the varied responses of seventy-nine experts to the question, "What is the probability (expressed as a percentage) of an attack involving a nuclear explosion occurring somewhere in the world in the next ten years?" While this question may sound specific, the geopolitical conditions leading up to such an event were not specified, assumptions about the attacker were absent, and what constitutes an "attack" was not defined, leaving each respondent free to decide what these factors might be. The wide variety of responses suggests that different experts answered differently based on their assumptions and what they were free to specify in their thought processes (but were not asked to report). As noted in chapter 1 and in the bias subsection above, such lack of specifics provided to the experts opens the door for biases to dominate, adding to the wide dispersion seen in Figure 3.1.

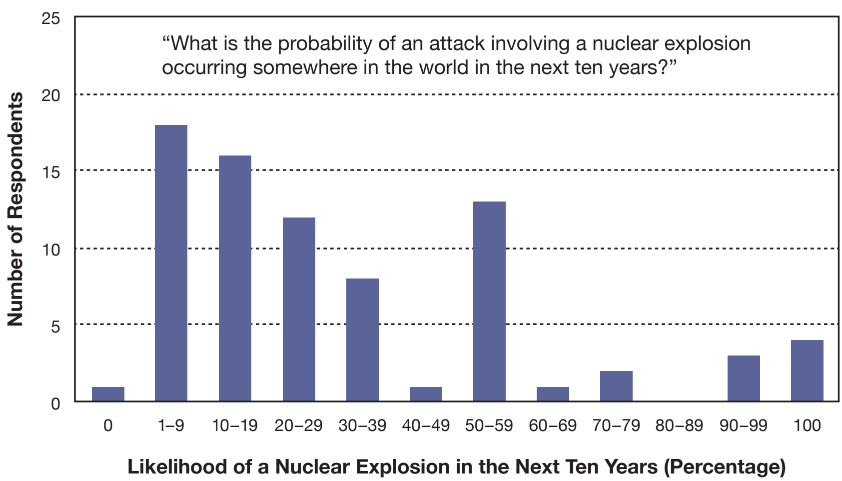

Figure 3.1. The Lugar survey, question 5.

While it is important to select a diverse group of experts to ensure the state of knowledge is represented, such a dispersion of responses could also indicate that some respondents did not know how to answer because of lack of expertise so they opted for the middle-percentage answers. However, even with expertise, experts may supply a middle response (e.g., 50 percent) to indicate their large uncertainty about the answer. It is not uncommon for experts who have strong biases regarding the probability of attack to respond with the extremes of 100 percent and 0 percent. Detecting such bias and getting experts to expand their thinking beyond their anchored views is what bias minimization elicitation is all about.

The nuclear terrorism and war literature contain some examples of decomposing the complex and ill-posed deterrence problem. Bunn,[27] Hellman,[28] and Mueller[29] decomposed the problem into separate events for evaluation. Each provided their own problem structuring for the conditions and assumptions of the events they chose. Each then provided their own estimates of the likelihoods of these events and descriptions of how to combine or propagate those estimates to obtain the final answers.

Their analyses of their versions of the problem could be called self-elicitations. Self-elicitations are very prone to biases when questions are not properly phrased and problem-solving is not monitored, as was the case in these authors' evaluations. The disadvantages of their analyses are that the authors' biased responses were driven by their personal agendas and that it is possible that not every author is an expert. The advantage of written self-elicitations is that authors tend to describe their thought processes, reasons for structuring the problem in a particular way, and reasons for their personal responses.

Decomposing a complex and/or an ill-posed problem into manageable parts and diligently defining the specifics of each question relating to those parts not only aids in eliciting pristine knowledge from experts, but it also helps determine whether different experts are answering slightly different versions of the same question. Differences in experts' assumptions, definitions, conditions, problem-solving processes, and interpretations of the question can result in different responses, such as those seen in Figure 3.1. Decomposing the problem and using formal elicitation methods helps the interviewer avoid those kinds of results.

Analysis Topics

After expert knowledge has been elicited, it must be analyzed. The analysis topics described in this section are part of formal elicitation design and implementation. The particular topics were chosen for inclusion because of their importance to the nuclear deterrence failure problem.

Selection and Motivation of Experts

For analysis results to have interpretive meaning about the current state of knowledge, the selected experts must be a representative subset of all such experts. To ensure proper representation, a random sample or other statistical sampling method should be used to select experts. However, there may be few experts in existence. In that case, the goal should be to get participation from as many as possible. If the entire set of experts is known to be composed of groups based on factors such as opposing views, varying levels of expertise or experience, and different backgrounds, the selected experts should represent those different groups. For example, a poorly designed selection would include only experts who work in Washington, DC, or only experts who hold strong anti–nuclear weapons views.

To avoid experts' nonparticipation or nonresponse, it is necessary to motivate their participation from the beginning. Motivations include flattery, compensation, and collaboration. Experts can be motivated and encouraged by reminders that their work is fundamental and is breaking new ground. Likewise, it is important to keep in contact with experts to encourage them to provide the requested knowledge in a timely manner. Motivation is difficult if a mail-in survey is the chosen elicitation setting. Lack of participation can undermine the care taken to obtain a representative selection of experts and can adversely affect conclusions drawn from the elicited knowledge.

Feedback: The First Analysis Step

After a representative set of experts is selected and the elicitation has been conducted, compiling and reviewing the experts' responses for clarity and errors is the first step in analysis. This step is likely to involve re-contacting the experts. At that time, they should be reminded of what analysis is planned for the knowledge they provided—of which they should have been informed when first interviewed. They can review their responses and reasons for them. This is the feedback process.

An analyst will be tempted to interpret the experts' responses in such a way as to make the analysis job easier. In doing so, the analyst introduces bias. For example, if the analyst wants to analyze the responses as average values, the experts should have been asked to provide averages. It is vital to plan ahead for the kinds of analyses anticipated so that proper questions and response modes can be provided to the experts. Response modes should be chosen based on how the experts think rather than for the convenience of the analyst.

Experts' Problem Solving and Cognition

Much of the wide dispersion (a type of uncertainty often measured by a variance) in the responses in Figure 3.1 could be understood if the experts had recorded their thoughts and problem-solving processes while answering the question. These activities are part of formal elicitation design. Querying the experts about their thinking and problem-solving processes is conveniently done in a face-to-face interview. It can also be done during the feedback process to clarify responses.

Probing into cognitive and problem-solving processes is important for determining whether an expert is answering the posed question or some modified or misinterpreted version. Often experts think about conditions, assumptions, cues, and experiences and use problem-solving methods that affect their responses, but these thoughts and methods may not be recorded. Changing one or more of these could significantly change an expert's response. If the analyst does not know details about how the experts answered a question, the analyst will not be able to draw proper conclusions or resolve disagreements among experts.

