14 Characterizing Uncertainty through Expert Elicitation

14.1 INTRODUCTION

How does a risk assessor address the consequence and probability of a risk when the desired knowledge and data are nonexistent or unavailable? The elicitation of scientific and technical judgments from experts, in the form of subjective probability distributions, can be a valuable addition to the evidence base used to support public policy decision making (Morgan 2014). The process of eliciting information from experts, especially in subjective probability distributions, is the subject of this chapter.

Risk analysis is evidence-based decision making that is intentional in its handling of the instrumental uncertainties that could affect decision making. Uncertainty is inevitable. Holes in our data and gaps in our understanding are among the most common causes of a lack of knowledge about a system of interest. Expert elicitation can be a useful way to characterize the uncertainty that results from a lack of knowledge or data.

Expert judgment is not data. Expert elicitation creates neither knowledge nor data. Expert elicitation does not remove uncertainty, it quantifies and characterizes it. Expert elicitation does not fill holes in our knowledge base, it, figuratively, characterizes the uncertainty about the size, shape, and depth of that hole. In so doing, it enables us to be intentional in the ways that we handle instrumental uncertainties in our risk assessments and our decision making. Expert elicitation combines subjective judgment with available knowledge and evidence to produce a type of information that focuses on the likelihood of the nature of an unknown quantity, event, or relationship. When this uncertainty is expressed as a subjective probability distribution it can be especially useful.

Risk managers frequently must ask questions which science alone is incapable of answering. For example, what will happen to accident rates when autonomous vehicles comprise a significant part of the traffic? How will sea level change affect the fortunes of a particular East coast port in this century? A subjective or degree-of-belief approach to probability is useful for many of these situations. We all have gotten quite used to making informal probabilistic judgments about the uncertainty in our own lives with this degree-of-belief approach and we do this with surprising ease and regularity. You don't take an umbrella if you believe it will not rain hard enough to need one. You cross the street if you believe you will not get hit. When you handicap your favorite team's chances to win their next game or the incumbent party's likelihood to retain the White House, you are addressing uncertainty through the subjective assessment of probability. Over time you learn where and when your instincts are more or less good, and you make decisions accordingly.

INFORMAL ELICITATIONS

Expert elicitation can be divided into informal and formal elicitations. Informal elicitations, which are not the focus of this chapter, include self assessment (you decide the values), brainstorming, causal elicitation without structured efforts to control biases, and discussion among peers and colleagues (EPA 1997). These informal techniques are, perhaps, the most common means of filling gaps in our knowledge for routine decision making. These techniques would be best suited for noninstrumental relevant quantities. The formal elicitations of this chapter are generally reserved for the most critical variables in high visibility and controversial analyses, that is, instrumental quantities that can affect decision making.

It is an altogether different situation, however, when the uncertainty we are assessing can have significant implications for others. It is one thing to say I do not think I need an umbrella and another to say we do not think sea level change will affect the fortunes of a specific port this century. In some of these situations, it is important to be more systematic and more careful about how we estimate or characterize uncertain quantities.

This chapter starts by distinguishing personal and professional opinions from expert judgment before considering the circumstances under which one might undertake an expert elicitation. Several types of quantitative descriptions of different uncertain values are presented before subjective probability distributions are discussed. Some distinct challenges presented by subjective probability distributions are considered before we address the elicitation protocol. This is followed by examples of elicitations of subjective probability distributions.

Subjective probability formation, as well as many other forms of decision making, are subject to reliance on a number of heuristics and biases. The more common of these are presented in a discussion of making judgments under uncertainty. The matter of calibrating experts is considered just before the chapter concludes by discussing some issues related to how to handle multiple experts in an expert elicitation.

14.2 PERSONAL OPINION, PROFESSIONAL OPINION, AND EXPERT JUDGMENT

Let us begin with some language. An opinion is a view or judgment formed about something, not necessarily based on fact or knowledge. "Personal" is of or concerning one's private life, relationships, and emotions rather than matters connected with one's public or professional career. Therefore, a personal opinion is a view or judgment formed about something based on one's private experiences. A professional is a person engaged or qualified in a profession. Therefore, a professional opinion is a view or judgment formed about something by a person engaged in a profession that is not necessarily based on fact or knowledge. Ideally, we would hope that professional opinions are based on facts and knowledge. An expert is a person who has a comprehensive and authoritative knowledge of or skill in a particular area. Judgment is the ability to make considered decisions or come to sensible conclusions. Therefore, an expert judgment is a considered decision or sensible conclusion reached by a person with comprehensive and authoritative knowledge in a particular area. Expert judgment shall be considered preferable to professional opinion which is preferable to personal opinion, while recognizing that all three can be valuable and any of them can be flawed. Relevant uncertainties can be instrumental or noninstrumental, and they can be quantified for risk assessments in any of these three ways.

Now let us make the distinctions concrete. If I estimate the price of a cubic yard of concrete to be a uniform distribution between $75 and $200, I am offering a personal opinion. I am not qualified as an estimator. If a cost engineer estimates the price to be normally distributed with a mean of $108 and a standard deviation of $11, that is a professional opinion. If the producer of the concrete says it will cost $117, that is an expert judgment.

In a risk assessment filled with relevant uncertainties there may be room for all three methods for characterizing uncertainty. Relevant but noninstrumental uncertainties may be estimated by personal opinion in some cases without doing any harm to decision making. Instrumental uncertainties should be characterized by professional opinion at a minimum. Expert judgment will be preferred if the decision or its outcomes are significant.

The vast majority of non-instrumental uncertainties can probably be adequately addressed by thoughtful and self-expressed personal and professional opinions supplemented by expert judgment. In a normal risk assessment, many uncertain values will be described by probability distributions for which the assessors estimate the defining parameters. Such uncertainties occur routinely, and they are treated routinely. The methods of the previous chapter represent a "high end" example of self-expressed opinions and judgments.

As part of the evaluation of an ecosystem restoration project downstream of Tenkiller Dam in Oklahoma, our team developed habitat evaluation procedure (HEP) models (U.S. Fish and Wildlife Service 1980) for catfish, bass, and trout. The models consisted of hundreds of variables about which there was uncertainty, either natural variability, knowledge uncertainty, or both. The model outputs, habitat units, were only one of the decision criteria for this project. Most variables in the HEP models were described by minimum, most likely, and maximum value estimates provided by two local resource agency employees. There was no formal elicitation process, professional judgment was sufficient. They were more than competent to answer questions like, "What is the minimum water temperature in this stretch of the river during the summer months?" In addition, the HEP models were not terribly sensitive to minor changes in the input variables. From a practical standpoint there were no real options for more sophisticated analyses as the budget for the work was quite small.

You do not need to complete a formal elicitation process for every uncertain variable in your risk assessment for which you lack sufficient data to choose a distribution in the manner described in the previous chapter. There will be times, however, when experts will be needed to characterize instrumental uncertainties in high profile or controversial risk assessments. In these situations, expert judgment will be needed and a formal process for harvesting that expert judgment will be warranted. That is when an expert elicitation may be called for.

So, when do you need to do an elicitation? The more the following list tends to describe your decision problem, the better served you will be by a formal elicitation process.

- You lack data and knowledge about instrumental quantities, that is, those that could affect the decision made or the outcomes of a decision.
- Your problem is complex and highly visible.
- Trust in your analytical work is an issue.
- Your risk assessment will provide the basis for public decision making.
- Your risk assessment will provide the basis for a formal rule or regulation affecting industry, nongovernmental organizations (NGOs), or others.
- Your organization has stewardship responsibility for some public value, for example, public health, public safety, water resources, homeland security, and the like.
- There are values in conflict for your issue.
- Your issue is a controversial one.
- There are many stakeholders with diverse views.
- Your issue is subject to intense media or other scrutiny.
- One or more stakeholders are likely to challenge your decision in a court or administrative setting.
- Human life, health, or safety may be affected by your decision.
- There is a clearly critical uncertain variable or two in your risk assessment model.

The ecosystem restoration project mentioned previously met none of these criteria and could be handled through the available sparse data and the professional opinions of those responsible for the analysis. As the importance of the decision grows, as indicated by the preceding list, so does the need for a more formal and credible process for estimating uncertain values in our models.

14.3 WHAT KINDS OF UNCERTAIN VALUES?

Scenarios, models, and quantities can be uncertain. The process is easier to understand if we focus on quantities and, so, we will. Expert elicitation is to be used when there is substantial uncertainty regarding the true values of a quantity. Risk managers often need to make decisions that address complex problems for which there is a lack of direct empirical evidence. Expert elicitation is a means of obtaining and synthesizing expert opinion when data are unavailable or not easily obtainable.

Expert knowledge, experience, and insight are especially useful when the quantities of interest are poorly understood because they are new, rare, or complex; when "hard" data are limited or unavailable, data collection is expensive, or formal modeling is infeasible (Ayyub 2001a,b; Meyer and Booker 2001). Morgan (2014) asserts that the elicitation of scientific and technical judgments from experts, in the form of subjective probability distributions, can be a valuable addition to other forms of evidence in support of public policy decision making, especially when our decision problems require judgments that go beyond well-established knowledge. Expert elicitation can make a valuable contribution to informed decision making. It can provide insights into the nature of instrumental uncertainty that are not available from any other source.

> The U.S. Army Corps of Engineers is relying on novel technologies to reduce the probability that Asian carp would reach Lake Michigan if a project is built at Brandon Road Lock and Dam. Little data exist on the performance of these technologies and the resulting uncertainty could affect the decision that is made or the outcome of a decision once made. This is an ideal situation for an expert elicitation.

Quantitative descriptions include: point estimates, paired comparisons, single-event probabilities, distributions of continuous uncertain quantities, conditionalization and dependence (Cooke and Goossens 1999), and parameters of distributions. Each is considered in turn below.

Assessments that result in point estimates are of limited, if not questionable, value because they give no indication of uncertainty, which, of course, is the entire point of characterizing uncertainty. Expert elicitations that produce point estimates can have the illusion of knowledge or fact, especially to decision makers and others outside the process. It is essential to have some characterization of the uncertainty in such instances.

Paired comparisons have been used when experts are asked to arrange a set of items in ordinal sequence according to some criterion like hazard potential, effectiveness, likelihood, and the like. The work can be tedious, comparing 20 items requires a total of 190 pairwise comparisons. These comparisons fail to produce a characterization of uncertainty, although methods for evaluating the degree of expert agreement and consistency are available (Cooke and Goossens 1999).

Single-event or discrete event probabilities are used to assess the probability that an uncertain event occurs or does not occur. The quantitative estimate takes the form of a point estimate, an interval, or a continuous distribution in the [0,1] interval. Estimating a probability using a probability distribution can be challenging. Examples include the probability a stock price rises today, a specific person contracts cancer, an earthquake occurs this year, it rains tomorrow, an aquatic nuisance species becomes established in a new watershed, and the like.

Distributions of uncertain quantities are used in problems concerned with the value that will be realized by an uncertain quantity that varies across a continuous range. This uncertainty is described by subjective probability distributions. Imagine that an expert is asked to characterize the uncertainty about the amount of oil that could be released in an oil spill, call this quantity X. The expert offers his subjective distribution over the possible values of X. The elicitation may take a number of forms; the expert may specify his cumulative distribution function (CDF), his probability density function (PDF), or he may estimate selected parameters of a distribution like the mean and standard deviation of a normal distribution, or selected quantiles of his distribution like the 5%, 50%, and 95% quantiles.

One must make known the background information upon which the uncertainty descriptions are to be based. Every distribution is conditional on some set of assumptions and some body of evidence. Different experts may offer distributions based on different sets of assumed conditions that may render the various estimates too disparate to be compared. Establishing this background information is an important function of the elicitation protocol. If a distribution's values are dependent on other variables whose values are not preidentified and specified, the subjectively identified distributions will be of limited value.