A simple example illustrates the importance of eliciting cognitive and problem-solving processes. Experts A and B both respond with high likelihoods of nuclear weapon use within the next ten years. However, after eliciting their problem-solving processes, it is discovered that expert A assumes a terrorist use while expert B assumes an interstate war. Further probing reveals that expert A considers the interstate war an unlikely situation for nuclear use and expert B considers nuclear terrorism unlikely. Thus, without knowing what the experts were assuming when responding to the nuclear use question, their apparent agreement is not the correct conclusion. Experts A and B were actually providing different answers based on different assumptions and cognitive processing.

The analyst is often faced with determining the degree of dependency among experts. This is important if experts' responses need to be aggregated (e.g., reporting an average response as done in the Lugar report). Experts who are highly dependent are expressing the same knowledge and cannot be counted as independent sources. It is difficult to determine the extent of overlapping or double-counted knowledge from a group of experts. Without details about how experts arrived at their responses, dependency determination becomes untenable. Experts who solve problems by using similar methods tend to produce similar responses, illustrating the importance of eliciting experts' cognitive processing for monitoring dependence.[30]

The analyst is a unique position to compare responses to multiple questions for each expert. Such analysis can check on the expert's self-consistency and understanding of the subject. It can also be used to indicate a change in a definition or assumption used by the expert and monitor biases.

Conditionality

Every piece of knowledge, model, answer, and problem is conditioned on things known and unknown, admitted and unaware. These conditions could be boundary conditions, scenarios, environments, settings, cases, domains, levels of detail (granularity), cues, rules, heuristics, or assumptions. A thorough, formal elicitation should uncover as many of these conditioning factors as possible, given constraints on time and budget. Different conditions considered in the expert's thinking often produce different responses. For example, two experts with the same experience, education, and viewpoints can produce different answers because one is using an assumption different from the other's. That assumption can be considered a different problem-solving process or a different model used by the expert.

Models applied to portions or the entire problem are also considered conditions because results may change if a different model is used. An example of such a model is Perrow's complexity theory, which describes the interaction of humans with technology.[31] Sagan's organizational theory uses a different conditional modeling—the interaction of humans with their environment (organization) and its influences (conditioning) on them.[32] Taleb (of black swan fame) recommends using expert knowledge but wisely warns about watching out for the assumptions and conditions found in modeling.[33]

An example of the importance of conditionality is found in probability theory. If conditional probabilities are not carefully and properly estimated, their combination can produce a result that violates probability axioms. This violation due to ill conditioning is the basis of Borel's paradox.[34]

Expert Resolution and Aggregation

Clarification gained by resolving differences in experts' responses by using their problem-solving processes is necessary for achieving a consensus of experts and for the analyst to aggregate experts' responses. An accurate consensus cannot be achieved if experts do not mutually understand the reasons behind their responses. An inaccurate consensus can arise because of biases, such as fatigued experts acquiescing just to end a meeting or following the party line in the presence of authority. The analyst could be inadvertently combining nonidentical responses if they do not know the experts' reasoning. Neither consensus nor aggregation may be necessary or practical (because they are too difficult to achieve) for the problem of assessing the risk of deterrence failure. A decision-maker would be better served by being given the full spectrum of responses (with uncertainties) from a diverse set of experts.

Aggregating experts' responses suffers from the same problem as any combination scheme: Should all experts' responses be considered equal? If not, how should experts be weighted? Recall that experts are identified by their peers; thus, those same peers are a reasonable source of weights. Self-identified weights are the next reasonable source; however, some experts can be overly modest, and some can be overly arrogant. The analyst or decision-maker should not determine weights for experts. Weights can differ from question to question, because some experts may be more knowledgeable than others in differing subject areas. Such determinations can be very complicated and time consuming. Assuming the experts are peer identified, then the simplest solution is to weight experts equally. This is the maximum entropy[35] solution and is recommended unless a good reason and a good method for discriminating among experts exist.

Cooke advocates aggregation of experts' responses through a process by which they are calibrated.[36] This calibration involves training and testing the experts—a time-consuming process. However, this approach is of limited use in subject areas in which data and experience are sparse and/or knowledge and theory are not well known. For calibration to be effective, feedback to the experts must be (1) immediate, (2) frequent, and (3) relevant to the subject. One of the few areas that meets these criteria is meteorology, which has theory, models, and huge amounts of data for forecasters to consult and improve their predictions. The nuclear deterrence problem is the opposite: it is data poor and theory poor. Thus, calibration is not recommended.

Drawing Conclusions

Usually the reasons for analysis are to summarize the elicited responses and to draw conclusions from them, often to inform decision- and policy-makers. Even though elicited expert knowledge is not a substitute for experimental, historical, or observational data, it can be analyzed and conclusions can be drawn from it. If there is ever a time when data might become available, elicited and analyzed expert knowledge can be considered a placeholder for those future data and can be compared and combined with the future data.

For highly qualitative responses, there may be little opportunity to analyze the information elicited by using statistical or data analysis methods. While qualitative knowledge can sometimes be grouped or categorized, this is subject to misinterpretation bias. If the responses are continuous numeric quantities, integers, ordinal, or categorical, then statistical analysis methods are useful for providing defensible conclusions inferred from experts' responses.

Decision-makers may be accustomed to seeing a central aggregated response from all the experts—a mean (the average of numerical values), median (the middle of the range), or mode (most frequent or common value). For example, the mean for the question in Figure 3.1 is 31 percent, which falls in the 30–39 percent bin. The median of seventy-nine values is the fortieth value, which falls in the 20–29 percent bin. The mode is the bin with the largest count, the 1–9 percent bin.[37] Because of how these three differ, the conclusion is that these data are not distributed symmetrically around a central value. Figure 3.1 visually confirms the lopsided loading of the data in the lower percentages. The wide dispersion of responses in Figure 3.1 is summarized by the large standard deviation—an uncertainty metric for dispersion—of 28 percent. Another common measure of dispersion uncertainty is the range, which is 100 percent.

Statistical methods can be used to determine whether the experts responded uniformly across the percentage scale as might be suspected in Figure 3.1. The answer here is no; significantly fewer than expected experts responded in the percentage bins labeled 0, 40–49, 60–69, and 80–89, and too many responded in the bins labeled 1–9, 10–19, 20–29, and 50–59.

It is anticipated that different experts may have different perspectives and perhaps strong personal agendas. Such differences can emerge from divisions or factions within a subject-matter community. For example, a discussion of nuclear war tends to divide viewpoints into factions based on the emotional response that concept evokes. That emotion translates to inducing bias as experienced from decades of elicitation efforts on sensitive and taboo topics, including nuclear weapons and war. The deterrence community also appears to be divided into factions regarding the effectiveness, or lack thereof, of nuclear weapons. The well-documented debate of two such factions can be found in the works of Sagan and Waltz.[38]

Analysts should be aware of such perspectives and should question experts about their preexisting (i.e., anchored) positions. Along with that, other questions about the experts' specific areas of research and experience provide information about how their responses may be biased. Statistical analysis may be able to determine whether or not these biases affect responses, by comparing responses among experts whose preexisting positions are established and whose problem-solving processes have been elicited.