There may be times when there is general agreement on the shape of the probability distribution to be used and the primary task is to simply estimate the parameters of the distribution. These parameters can be estimated as points, intervals, or distributions. The latter two options lead to second degree distributions, with uncertain parameters. First order distributions are well-served by point estimates of the parameters. For a normal distribution the task is to estimate the mean and standard deviation. It is also common practice to estimate quantile parameters for a distribution, for example by estimating the 5%, 50%, and 95% cumulative probabilities for the chosen distribution. The distributions yielded by these methods are subjective probability distributions, the subject of the next section.

14.4 SUBJECTIVE PROBABILITY DISTRIBUTIONS

The point of expert elicitation is to capture an expert's characterization of the uncertainty about some event or quantity. That quantity may be a probability of a single event or the frequency of an event in a population, or it may be any nonprobability value. A probability distribution can be constructed from p-x pairs, where p is a cumulative probability and x is a realized value of a random variable X. When X is, itself, a probability the process can be challenging.

In an expert elicitation, probability is viewed as a statement of an individual's belief, informed by all the formal and informal evidence that they have available (Morgan 2014). Probability, the p in the p-x pairs, is always interpreted as personal or subjective probability in expert elicitation, that is, it is a measure of a person's degree of belief in something. For example, if an expert says there is a probability of 0.3 that if an oil spill occurs at a specific location it will be less than 10,000 gallons, this expresses the expert's personal degree of belief that a spill in that range will occur. Subjective probabilities differ from one expert to the next because experts all have different extents of factual knowledge and generic experience in their domains of expertise.

These subjective probabilities are not preformed numbers lying hidden in our brain cells waiting to be revealed to the world. They are values that must be carefully constructed when needed. According to the subjectivist view, the probability of an event, such as $P(X) > x$, is a measure of a person's degree of belief that it will occur. Thus, probability is not a property of the event but a property of the expert's judgment (Morgan and Henrion 1990). Because experts may vary in their judgments, my subjective probability of an event may be different from your subjective probability of the same event. Thus, there are no "correct" subjective probabilities, even for experts. Subjective probabilities may be self-expressed or elicited.

Subjective probabilities are often used in risk assessments. They cannot be arbitrary. To be useful they must be coherent, that is, they must obey the laws and theories of probability such as those laid out in Chapter 12. Lindley et al. (1979) posited that the individual expert has a set of coherent probabilities that can be distorted by the elicitation process, so that process must be carefully constructed. Savage (1954) and DeGroot (1970) have argued that a person who wants to make rational decisions when faced with knowledge uncertainty must act as if they have a coherent set of probabilities. Some, like Gigerenzer and Todd (1999), are more forgiving about coherence of the individual and are willing to live with reasonable adaptive inferences about our real-world environments.

If subjective probabilities are not based on formal and coherent constructions of the theory and may in fact be noncoherent, then what status do these probabilities have? Winkler (1967) said there is no built-in prior distribution there for the taking, and different elicitation techniques may produce different distributions. He goes on to suggest that these different distributions are equally valid, even though research has shown that some forms of questioning can lead to invalid responses.

This all means that the elicitation technique matters because experts' statements of probability are likely to be made in response to the question asked rather than based on preanalyzed and preformed coherent beliefs. The purpose of elicitation is to represent the expert's knowledge and beliefs accurately in the form of a good probability distribution (O'Hagan et al. 2006).

14.5 THE ELICITATION PROTOCOL

People do not have well-formed subjective probabilities in their heads. Asking people to make up numbers out of the blue is not a good idea. There is a great deal of literature that says people are not especially good at estimating probabilities, especially small ones. We tend to be amazingly overconfident in our estimates of uncertain things, and we tend to rely on the heuristics discussed later in this chapter. Using elicitation techniques that avoid reliance on heuristics and biases that force people to think about the numbers and calibrate themselves can be useful. If there are systematic ways we tend to err and our goal is to improve the elicitation process and its results, perhaps we can avoid the errors with well-structured formal protocols for assessing uncertainty and eliciting probabilities.

ROLES IN ELICITATIONS

Decision makers will use the results of the elicitation (the client).

Substantive experts have the knowledge about the uncertain quantities.

Normative experts provide probabilistic training, assess the process, validate results, and provide feedback.

Facilitators are experts in the elicitation process and manage the elicitation.

Source: O'Hagan et al. (2006).

There are numerous ways to configure an elicitation. The number of experts, the elicitation method (individual or group), elicitation instrument (interview or questionnaire), and method of aggregation can all be varied. It is useful to have enough experts to get a good characterization of the uncertainty without a great deal of duplication. Six to 12 experts are considered a reasonable number of experts. O'Hagan et al. (2006) opine that group elicitation is best suited to elicitations for tasks with a definitive and demonstrable correct answer. Otherwise individual elicitations may be best. Face-to-face interviews offer obvious advantages for communication, clarification, and completeness, but questionnaires serve a useful purpose when assembling the experts in one place is not an option. Aggregating experts is considered in Section 14.8.

There are any number of elicitation protocols described in the literature. Three are briefly summarized here. The first is a seven-step method proposed by Clemen and Reilly (2001). The steps are:

1. Background identification of variables requiring expert assessment; review of scientific literature
2. Identification and recruitment of experts
3. Motivating experts—establish rapport with experts and generate enthusiasm for the process
4. Structuring and decomposition—explore the experts' understanding of causal and statistical relationships among the relevant variables
5. Probability assessment training—explain the principles of the assessment, the biases and heuristics, and ways to counter them; provide opportunities for experts to practice making probability assessments
6. Probability elicitation and verification—experts make required assessments under the guidance of a facilitator; assessments checked for consistency and coherence
7. Aggregation of experts' probability distributions—if multiple experts were used, it may be necessary to aggregate their assessments.

The second is the Stanford/SRI Assessment Protocol (Morgan and Henrion 1990), which consists of five phases:

1. *Motivating*: Establish rapport with experts, explain need and process, explore motivational bias
2. *Structuring*: Develop unambiguous definition of quantity to be assessed in a form meaningful to experts
3. *Conditioning*: Get experts to think fundamentally about their judgments while avoiding cognitive biases
4. *Encoding*: Elicit and encode the expert's probabilistic judgments
5. *Verifying*: Test judgments to see if they reflect the expert's beliefs

The third is a five-step model for elicitation developed by O'Hagan et al. (2006):

1. *Background and preparation*: The client identifies variables to be assessed

2. *Identify and recruit experts*: The choice may be obvious or it may require some effort; six criteria for experts:
   a. Tangible evidence of expertise
   b. Reputation
   c. Availability and willingness to participate
   d. Understanding of the general problem area
   e. Impartiality
   f. Lack of economic or personal stake in the potential findings
3. *Motivating and training the experts*: Assure the experts that uncertainty is natural; training should have three parts:
   a. Probability and probability distributions
   b. Introduction to most common judgment heuristics and biases as well as ways to overcome them
   c. Practice elicitations with true answers unlikely to be known by the experts
4. *Structuring and decomposition*: Spend time exploring dependencies and functional relationships that meet agreement by experts; precisely define quantities to be elicited
5. *The elicitation*: An iterative process with three parts:
   a. Elicit specific summaries of expert's distribution
   b. Fit a probability distribution to those summaries
   c. Assess adequacy: if adequate, stop; if not, repeat the process with experts making adjustments.

Elicitations are best in a face-to-face setting that allows interactions between the expert and the facilitator. All protocols have several stages, but the elicitation always comes at or near the end of the protocol. Its success depends on preparation and foundations established in the preceding steps. Each of these protocols is described in greater detail in the reference documents.

14.6 ELICITING SUBJECTIVE PROBABILITY DISTRIBUTIONS

The elicitation step is where the information is developed. The expert's statements of probability are judgments made in response to the facilitator's questions. Once again, for emphasis, they are not preanalyzed and preformed beliefs of the expert. Thus, the matter of how one asks the questions is of some importance. The basic choices are to provide an x-value for a variable X and ask for the corresponding cumulative probability, p; or to provide a p-value and ask for the corresponding x-value.

To better frame the discussion, assume we are interested in estimating the size of a crude oil spill in a waterway that serves several refineries with a subjective probability distribution. A cumulative probability distribution that represents the expert's beliefs about the total number of gallons spilled in an event might look like the hypothetical example in Figure 14.1.

This is a free-form estimate based on an expert's estimation of the quartile values shown in Table 14.1. The curve of Figure 14.1 was constructed by linearly

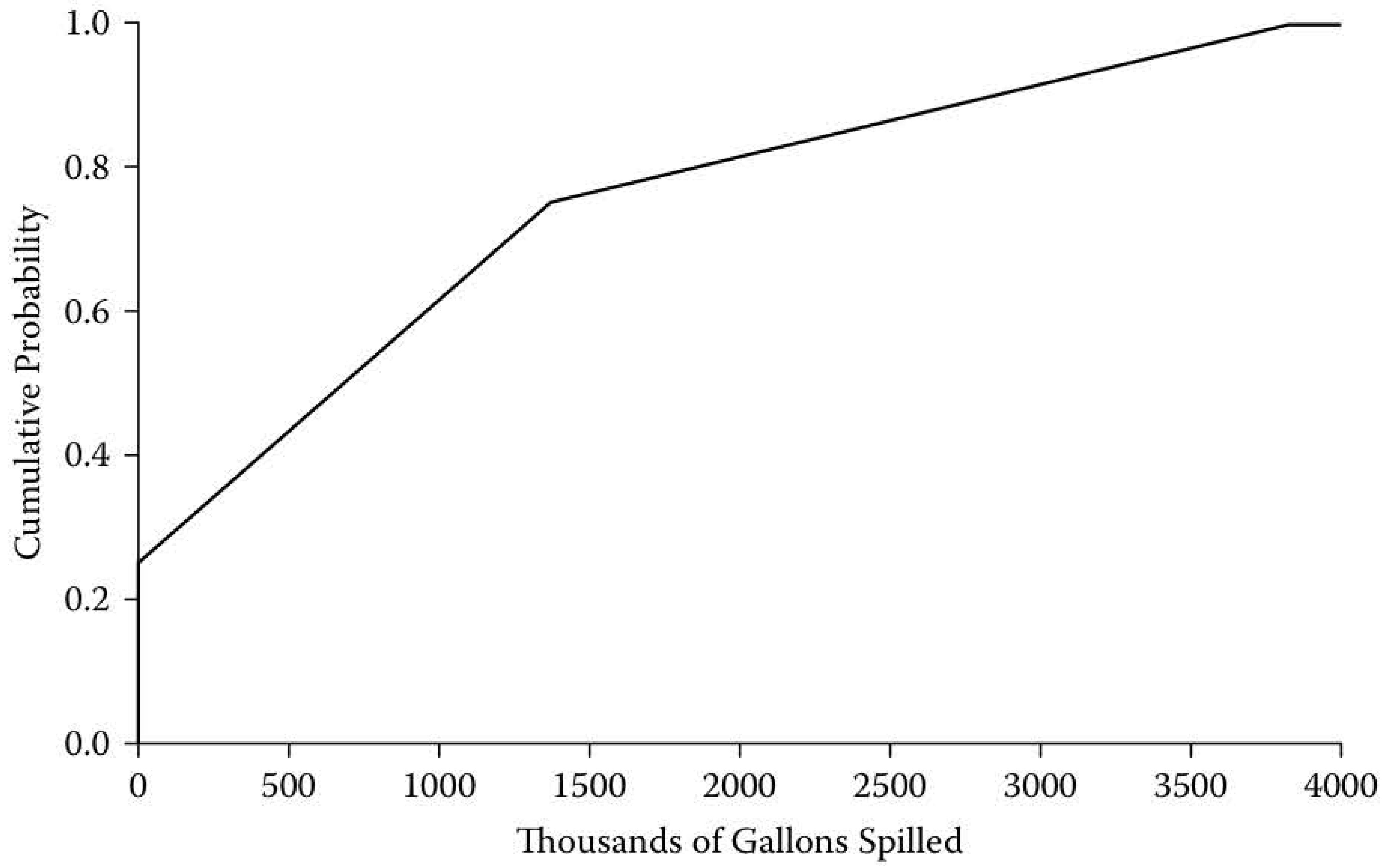

FIGURE 14.1 CDF of the size of an oil spill in the Delaware River.