Linguistic responses and qualitative descriptive answers are more difficult to analyze than quantitative responses such as those shown in Figure 3.1. However, the knowledge gained from these responses is more detailed than that obtained by forcing experts to collapse their knowledge into a single numeric response. Some linguistic responses can be categorized and category responses counted, permitting some analysis. The analyst must accept the fact that some responses cannot be graphed, counted, or analyzed in any manner. In such cases, thorough and unaltered documentation of responses accurately captures the current state of knowledge for that question.

Reviewing information from the nuclear deterrence and war literature illustrates some of the issues regarding drawing conclusions, uncertainty, and conditionality. Table 3.3 (discussed in chapter 1 and reprinted in this chapter for convenience) presents a set of estimates of nuclear war or terrorism from various authors who have written about the subject (and who may or may not be considered experts). At first glance, these authors appear to be estimating the same thing—the probability of nuclear war—but with widely different results. The table divides the sources into two different subjects (conditions), war and terrorism. Different response modes are used: some of the estimates are percentages, some are odds, and some are ratios (1 in *n* or *x* in *n*). In addition, the estimates have different time conditions: four estimates apply to the next decade, two are per year, one is per attempt, and two are specific to the time during the Cuban missile crisis. It should be noted that lack of specificity is a type of uncertainty because the analyst looking at this table faces the conundrum of how to compare results from unspecified conditions to the results from specified ones.

Table 3.3. Individual estimates of the probability of nuclear war

	Question	Estimate	Author	Year
War	Probability that the Cuban missile crisis could have escalated to (nuclear) war?	Between 1 in 3 and even (war)	John F. Kennedy	1962
		As large as 1 in 100 (nuclear war)	McGeorge Bundy	1988
	Probability of a future Cuban missile-type crisis that results in at least one nuclear weapon being used?	2 in 1,000 to 1 in 100 per year	Martin Hellman	2008
Terrorism	Probability that terrorists will detonate a nuclear bomb?	More likely than not (on America)	Graham Allison	2004
		50–50 odds within the next decade	William Perry	2004
		Less than 1 percent in the next 10 years	David Albright	2005
		29 percent probability within the next decade	Matthew Bunn	2007
		10–20 percent per year against a US or European city	Richard Garwin	2007
		Less than 1 in 1,000,000 (per attempt)	John Mueller	2008

The analyst might be able to resolve other response differences by determining each author's viewpoint, understanding what information the author used, discovering how the author structured or modeled the problem, and gaining insights into the author's cognition by reading the author's papers. Without such conditioning information, the analyst can only compare "apples to apples." The four estimates that terrorists will detonate a nuclear bomb in the next decade have a large, unexplained, range of 1 percent to 50 percent. The other estimates cannot be included with these unless and until the conditional factors inherent in them are known, putting them in the same terms as the first four. The Bundy and Kennedy estimates can be compared to each other but not the rest.

Informing Decision-Makers

Quantifying or summarizing results from elicitation and analysis should be done in a form useful for and understandable to decision- and policy-makers. Determining that format may involve an elicitation with the decision-maker. While top-level managers rely on executive summaries, details should be made accessible for their staff and for future updates as knowledge changes.

Returning to the data in Figure 3.1, quoting the mean response of 31 percent to a decision-maker without the uncertainty does not convey an adequate summary of these data. In this particular case, the histogram in Figure 3.1 does provide an appropriate summary. However, a decision-maker who is unfamiliar with histograms (or who is uncomfortable with graphs and bar charts) should be given verbal descriptions and explanations of the data, using that decision-maker's usual terminology, rather than shown Figure 3.1. Another disadvantage of Figure 3.1 is the choice of intervals for the bins. Of note is the large count in the bin for 50–59 percent. A reason for this may be that some experts opted for the 50 percent response. The decision-maker should be given the 50 percent count instead of mixing it with other responses in the 50–59 percent bin.

There are creative and informative visual displays for data and information available with apps, such as the word frequency generator, Word Cloud, from Microsoft. Most of these modern tools have their origins with those in Edward Tufte's seminal books.[39]

Eliciting Problem Structure

Assessing the risk of nuclear deterrence failure is an ill-posed, complex, multifaceted, knowledge-sparse problem spanning multiple subject areas. These characteristics make structuring this problem a challenge. A problem structure can be described generally as a recorded representation of a problem in the real world, organized into a useful format of pieces, facets, or aspects (often designated with boxes), which are interconnected according to some order, association, hierarchy, time flow, or logic. A problem structure:

- Defines the boundaries and scope of the problem (which facets and subject areas will be included and excluded)
- Defines the top-level or bottom-line question (what is the risk of deterrence failure?)
- Provides a logic flow that cohesively connects all aspects of the problem to answer the top-level question. Such a flow could be a timeline (e.g., an event sequence), a hierarchy (e.g., general to specific parts), or specified relationships (e.g., dependencies, influences and conditions, and mathematical models), to name a few possibilities.
- Guides the formation of questions about smaller problem aspects or parts
- Provides a mechanism for capturing and recording experts' thoughts and problem-solving processes in an elicitation
- Guides the use of the decomposition principle for an elicitation
- Provides the relationships, connections, and associations of problem parts and aspects for analyses
- Provides a framework or skeleton on which all the available and applicable data, information, and knowledge are attached

For purposes here, the term *framework* refers to the general problem outline, concept, and scope, while *structure* refers to establishing order, organization and arrangement, logic flow, and connections and interrelationships of problem aspects and parts. A framework is part of the structure and is related to it like a skeleton is to a body.

Applicability of Established Structures

Determining problem structure for an ill-posed, knowledge-sparse, multifaceted problem, such as the risk of nuclear deterrence failure, is challenging because many of the established structuring methods may not be applicable. Established structures from risk and reliability analyses include fault trees, reliability block diagrams,[40] and event trees[41] and are designed to represent physical systems. Such systems have definite structure and are designed for specific modes and environments for operation. Complex or ill-posed problems involving human behavior, such as the problem of nuclear deterrence, may not be so easily decomposed into discrete parts the way a physical system can be. Defining what constitutes the whole—the "system" and its boundaries—for this problem is also a challenge. Even determining what constitutes success or failure may not be clear, precise, or crisp in the deterrence "failure" problem.

Other established structures follow timelines and logic flow sequences in operations and processes. Examples indicating the wide variety of such structures include computational algorithms, flowcharts, manufacturing processes, communication networks, PERT charts,[42] Gantt charts, electrical circuit or wiring diagrams, blueprints, chemical and physics reaction sequences, and assembly processes. The structures for these problems usually involve human interaction with physical systems and physical processing, so the physical system supplies the structure. Again, these structuring methods do not readily apply to the nuclear deterrence problem, which is not a physical system. In addition, it may be difficult to define and prescribe predominantly human processes or sequences because of unknown behaviors, politics, etc. Multiple parallel activities may cease and restart for unknown reasons.