TABLE 14.1
Oil Spill CDF Elicited from Experts

Cumulative Probability	Thousands of Gallons Spilled
0	0
0.25	5
0.5	668
0.75	1,353
1	3,850

interpolating between the points provided. Note that to define such a distribution, the facilitator can elicit values for x_i, in this case gallons of oil spilled, associated with a given probability (p_i), or the facilitator could provide the x value and ask for associated p values.

An alternative elicitation would be to estimate the uncertain parameters of a specific distribution. This is somewhat similar to the point-estimate elicitation in concept. For this example, imagine it has been determined the size of an oil spill is best represented by an exponential distribution. This assumption imposes a structure on the oil spill problem. The experts are needed to estimate the parameters of the distribution. In this instance, the exponential has only one parameter, the mean. Assuming our experts have estimated this value to be 988,000 gallons, the exponential distribution would be the first-degree distribution shown by the line graph of Figure 14.2. If our experts estimated the mean with a uniform distribution with a minimum of 750,000 gallons and a maximum of 1,500,000 gallons, the second-degree distribution so formed would look like the bar graph of Figure 14.2.

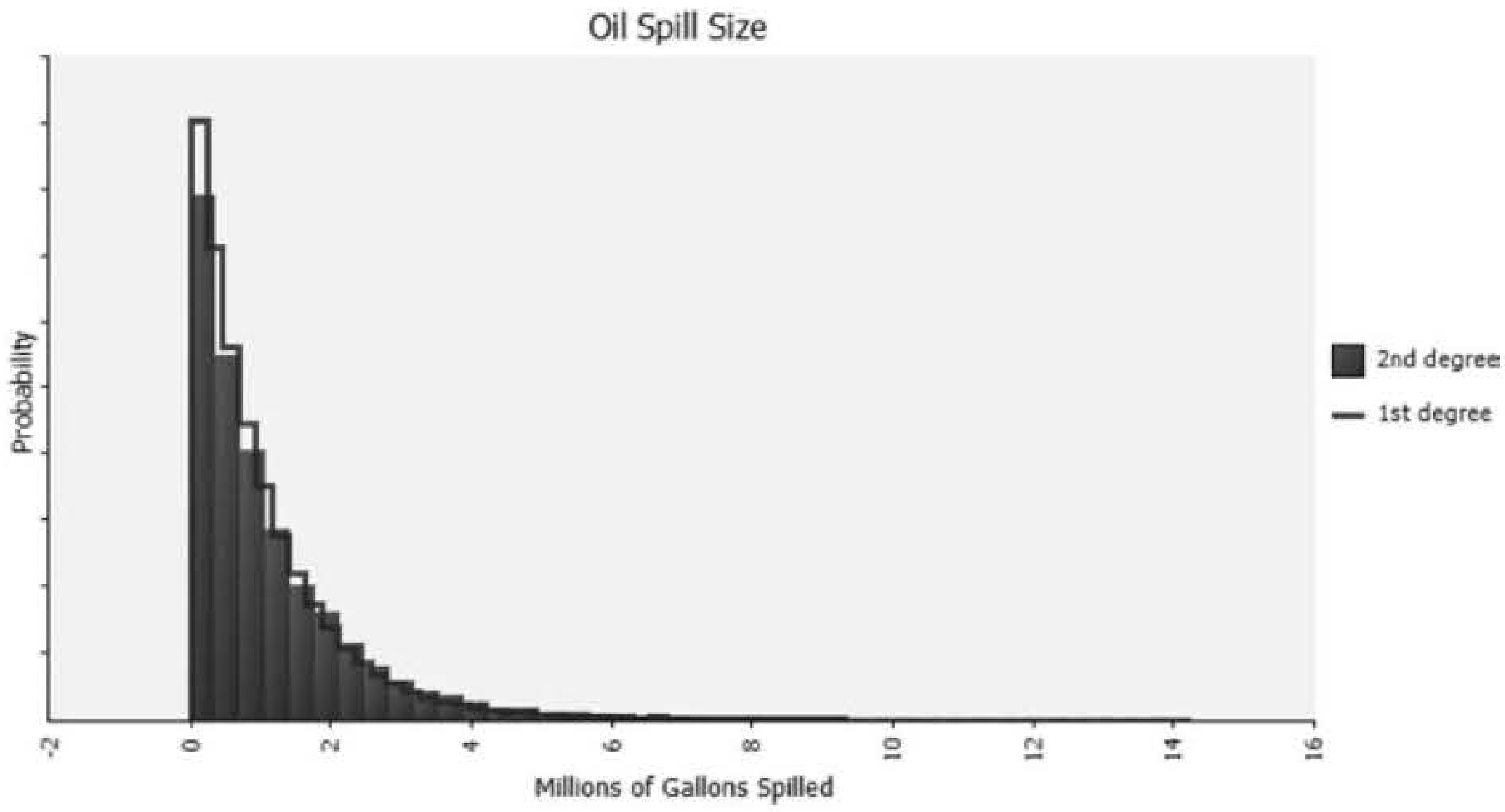

FIGURE 14.2 First and second degree exponential distributions describing the size of an oil spill in the Delaware River.

The p-x pairs of the CDF are a more common way of eliciting information and will be our elicitation focus from now on. That will subsume many of the issues associated with the other elicitations, because it is the more involved process. The size of an oil spill is a continuous random variable. Assuming we are going to elicit the expert's CDF, there are an infinite number of points that could be elicited. The strategy is to elicit information about a relatively few number of individual (p, x) pairs that enable us to reproduce the expert's beliefs in the form of a probability distribution.

The information sought is called a summary. The idea is that a few well-chosen summaries will enable us to identify the expert's beliefs with a high degree of precision. Then, using these summaries, a curve is fitted to them. This can be done by interpolation as shown previously or by fitting a general form function to the points.

The summaries we obtain can, as noted, be based on probabilities or on quantity values. Summaries based on probabilities may be the most common form of elicitation. A basic strategy is to seek specific quantiles that can be used to construct a CDF. For example, one might elicit an x-value for the median value, $p = 0.5$. This divides all possible values in half. The expert may then be asked to find the midpoint of each remaining range, producing the median and then the quartiles. If desired, this process can be extended to the octiles, the 12.5, 37.5, 62.5, and 87.5 percentiles. It is usually difficult for the expert to go beyond these values. Examples of questions that yield the median, first and third quartiles for the oil spill follow:

Can you determine an oil spill size such that any spill is as likely to be larger than this number as smaller than this number?

Suppose we have learned the oil spill is smaller than your median. Can you now determine a new oil spill size (lower quartile) such that it is equally likely that this oil spill is larger than or smaller than this value?

Suppose we have learned the oil spill is larger than your median. Can you now determine a new oil spill size (upper quartile) such that it is equally likely that this oil spill is larger than or smaller than this value?

For contrast, an alternative formulation of the second question might be, "Give an oil spill size such that you think an actual oil spill has a 25% chance of being less than it." There can be many formulations of your specific questions. As long as the answers yield p-x pairs, a distribution can be elicited. However, there are a great many nuances and lessons learned that are documented in a large and growing literature. The O'Hagan et al. text is essential reading for anyone who intends to work in this area. The Ayyub (2001a,b) text is another essential reference for those interested in more details on the topics introduced in this chapter.

Fixed-value methods work a little differently than the quantile method illustrated above. If we are deriving a CDF we might ask, "What is the probability that an oil spill will be 5,000 gallons or less?" Successively larger or smaller spill sizes probabilities can be sought to flesh out the shape of the CDF. My experience, and that of others, suggests that an expert can fall into a biased rhythm of linear interpolation/extrapolation if the facilitator's questions have a pattern to them. For example, asking probabilities of oil spills in 5,000-gallon intervals or even in a monotonically increasing (or decreasing) way can lull an expert into an easy pattern of answering that does not require careful thought about the answers. Varying the pattern of your questions can encourage a more analytical response.

For example, one might ask the probability that a spill is 5,000 gallons or less, followed by 150,000 gallons or less, then 70,000 gallons or less, followed by 1,000,000 gallons or less, and so on in an unpredictable way. The idea is to vary the pattern of the questions to avoid simple interpolation or other biased answering schemes and to require more careful thinking about the probabilities. One could vary the question by asking the probability that a spill is 70,000 gallons or more, instead of less. Such questions can help identify problems with coherence in the expert's assessment of the uncertain quantity. Bounding the CDF, that is, establishing minimum and maximum values when appropriate, is usually going to be a sensitive aspect of the elicitation. Given the anchor-and-adjust heuristic, it is especially important to probe the limits carefully (see box).

BOUNDING A DISTRIBUTION

When eliciting an expert's distribution, it is often necessary to establish a minimum and maximum value for the quantity of interest. Neither data nor experience are likely to reveal absolute extreme values, when logical limits do not exist (0 is a natural limit for many quantities and the [0,1] interval applies to a number of problems).

When experts must estimate these values, it is important to find nonthreatening ways to challenge their estimates. In the oil spill example, one might ask, "If you were gone for a year and returned to learn there had been a spill of X (some value noticeably larger than the expert's maximum estimate) gallons, how could you explain that?" If the expert offers a plausible explanation, offer him the opportunity to amend his estimate.

Probe the limits carefully.

It is also possible to ask questions that approximate a PDF for an unknown variable. In this instance we might ask, "What is the probability that if an oil spill occurs it will be between 10 (or whatever the agreed upon minimum size of an oil spill is) and 5,000 gallons? Greater than 5,000 gallons but less than 25,000 gallons?" and so on. The notion for the PDF elicitation is to seek the probability of a quantity being within a fixed interval. The challenge is to assure that all of those fixed interval probability estimates sums to one.

An alternative approach to elicitation is to ask the expert to draw points for the CDF. This is generally not a terribly realistic way to approach the problem, but it has been used in the past.

Estimates of probabilities can sometimes be better analyzed by decomposing the problem. An oil spill problem can be decomposed by focusing on the vessel's size or the cause of the spill. This can increase the number of elicitations required, or it can entail elicitation of multivariate distributions, which are far more complex. If a problem is to be decomposed, the decomposition can be prescribed by the investigators or by the experts.

To this point, the discussion has emphasized the expert's unique probability distribution. It is worth emphasizing that the elicitation could elicit p-x pairs, rather than parameters, for well know distributions like the beta or Pert distributions. When this is done, some common values sought are: the 5%, 50%, 95% quantiles; the 1%, 50%, 99% quantiles; or the terciles (33%, 67%). The X variable may or may not be bounded by minimum and maximum values.

Each of the elicitation protocols provides for probability training or conditioning of the experts. This is going to be especially important if some of the uncertain values you seek are probabilities that will be estimated by a subjective probability distribution. Consider the question, "What is the probability that Asian Carp will spread beyond Lake Michigan?" (See box.) Can you determine a probability of spread such that the uncertain probability of spread is as likely to be larger than this number as smaller than this number? If that does not make your head feel like it will explode, little will. Great care must be taken in assuring that your experts are comfortable answering such questions.

EXAMPLE ELICITATION QUESTIONS

When thinking about the number of Asian Carp that could arrive at Brandon Road by the year 2021, please rate the following population sizes by their relative probability of arriving (larger numbers indicate a higher probability):

Negligible Population__________
Small Population _____________
Medium Population ___________
Large Population ______________
TOTAL 100 Points

The following questions were answered by having the expert provide minimum, 33rd percentile, 67th percentile and maximum values. Considering the currently existing system that is in place, assume that a large population of Asian Carp have arrived at the Brandon Road Lock and Dam. The number of fish that will make it all the way from downstream of the lock and dam into Lake Michigan in a year will vary from year-to-year, and that number is uncertain. What do you believe is the number of fish that could pass from below the dam to Lake Michigan in a year?

Assume that Asian Carp have a colony in Lake Michigan. What is the probability that Asian Carp will spread beyond Lake Michigan?

Repeated after each answer was the following question: What were the main considerations in shaping your answer?