Established structures from decision sciences may be somewhat applicable because they deal with human thinking and actions in decision-making and problem solving. These structures include decision trees[43] and influence diagrams.[44] Because of these structures' popularity, many software packages exist to allow users to create them. Decisions, actions, and causalities are specified in the structuring of a decision problem. Connections between these are limited to specific relationships according to the mathematics used. The mathematical framework is utility theory, which has its origins in game theory.[45] However, it is difficult for humans to think and behave in accordance with this mathematics. Thus, while the diagrams and interrelationships (e.g., influencing factors) may be useful for the nuclear deterrence problem, the mathematics used to perform the analysis and glue the structure together to arrive at the top-level answer may not be appropriate.

For the reasons mentioned (and others), established structures may not be applicable to assessing the risk of nuclear deterrence failure. The goal for structuring this problem is to take the ill out of the ill-posed problem or at least understand the difficulties and what knowledge would be needed to overcome them. An alternative to applying established structures for the assessment of the risk of deterrence failure is to elicit problem structure consistent with the way the experts think about the problem. Reasons why this alternative is attractive for the nuclear deterrence problem follow.

Reasons for Eliciting Problem Structure

Expert knowledge will be the primary source of knowledge for the nuclear deterrence failure problem, and the structure should be consistent with experts' thinking about the problem, according to elicitation principles. However, experts from the different subject areas involved may not agree on how their portions of the problem should be structured. If those differences are not resolvable, then reasons for those differences can be documented.

Experts may think about their portions of the problem using relationships and connections not easily accommodated by established structures. These relationships include feedback loops, complex associations spanning or crossing different facets or dimensions, partially or ambiguously defined influences, vague or indeterminate conditions and dependencies, and complicated networks. Network structures (e.g., Bayesian networks) permit conditional probability types of dependencies with a hierarchical structure, but the logic flow defined by the mathematics is cumbersome and is not easily understood by experts (or analysts) outside of the Bayesian analysis community. Experts should be permitted to define whatever conditional relationship or network necessary without being forced to fit them into a prescribed mathematical rule set or axioms.

When encountering problems with relating poorly known, interacting, continuous processes not suited for established structures, experts do not think of problem features as discrete boxes with definitive connections. For example, a physicist or chemist resists structuring the kinetics of an explosion into sequences of well-defined boxes. This is because of the lack of detailed fundamental knowledge required to "box" and because of the complexities (some poorly known) of the processes involved. The problem of assessing the risk of nuclear deterrence failure may suffer from the same difficulties.

However, even for difficult, amorphous, or ill-posed problems, experts tend to think in terms of some sort of problem structure or framework based on the logic behind their understanding. That structure may be loosely defined, choppy, disjoint, approximate, general, vague, and difficult to record on paper, a whiteboard, or a computer pad/tablet. Detailed probing into the expert's thinking may be required to elicit a rough draft that mimics the expert's thoughts about their portion of the problem. During the elicitation reasons for the "ill" nature can be discovered, investigated, and documented. As more knowledge becomes available in time, that understanding and documentation can be updated.

For the nuclear deterrence failure problem, it would be interesting to determine whether any expert has a structure and logic flow in mind for the whole problem. If such organization exists in an expert's thinking, it may be at only a general level, oversimplified, or beyond the expert's subject proficiency. Examples of this in the literature include Bunn's general structure cutting across multiple areas of expertise without eliciting from different experts[46] and Hellman's acknowledgment that his structure, a mathematical model, is not formulated from any expertise and is for illustration purposes (see chapter 8). Instead, it is anticipated that experts may have only structural ideas about their particular subject-matter portion of the whole problem. Different experts can work together to construct the whole problem during a group elicitation. Utilizing the decomposition principle goes a long way toward understanding aspects of an ill-defined problem structure.

Structure in the Knowledge

Whether an expert-supplied or an established structure is used, the data, information, and knowledge used to populate the structure may have internal patterns, association structures, and redundancy or dependency relationships. In other words, the knowledge can have a structure that is worth understanding and using.

Understanding and using any structure in the knowledge is a separate exercise from structuring the problem. Knowledge structuring is more of an analysis activity than an elicitation activity. Nonetheless, experts must work closely with analysts in seeking understanding of the knowledge structure. Neural networks, factor analysis, cluster analysis, statistical covariance, and correlation structures are some commonly used techniques to uncover data structures. Although many of these require large amounts of numerical data, some can still be used for smaller amounts of more general knowledge.

For example, an expert examining the results of a neural network or factor analysis of historical events data might be able explain the data structure found from this analysis by seeing an association or reason that was previously not considered. That reason or association would then be an added feature to the problem structure.

Analysis for structure in the data, information, or knowledge (e.g., historical record) is recommended, when possible, because understanding the data/knowledge structure often provides insights into the problem structure. Even organizing all the available data, knowledge, and information into files, spreadsheets, or perhaps databases reveals problem structure. For the nuclear deterrence problem, it is unlikely that much analysis would be possible because of the sparse amount of data, information, and knowledge available. However, some collection and organization of the applicable data, information, and knowledge will be necessary for simple bookkeeping. This effort can reveal structure in the knowledge, which might, in turn, be useful for considerations about the problem structure. If the structure in the knowledge is inconsistent with the problem structure, the reasons for this conflict should be understood.

Eliciting a Structure

The formal elicitation principles from the first section of this chapter have been applied to eliciting a structure from experts.[47] Eliciting a problem structure is an iterative process; it is common to start, stop, restart, redo, and rework. What follows is a brief description of how to elicit a problem structure and some of the difficulties involved relating to the problem of assessing the risk of deterrence failure.

Elicitation can be done with each expert or with a group of experts. The former is advantageous for understanding how each expert views their portion of the ill-posed problem. The latter is advantageous for the deterrence problem because different experts will be needed for different aspects of the problem. In a group setting, these experts can discuss how their different areas fit together to complete the whole problem structure.

Such interactions often reveal new understandings that cut across different aspects of the problem.

The first step in eliciting problem structure is to ask the expert(s) to simply write down some of the fundamental components, issues, or aspects of the problem. For nuclear deterrence experts, this would include eliciting their areas of expertise and experience. Defining the problem scope—what may or may not be included—also starts here. Usually this first set of items supplied is at a very general level of detail, representing the basic problem features, facets, subject areas, and historical record. For the deterrence problem, these items could include a time frame (past and future), participants involved (states, groups, leaders), sociopolitical perspectives and agendas, technologies available at the time (including communications, manufacturing, transportation, and detection), scenarios or sequences, and intelligence-gathering capability. Many iterations and refinements might be needed just to get the fundamentals listed down on paper, with no particular organization. Using the decomposition principle helps experts clarify their thinking about the problem while drilling down to the level of detail of their knowledge.

The interviewer should continuously record the expert's verbalizations as the expert works and encourage the expert to think out loud. Elicitation probing methods should be used to get the experts to supply reasons behind their thinking. The interviewer may have to encourage experts to think about the unthinkable (e.g., nuclear war), to think beyond their experience (e.g., the use of nuclear weapons), and to go outside their comfort zones, countering anchoring bias.

It may or may not be appropriate to instruct experts to "box" their supplied information. Whether to do so is the experts' choice. It is appropriate to permit experts to separate or group some items even at this early stage. For example, an expert may be recording multiple activities and events that can be organized into different scenarios leading to potential nuclear weapon use.