Think about whether probability values will be sought as decimals, percentages, or odds. Choose a method that suits the expert assessor's cognitive styles, skills, and preferences (Morgan and Henrion 1990). I have worked with experts who did not finish high school (towboat operators) and Ph.D.'s (epidemiologists and engineers). No one-size probability elicitation fits all needs. It is best to establish the preferred or most promising question style in the foundational work that precedes the elicitation, including the practice elicitations.

14.7 MAKING JUDGMENTS UNDER UNCERTAINTY

To develop useful elicitations procedures and to obtain useful probability estimates, it helps to understand how people form judgments about uncertain values. Research suggests that experts and laypersons alike rely on the use of some common heuristics in forming judgments about uncertain quantities (Tversky and Kahneman 1974). Knowing about these shortcuts and their pitfalls is the first and most important step in avoiding them. Elicitation processes that avoid reversion to these heuristics or at least minimize their impact are better than those that do not.

A heuristic is a mental shortcut we use to solve a particular problem. It is a quick, informal, and intuitive algorithm we use to generate an approximate answer to a reasoning question. Heuristics enable us to quickly make sense of a complex environment, but they sometimes fail to correctly assess the world. When they do, heuristics can lead to a cognitive bias that results in an incorrect conclusion.

Our brains are wired in such a way that it is easier to think dramatically than it is to think quantitatively, and especially probabilistically. Probability is not intuitive, as we have seen. It is not easy to think in probabilistic terms. Instead, we tend to use some rather crude cognitive techniques to help us along. The most common of these heuristics and biases are discussed in the following sections.

14.7.1 Overconfidence

Mark Twain once said, "It ain't what you don't know that gets you into trouble. It's what you know for sure that just ain't so." This is the essence of overconfidence; it is the tendency for people to be more confident than the evidence warrants. Overconfidence is often said to be the most pervasive bias people have. When eliciting subjective probabilities, it can lead to estimates that are closer to zero or to one than they should be and to distributions that are narrower than they should be. There is some evidence to suggest overconfidence increases with expertise and knowledge. Many studies of lawyers, doctors, nurses, managers, entrepreneurs, investment bankers, and others have found that they tend to put great trust in their opinions and overestimate their expertise (Riordan 2013).

Calibration exercises can be effective in making subject matter experts aware of the extent of their overconfidence. Decomposing a problem into several component events may help reduce the effects of overconfidence. It is easier to be overconfident about a general issue than it is to be overconfident about a detailed and specific issue. The strategy for offsetting this bias means conducting more specific elicitations. Alternatively, one might ask an expert what sources of information they rely on for their judgments and whether these are fact-based or simply beliefs and opinions.

14.7.2 Availability

Estimates of uncertain quantities are driven by the ease with which previous similar events or instances can be recalled; it can be enhanced by the frequency of events recalled (Hardman 2009). We tend to overestimate the probability of events that come to mind easily and underestimate the probability of events that are not so readily recalled. If a flood just occurred, people tend to overestimate the probability of flooding. If there have been no floods for a few decades, people underestimate the probability of flooding. The more easily examples come to mind, the more probable we think the event is. Similarly, the easier it is to recall instances of a class of events, the larger we estimate that class to be. Most of us use availability instead of the more complex processes we might use to estimate probabilities.

Media coverage affects availability. When low-frequency events receive extensive media coverage—terrorist attacks, levee failure in New Orleans, plane crashes, oil-drilling disasters—it is common to overestimate their probability of occurrence because it is easy to call examples of these things to mind. On the other hand, the probabilities of more common events that do not get the same attention—seasonal flu, automobile accidents, choking on food—are often underestimated.

> Members of a team overestimate their own contributions relative to those of others because their own contributions come easily to mind.

Recent events or events that have affected the expert personally may be emotionally paper-clipped memories that are easier to recall. This can cause a temporary increase in the estimates of the likelihood of these events. If a family member is diagnosed with leukemia, it is more likely that a person will overestimate the probability of this disease. The flipside is also true. If there have been no recent occurrences (of home fires or automobile accidents, for example) or it is a problem that has not touched the expert personally (e.g., identity theft or leukemia), probabilities may be underestimated.

Co-occurrences of events can also distort estimates through an availability bias. I grew up at a time when many family members and neighbors associated space launches and extreme weather events. This connection of events was more illusion than cause and effect. After a while, the hurdle for unusual weather was dropped and the window of opportunity for an occurrence was extended, all of which had the effect of a self-fulfilling prophecy. Eventually people saw more co-occurrences than actually happened. A heavy rain, an extreme temperature within weeks of a launch, or a heavy snowfall in the same year could be associated with a space launch. These events are more strongly associated in the mind than in fact. Beliefs about relationships can affect judgments for all kinds of uncertainties. If availability misleads the expert to misinterpret the strength of a relationship, it can affect probability estimates (Gilovich, Griffin, and Kahneman 2002; Sedlmeier and Betsch 2002). Linked events in a scenario, for example, are easier to imagine than the individual events considered in isolation.

The first step in neutralizing the availability bias is to be consciously aware of it. Try not to look at information chronologically. Look at trends from a variety of angles. Continue to question the extent to which you rely on easily recallable events. Consider a longer timeframe. Investigate the history of like events before and after those that come to mind easily. Consult a wider base of evidence. Question commonly accepted views to build a wider perspective of the quantity you are addressing. Expand your focus and include information in addition to the vivid information that captures your imagination. Constructing a comprehensive body of evidence as part of the elicitation read ahead may help offset the availability bias.

14.7.3 Representativeness

The representativeness heuristic manifests itself in a variety of ways. Tversky and Kahneman (1983) define representativeness as "an assessment of the degree of correspondence between a sample and a population, an instance and a category, an act and an actor, between an outcome and a model." A person following this heuristic evaluates the probability of an uncertain event by the extent to which it is similar in essential properties to its parent population. Consider an event, A. Instead of calculating P(A), those laboring under the representativeness heuristic will try to fit event A into some more familiar class of events and use that probability instead.

Representativeness bias is especially relevant to the estimation of conditional probabilities. Common representativeness biases are summarized in the following sections. The best way to confront and address the representativeness heuristics that follow begins with awareness of the relevant heuristic and proceeds most logically to doing the math. Not every expert will be adept at the relevant math so it may be helpful to have that expertise available on the elicitation team.