At any point, the expert may want to begin denoting associations, sequences, relationships, influences, causalities, or dependencies among items recorded on paper. Again, these relationships should be identified and designated in whatever form or format the expert desires. Colors, shapes, lines, arrows, highlighting, using different pages, or cutting and pasting are a few helpful methods. For example, an expert may have listed several socioeconomic and political factors necessary for any state or terrorist group to consider when committing to the acquisition of a nuclear weapon. The expert now wants to distinguish and organize these factors according to which particular state and which particular group.

Connections or associations among items may be difficult to define and characterize because of the uncertainty in their relationships. The difficulties and uncertainties expressed by the expert should be recorded. To aid the expert in these determinations, some common relationships among two generic items, A and B, include:

- **Cause and effect (A causes B).** For example, a 9/11 terrorist-type attack (A) causes Americans to become incensed (B).
- **Dependence (A is conditioned on B).** For example, a country will not impose economic sanctions (A) unless the United Nations agrees (B).
- **Implication (A implies B).** For example, Israel's past policy of preemptive strikes (A) implies it will strike preemptively again (B).
- **Subset (A is included in B).** For example, an attack on a NATO nation (B) is an attack on the United States (A).
- **If–then rule (if A, then B).** For example, if the United States determines who originated the attack (A), then it will retaliate against them (B).
- **Series or intersection (A and B).** For example, the Joint Chiefs will transport troops (A) and send a carrier group (B) to the area.
- **Redundancy or union (A or B).** For example, the Army will either deploy special forces (A) or use drones (B).
- **Correlation (A behaves like B or the opposite of B with or without known causality).** For example, as world economics gets worse (A) the likelihood of attacks (B) increases.
- **Inference (A is inferred by B).** For example, examining the debris and isotopes from a nuclear blast (B) provides evidence to infer its country of origin (A).

During the initial portion of the elicitation for problem structure, the expert should be thinking freely and freely recording aspects, features, and issues of the problem, including the first round of relationships and associations. Any difficulties in formulating or recording these should be noted and completion should be postponed. Likewise, focus on organization or logic flow is not necessary yet and may still be too ill posed. Organization and flow may become clearer as the elicitation progresses.

To distinguish details from general items, an iterative course in the elicitation is helpful. Start with the most general level of detail and then elicit more specific issues, facets, ideas, etc. However, getting specific can quickly burden and complicate the expert's thinking, resulting in inconsistency and in reaching knowledge voids or gaps. An alternative strategy is to stop drilling down in detail and generalize once more. Guide the expert, without fatiguing them, to iterate between thinking about the general to the specific and back again as often as required. The reason for this is to aid the expert in keeping the bigger picture in mind while decomposing the problem into details. For example, the bigger picture might be a particular assumed political environment, affecting the detailed issues, events, and outcomes within it.

Permit the expert to leave holes, blanks, and question marks as placeholders for things not easily characterized or known. These voids can be addressed in a later iteration or after the expert has had a chance to ponder, calculate, or research. Other experts may have to be used to fill in these gaps. Alternatively, these holes, blanks, or questions may never get completed because the knowledge simply does not exist. This lack of knowledge is part of the uncertainty inherent in the problem. The same is true of describing associations. Some may remain vague or ill defined. A simple notation suffices such as "I know A is somehow related or important to B, but I just don't know what that relationship is."

The experts should not try to complete the structure in one elicitation session or even one day. Time between sessions gives the experts a chance to rethink and reorganize, preventing cognitive overload. It is not uncommon for the expert to return to the next elicitation session and completely start over. However, the previous work should not be discarded.

It may be possible to establish some major general features in one session and then develop the structures for each of these in subsequent sessions. The level of detail may not be the same for all features of the problem. Some aspects of the problem may be known in great detail. Others may be listed at only the most general level, with nothing known in detail. For example, the actions of some newly formed terrorist faction would difficult to detail.

An expert may designate some issues, relationships, or portions of the problem for other experts to structure. Bringing in new experts brings in new knowledge, but it can also bring in disagreements about how to structure the problem. Resolution of disagreements between experts takes time; however, it usually provides valuable insights for the interviewer, analyst, and the experts. Some disagreements may not be resolved. Those unresolvable differences reflect the large uncertainty in the state of knowledge for that issue.

Some Difficulties in Eliciting a Structure

A few difficulties involved for ill-posed problems such as the risk of nuclear deterrence failure are described below.

Experts may run into dead ends where their thoughts cannot be depicted because of complexities or lack of knowledge or because they have not thought about how to structure aspects of the problem before. Dead ends are legitimate. There is a difference between forcing experts to supply knowledge that does not exist and asking them to use their expertise beyond their personal experience or comfort zone. The former results in biased, fictitious responses, whereas the latter minimizes anchoring bias. For example, asking experts to consider circumstances according to their knowledge for when a state leader might detonate a nuclear weapon on US soil may be uncomfortable but can be within the expert's capability. Demanding that the experts read the leader's mind is unreasonable.

The unknown or little known details (high uncertainty issues) can hinder thinking and even contribute to cognitive overload. The same is true for poorly understood relationships, such as degrees of association or dependency. For example, an expert may state something like "I just don't know why country A nearly always votes like country B in the United Nations, but it just does."

The expert may have to explore various ways of depicting the problem, which can be frustrating and time consuming. The expert may find it difficult to think aloud or record on paper their thoughts about the structure. These difficulties are not necessarily due to some inability of the expert, but they stem from the complexity, knowledge-poor nature, and high uncertainty inherent in the problem.

The elicited structure may make sense only to a single expert, reflecting their way of thinking about the problem. Their structure is conditioned on the way that expert thinks. That conditioning makes it difficult to combine structures from different experts or to combine substructures of parts of the problem elicited from different experts. After the structuring elicitation(s), it is permissible and often beneficial for experts to see how others view the same problem or parts of it. Facilitated group elicitations can accomplish this as long as bias minimization techniques are used.

Because the elicited structure is personal and expert specific, experts may request that their names be kept anonymous or not associated with specific details. Honoring such requests is part of good formal elicitation practice.

At the end of the elicitations of problem structure, there may be multiple versions from multiple experts. Each may have holes, blanks, and unresolved questions. For the problem of assessing the risk of nuclear deterrence failure, if this is not the result, something went wrong with the elicitations. High uncertainty, especially from lack of knowledge, manifests itself in what appears to be an inconsistent mess (or even a waste of effort) for problem structure. Getting the experts to think deeply and deliberately is necessary to understand and capture the current state of knowledge about the problem—as poor as that current state of knowledge may be.

It is possible that the final expert-supplied structure(s) may not completely specify how all the pieces of the problem go together so that the likelihood and consequence constituents of risk can be assessed. Even with this situation, the risk constituents can be determined conditioned on the fact that pieces are missing or aspects are temporarily removed. A conditional risk assessment is better than no assessment. Those conditions made to assess risk should be noted as the focus for future investigation and understanding when or if the required knowledge becomes available. Only then can an unconditional risk assessment be completed.

Alternatives to Established Approaches

Eliciting problem structures from experts is one of the alternative approaches for problem structuring. Other approaches could prove useful for the risk of nuclear deterrence failure. Some suggestions follow, including enhancements to established methods and new, untested ideas.

Fuzzy Sets and Logic

Since 1965, when Zadeh published his landmark paper,[48] many mathematical and logic-based applications, based on crisp sets and binary logic, have been enhanced by using fuzzy sets and logic. For example, probability-based decision analysis, reliability, and risk analysis have become fuzzy decision analysis, fuzzy reliability, and fuzzy risk analysis. Fuzzy sets and logic accommodate a different type of uncertainty than probability theory does. That uncertainty is called by many names: an imprecision uncertainty, the uncertainty of classification, linguistic uncertainty, and rule-based relationships (e.g., if–then rules) uncertainty. Uncertainty of classification is found in formulating the "boxes" and in determining their connections in problem structuring. Linguistic uncertainty is applicable to the quantification and interpretation of words (e.g., *better*, *not likely*, *maybe*). These uncertainties, along with rule-based relationships, could be prevalent in the risk of the nuclear deterrence failure problem.

While it is doubtful that an expert would have experience with fuzzy sets, their elicited thinking may be conducive to its use. It is the job of the interviewer and/or analyst to make the expert aware of fuzzy constructions when the expert appears to be thinking about the kinds of uncertainties and relationships best handled with Zadeh's fuzzy mathematics.

Structuring methods (including established ones) need not be restricted to binary outcomes (e.g., failure or success) or crisp logic. Actually, human thinking, decisions, and actions tend to follow fuzzy logic better than crisp logic. This is because fuzzy logic permits degrees of performance, likelihood and consequence, and partial decisions and actions. For example, an expert having difficulty deciding how several events are related can describe multiple connections of varying strengths and degrees. Many connections listed in the "Eliciting a Structure" section contain words that lack binary meaning. Fuzzy connections are not restricted to sum to 1.0 as in probabilistic event trees.

Fuzzy structures are potential alternatives to established structures based on crisp logic.[49] Using fuzzified versions of established structures requires more elicitation time because of the different types of uncertainties involved in characterizing degrees of associations and many rules governing those.

It is common practice to characterize the constituents of risk using green, yellow, and red shading to indicate low, medium, and high levels, respectively, for likelihood and consequences, as shown in Figure 3.2. However, this representation actually depicts fuzzy sets for the risk constituents. For instance, the risk denoted by the X has degrees of both yellow and green but is mostly green. Thus, X partially belongs to the yellow (medium) set and more to the green (low) set. The risk at X cannot be precisely assigned to either the low or the medium sets. The same is true of the risk denoted by the asterisk, which has most membership in the red (high) set but some in the yellow (medium) set.

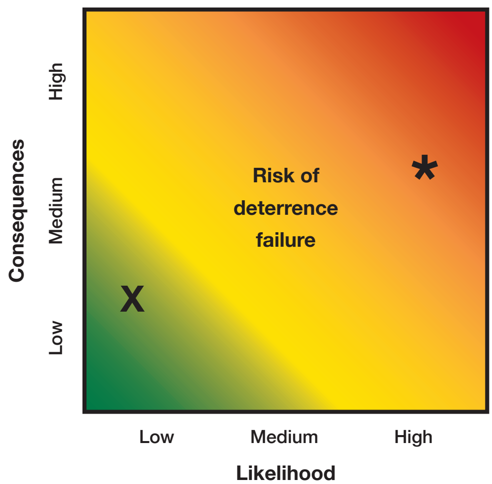

Figure 3.2. Fuzzy shades for the constituents of risk.

Uncertainty Perspective

Recent developments in risk assessment have included the use of possibility theory instead of probability theory because of the type of uncertainty that possibility addresses (which probability does not) and because humans are poor probabilistic thinkers.[50] Established problem structuring approaches (e.g., trees) used in probabilistic risk assessments can be modified for using possibilities instead of probabilities.

Because the deterrence problem has such a high degree of uncertainty attributable to lack of knowledge, one could imagine a structure for the problem based on these uncertainties. The exact form or nature of this uncertainty-perspective structuring is speculation at this point; however, experts could again be called on to determine it. The idea is that experts would be asked to view the problem and its aspects in terms of uncertainties instead of the event/issue perspective. The challenge would be that experts should be comfortable with thinking about uncertainties, and most experts are not.

Regardless of the problem structure used for assessing the risk of deterrence failure, managing the different types of uncertainties will be a challenge.

Information-Gap Structure and Triad Principles

The JASON study correctly concluded that rare events could not be predicted because of the lack of data and the lack of specific information about such events.[51] This conclusion actually is based on a type of uncertainty called nonspecificity.

Nonspecificity is the uncertainty from relying on the general to determine the specific. For example, with so few and so varied kinds of attacks (events), the best an expert could predict for the future would be a rate or average time until the next attack but with no specification about what, where, or how it would happen. As the JASON study concludes, predictive capability suffers from this uncertainty. This conclusion is one of three important principles composing the triad.[52]

The triad involves three dependent concepts: predictability, robustness to uncertainty, and fidelity of models or theory to data. Simply stated, for any given problem with an information-gap structure, all three cannot be simultaneously optimized.[53] Trade-offs or sacrifices must be made for one or two to improve the third.

The information-gap structure is a decision-making and an uncertainty-structuring approach. It is general enough to permit the use of general information theories (including probability) to characterize the different types of uncertainties. The information-gap approach focuses on the relationships in the triad, and their trade-offs could prove useful for structuring the conclusions and results of a risk assessment for presentation to decision- and policy-makers.

Structuring Knowledge Sources

Research is currently in progress focusing on the uncertainty involved in structuring the data, knowledge, and information available for populating a problem structure.[54] For knowledge-poor problems, additional knowledge is sought from similar and relevant problems. Using other knowledge sources induces additional uncertainty based on how close these other sources are to the problem of interest. For example, an important issue in the deterrence problem is assessing close calls. Another knowledge source that could be useful for understanding nuclear close calls would be to understand close calls in historical military attacks (see chapter 2). Knowledge source structuring can be considered a knowledge-integration-structuring approach and is described in chapter 8.

Assessing Risk with Expert Knowledge

Regardless of the structuring approach for the problem or for the knowledge, the high uncertainty and knowledge-poor nature of the risk of deterrence failure problem necessitates eliciting knowledge from multiple subject-area experts. There is a long history of using expert knowledge in risk assessment. Perhaps the best-known and earliest use of expert knowledge in data-sparse applications was the WASH-1400 study, also known as the reactor safety study, considered the birth of probabilistic risk assessment.[55] Experts contributed their knowledge and expertise for this study but did so without formal elicitation. To remedy this, the US Nuclear Regulatory Commission sponsored research into the development of formal elicitation and analysis techniques. WASH-1400 was replaced in 1991 with NUREG-1150, which used formal elicitation.[56]

Evaluating Likelihood

As shown in Figure 3.2, the first constituent of risk is determining the likelihood, which is often (perhaps too often) expressed as a probability. When data, history, or models are available, they can produce estimates for likelihood or probability. When they are lacking, expert knowledge becomes the source for estimating likelihood. Recall, it is best to avoid asking experts for probabilities, especially if they are not accustomed to thinking in those terms. However, it is often reasonable to elicit general likelihoods.

Likelihoods are estimated or elicited for the many issues, items, and parts of the problem in a risk assessment. Hence, they are common to the entire problem structure. The problem structure specifies how these likelihoods are to be combined together, resulting in the overall likelihood required for calculating risk. Uncertainties attached to these likelihoods must also be combined through the structure. Again, the source of these uncertainties may be solely from the experts' experience and knowledge.

Evaluating Consequences

The second constituent of risk is determining the consequences, as shown in Figure 3.2. A common form, quantity, or standard of these is less obvious because consequences stem from different subject areas: loss of life, damage to property, cost, time, and perception. A utility or utility function is often formulated to transform these different consequences to a common scale or measure of value or worth.[57] Sometimes a dollar value is used as a common measure of utility.

Consequences of deterrence failure are particularly devastating—nuclear weapons exchange or nuclear war. While these are difficult to evaluate and estimate, comparative techniques, such as Saaty's Analytic Hierarchy Process, and formal elicitation techniques aid the expert in thinking about the unthinkable.[58]

Summary

Assessing the risk of nuclear deterrence failure is a complex problem covering multiple subject areas. Common to these subject areas are sparse or lacking data, lacking theory or models, high uncertainty, and involvement of human behaviors and decisions. Because of these difficulties, analysts must rely on the use of experts and formally elicited expert knowledge. Established problem structuring and framework methods (e.g., logic or block diagrams) may not be appropriate and may be inconsistent with the way experts think about the problem or their portions of it.

An alternative approach for structuring, framing, and/or organizing the ill-posed deterrence problem is to elicit the structure from the experts. The same formal elicitation techniques briefly described in the first part of this chapter also apply to eliciting problem structure described in the second section. These bias minimization techniques help ensure that the knowledge gathered is of the best quality.

Qualitative or quantitative knowledge can be accommodated, permitting some analysis and drawing of conclusions. Elicited uncertainties can also be qualitative or quantitative. There are theories that characterize these various kinds of uncertainties consistent with experts' thinking; however, these theories present some analytic difficulties when used together.

For the challenging problem of assessing the risk of deterrence failure, an analyst should rely on an expert-oriented structuring of the problem and should use all available sources of data, knowledge, and information. The integration approach necessary to analyze such a structured problem and to draw conclusions is discussed in chapter 8.

In summary, assessing the risk of nuclear deterrence failure relies on the existing state of knowledge of the experts in its subject areas. Eliciting that knowledge with established formalism for minimizing biases is feasible, as outlined in this chapter. What is described is an expert-oriented, expert-driven methodology. Because knowledge is constantly evolving, it is necessary to periodically elicit experts to update how their understanding and cognitive processing has changed with new information and knowledge. For this updating, it is vital to retain all material gathered in all the elicitation sessions.

Acknowledgments: Special thanks go to James Scouras for the many hours of review, discussion, and comments devoted to improving this chapter during its development. His insights and knowledge about nuclear deterrence were a valuable contribution and are much appreciated. Additional thanks go to Andrew Bennett, Edward Toton, and Mary Meyer for their constructive thoughts, knowledge, and discussion in the planning and revising of this chapter.

Notes

1. Mary A. Meyer and Jane M. Booker, *Eliciting and Analyzing Expert Judgment: A Practical Guide* (Philadelphia: Society on Industrial and Applied Mathematics, 2001), 429.
2. Bilal M. Ayyub, *Elicitation of Expert Opinions for Uncertainty and Risks* (Boca Raton, FL: CRC Press, 2001); and R. M. Cooke, *Experts in Uncertainty—Opinion and Subjective Probability in Science* (New York: Oxford University Press, 1991).
3. Meyer and Booker, *Eliciting and Analyzing Expert Judgment.*
4. Meyer and Booker, *Eliciting and Analyzing Expert Judgment.*
5. Jane M. Booker, *Guidelines for Formal Elicitation and Analysis of Expert Knowledge*, 2009, available from the author.
6. Meyer and Booker, *Eliciting and Analyzing Expert Judgment.*
7. Robin L. Dillon and Catherine H. Tinsley, “How Near-Misses Influence Decision Making Under Risk: A Missed Opportunity for Learning,” *Management Science* 54, no. 8 (2008): 1425–1440, https://doi.org/10.1287/mnsc.1080.0869.
8. Peter Vincent Pry, *War Scare: Russia and America on the Nuclear Brink* (Westport, CT: Praeger Publishers, 1999).

9. Scott D. Sagan and Kenneth N. Waltz, *The Spread of Nuclear Weapons: An Enduring Debate* (New York: W. W. Norton, 2002).
10. Pry, *War Scare*.
11. Stanley Le Baron Payne, *The Art of Asking Questions: Studies in Public Opinion, 3* (Princeton, NJ: Princeton University Press, 1951).
12. Meyer and Booker, "Selecting the Question Areas and Questions," chap. 4 in *Eliciting and Analyzing Expert Judgment*; and Meyer and Booker, "Refining Questions," chap. 5 in *Eliciting and Analyzing Expert Judgment*.
13. Cliff Joslyn and Jane M. Booker, "Generalized Information Theory for Engineering Modeling and Simulation," chap. 9 in *Engineering Design Reliability Handbook*, ed. Efstratios Nikolaidis, Dan M. Ghiocel, and Suren Singhal (Boca Raton, FL: CRC Press, 2004).
14. L. A. Zadeh, "Fuzzy Sets," *Information and Control* 8, no. 3 (1965): 338–353, https://doi.org/10.1016/S0019-9958(65)90241-X.
15. Nozer D. Singpurwalla and Jane M. Booker, "Membership Functions and Probability Measures of Fuzzy Sets," *Journal of the American Statistical Association* 99, no. 467 (2004): 867–877, https://doi.org/10.1198/016214504000001196.
16. Sherman Kent, "Words of Estimative Probability," *Studies in Intelligence* 8, no. 4 (1964): 49–65.
17. Charles Weiss, "Communicating Uncertainty in Intelligence and Other Professions," *International Journal of Intelligence and Counterintelligence* 21, no. 1 (2007): 57–85, https://catalog.archives.gov/id/7282770.
18. Meyer and Booker, *Eliciting and Analyzing Expert Judgment*.
19. Meyer and Booker, *Eliciting and Analyzing Expert Judgment*.
20. Thomas R. Bement, Jane M. Booker, Sallie Keller-McNulty, and Nozer D. Singpurwalla, "Testing the Untestable: Reliability in the 21st Century," *IEEE Transactions in Reliability* 52, no. 1 (2003): 118–124, https://doi.org/10.1109/TR.2002.807239.
21. Bruno de Finetti, *Theory of Probability*, vol. I and II, trans. A. Machi and A. F. M. Smith (New York: John Wiley & Sons, 1974).
22. Jonathan Baron, *Thinking and Deciding*, 3rd ed. (New York: Cambridge University Press, 2000), 102.
23. T. J. Ross, Jane M. Booker, James R. Langenbrunner, and F. M. Hemez, "The Quantification of Inference Uncertainty in the Absence of Physical Tests," LA-UR-08-1705, presented at the Los Alamos National Laboratory Risk Symposium, Santa Fe, NM, March 11–13, 2008.
24. Meyer and Booker, *Eliciting and Analyzing Expert Judgment*.
25. Kent, "Words of Estimative Probability."
26. Richard G. Lugar, *The Lugar Survey on Proliferation Threats and Responses* (Washington, DC: US Senate, 2005), https://irp.fas.org/threat/lugar_survey.pdf.

27. Matthew Bunn, "A Mathematical Model of the Risk of Nuclear Terrorism," *Annals of the American Academy of Political and Social Science* 607, no. 1 (2006): 103–120, https://doi.org/10.1177/0002716206290182.

28. Martin E. Hellman, "Risk Analysis of Nuclear Deterrence," *The Bent of Tau Beta Pi* 99, no. 2 (2008): 14–22, https://ee.stanford.edu/~hellman/publications/74.pdf.

29. John Mueller, "The Atomic Terrorist: Assessing the Likelihood," presented at the Program on International Security Policy, Chicago, IL, January 15, 2008.

30. Jane M. Booker and Mary A. Meyer, "Sources and Effects of Interexpert Correlation: An Empirical Study," *IEEE Transactions on Systems, Man, and Cybernetics* 18, no. 1 (1988): 135–142, https://doi.org/10.1109/21.87061.

31. Charles Perrow, *Normal Accidents: Living with High-Risk Technologies* (Princeton, NJ: Princeton University Press, 1999).

32. Scott D. Sagan, "Learning from Normal Accidents," *Organization & Environment* 17, no. 1 (2004): 15–19, https://doi.org/10.1177/1086026603262029.

33. Nassim N. Taleb, *The Black Swan: The Impact of the Highly Improbable* (New York: Random House, 2007).

34. Nozer D. Singpurwalla, "Appendix C. Network Survivability and Borel's Paradox," in *Reliability and Risk* (West Sussex, UK: Wiley & Sons, 2006), 345–348.

35. Maximum entropy can be interpreted as the solution that yields the most (maximum) uncertainty.

36. Roger M. Cooke, *Experts in Uncertainty—Opinion and Subjective Probability in Science* (New York: Oxford University Press, 1991).

37. The mean was calculated by multiplying the midpoints of the bins in Figure 3.1 by the bin counts. If the original seventy-nine values were available, the mean would be the average of those prior to binning. The median of seventy-nine values is the fortieth ordered value, which falls in the 20–29 percent bin. The mode is the bin with the largest count, even though the first and last bins are not the same width as the other bins. Without the raw data and given the unequal bin sizes used, there is no other way to determine the mode.

38. Scott D. Sagan and Kenneth N. Waltz, *The Spread of Nuclear Weapons: A Debate* (New York: W. W. Norton, 1995); and Scott D. Sagan and Kenneth N. Waltz, *The Spread of Nuclear Weapons: An Enduring Debate* (New York: W. W. Norton, 2002).

39. Edward R. Tufte, *The Visual Display of Quantitative Information* (Cheshire, CT: Graphics Press, 1982, first ed.; newer editions available).

40. Harry F. Martz and Ray A. Waller, *Bayesian Reliability Analysis* (New York: Wiley & Sons, 1982).

41. Hiromitsu Kumamoto and Ernest J. Henley, *Probabilistic Risk Assessment and Management for Engineers and Scientists*, 2nd ed. (New York: IEEE Press, 1996).

42. D. G. Malcolm, J. H. Roseboom, C. E. Clark, and W. Fazar, "Application of a Technique for Research and Development Program Evaluation," *Operations Research* 7, no. 5 (1959): 646–669, https://doi.org/10.1287/opre.7.5.646.

43. An example of a hand-drawn decision tree can be found at https://commons.wikimedia.org/wiki/File:Manual_decision_tree.jpg.
44. Ronald A. Howard and James E. Matheson, eds., "Influence Diagrams," *Readings on the Principles and Applications of Decision Analysis*, vol. II (Menlo Park, CA: Strategic Decisions Group, 1984).
45. John von Neumann and Oskar Morgenstern, *Theory of Games and Economic Behavior* (Princeton, NJ: Princeton University Press, 1944).
46. Bunn, "A Mathematical Model of the Risk of Nuclear Terrorism."
47. Jane M. Booker and Laura A. McNamara, "Expert Knowledge in Reliability Characterization: A Rigorous Approach to Eliciting, Documenting, and Analyzing Expert Knowledge," chap. 13 in *Engineering Design Reliability Handbook*.
48. Zadeh, "Fuzzy Sets."
49. James J. Buckley, Thomas Feuring, and Yoichi Hayashi, "Fuzzy Hierarchical Analysis Revisited," *European Journal of Operations Research* 129, no. 1 (2001): 48–64, https://doi.org/10.1016/S0377-2217(99)00405-1.
50. John L. Darby, *Estimating Terrorist Risk with Possibility Theory*, Report No. LA-14179 (Los Alamos, NM: Los Alamos National Laboratory, 2004), https://www.osti.gov/servlets/purl/836683; and Didier Dubois and Henri Prade, *Possibility Theory* (New York: Plenum Press, 1988).
51. *Rare Events*, JASON Report JSR-09-108 (McLean, VA: MITRE Corporation, October 2009), https://irp.fas.org/agency/dod/jason/rare.pdf.
52. François M. Hemez and Yakov Ben-Haim, "The Good, the Bad, and the Ugly of Predictive Science," presented at the 4th International Conference on Sensitivity of Model Output, Santa Fe, NM, March 8–11, 2004.
53. Yakov Ben-Haim, *Info-Gap Decision Theory: Decisions under Severe Uncertainty* (Oxford and Amsterdam: Academic Press, 2001).
54. James R. Langenbrunner, Jane M. Booker, François M. Hemez, and Timothy J. Ross, "Inference Uncertainty Quantification Instead of Full-Scale Testing," presented at the 10th Non-Deterministic Approaches Conference, Schaumburg, IL, April 7–10, 2008.
55. *Reactor Safety Study: An Assessment of Accident Risks in U.S. Commercial Nuclear Power Plants*, WASH-1400 (NUREG 75/014) (Washington, DC: US Nuclear Regulatory Commission, 1975).
56. *Severe Accident Risks: An Assessment for Five U.S. Nuclear Power Plants*, NUREG-1150 (Washington, DC: US Nuclear Regulatory Commission, 1991), https://www.nrc.gov/reading-rm/doc-collections/nuregs/staff/sr1150/index.html.
57. Paul Anand, *Foundations of Rational Choice Under Risk* (Oxford: Oxford University Press, 1993).
58. Thomas L. Saaty, *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation* (New York: McGraw-Hill, 1980).