Koehler (1996) found that several of the problems associated with representativeness can be improved by eliciting frequencies instead of single-event probabilities. For example, instead of asking the probability for "Linda is a teller and an activist," we might ask, "Of 100 individuals like Linda, how many would be bank tellers and how many would be active in the feminist movement and bank tellers?" (Fiedler 1988). Frequency formats help reduce representativeness bias as compared to probability of single-event question formats. Unfortunately, they also increase availability problems.

14.7.3.1 Conjunction Fallacy

Tversky and Kahneman (1983) offered this oft-repeated example of the conjunction fallacy:

> Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice and also participated in antinuclear demonstrations.
> Which is more probable?
> Linda is a bank teller.
> Linda is a bank teller and is active in the feminist movement.

Most people (usually 80% or more) select the second answer, and this is impossible. How is it possible that Linda is more likely to be a teller (A) and active (B) than that she is just a teller (A). The second answer above is a joint probability, the conjunction of two conditions (A and B). There is no way there can be more people who are tellers and activists than there are people who are just tellers. Surely some tellers are not activists. So, it must be more likely that she is a bank teller.

What is at work here is that the description of Linda is unrepresentative of what most of us think a bank teller would be. This tends to decrease our belief that she is a teller. However, adding the active feminist detail provides a description of Linda that is more representative of an activist. We tend to rely more on the representativeness than we do on the simple facts of probability to gauge the likelihood of the right answer. In other words, we replace a difficult question with an easy question. Instead of calculating a probability we ask does Linda seem like a feminist? One take away from this is do not be misled by highly detailed scenarios. Greater specificity and detail in a scenario lowers its chances of occurring, even if the scenario seems to represent the most probable outcome.

14.7.3.2 Base-Rate Neglect

A base rate is a relative frequency. People will often ignore information about the base rate and estimate probabilities based solely on the information about the event or instance before them. If you have ever been to the doctor and been instructed to have a laboratory test for a rare and serious disease, the chances are good that your emotional response illustrated base-rate neglect.

The simple fact is, absent other specific information, it is highly unlikely you have the rare disease. When the doctor asks you to have the test, you ignore this fact and focus on the fact that they have asked you to get this test. We are likely to overestimate the probability that we have the disease.

Tversky and Kahneman (1982) offered the following example. Suppose the percentage of homosexuals who have a disease is three times higher than the percentage of heterosexuals who have the disease. Pat is diagnosed with the disease. This is all that you know about Pat; you do not even know if Pat is male or female, much less Pat's sexual orientation. What is the probability that Pat is homosexual?

The disease is more representative of homosexuals than heterosexuals, so the representativeness bias would cause most people to overestimate the probability Pat is homosexual. The actual answer depends on the real numbers, so let us use a population of 100%, 10% of whom are homosexual (10). Of them 3 have the disease (30%). There are

90 heterosexuals with a 10% rate of disease (9). Three of 12 people (3 + 9) who have the disease are homosexual, so 25% of all diseased people are homosexual, and the probability that Pat is homosexual is 25%. Base-rate neglect pushes the probability estimate closer to 75% for many people (75% is three times 25%, reflecting the higher prevalence for homosexuals). Whenever possible, pay attention to base rates. They are especially important when an event is very rare or very common.

14.7.3.3 Law of Small Numbers

People fail to understand the nature of randomness. We often expect the properties of a population to be manifest in a small sample. We are inclined to think sample data are representative of the population and do not understand how the variability in a small sample can be very unrepresentative of the population's true parameters. People who have lived a long time on the coast tend to think they have seen just about every kind of storm. So, when they are warned to evacuate, many think they can ride it out like they have the vast array of other storms they have seen in their lifetime. Although 70 years on the coast may be a long time in human terms, it is a trivially small sample in geologic time, and many people have lost their lives to the law of small numbers by overestimating their experience and underestimating the strength of storms.

Take another example exaggerated for effect. The United States is a country of over 300 million people. Imagine we take a random sample of 10 people and use it to profile the age, gender, and race attributes of the United States. If five of our people are, by chance, African-American, two are Asian, and three are white, we would have a very biased view of the population. We would overestimate the probability of some subpopulations and underestimate others like the Hispanic and white populations.

APOPHENIA

You are running late for work, because your alarm did not go off. You step in the dog's vomit in the hall on the way to the kitchen where you learn there is no coffee in the house. The disposal unit mangles a spoon and comes to a whining stop when you hit the switch. You hurry to the car without your keys and have to make a return trip to the house. Any one of these irritants alone would not be too unusual, but together they outline the universe's plot against you. That is apophenia, the perception of connections and meaning among unrelated events.

Our brains are pattern detecting machines. They uncover meaningful relationships among the constant barrage of sensory input we face. The gambler's fallacy is a misperception of probability, where we think a number is due because it has not come up recently. Conspiracy theories, like the belief the twin towers of 9/11 were destroyed by the government are confabulations based on misperceived patterns. Despite a lack of evidence of a causal connection, many parents do not vaccinate their children because they believe they cause autism.

Source: Adapted from Poulsen (2012).

Things that by chance show up in fewer than representative numbers in our small samples get underestimated. Likewise, things that show up more in the sample than the population are overestimated. The tendency to expect samples to be representative of populations can distort our view of reality.

14.7.3.4 Confusion of the Inverse

People tend to confuse conditional probabilities with their inverse. They tend, for example, to think the P(Test positive|Disease positive) is the same as P(Disease positive|Test positive). Even doctors have been known to fall prey to this error. A doctor may see a positive test as being representative of having the disease, and they may see having the disease as being representative of a positive test. This could simply be semantic confusion or it could be failure to apply Bayes' Theorem. In either event it can lead people to misstate probability estimates.

Consider Table 14.2 for an example. The probability that a person is disease-positive is 1%. The test accurately predicts the disease when present (sensitivity) 80% of the time. It correctly predicts the absence of disease 90% of the time. Imagine a patient is given the test for the disease and it comes back positive. What is the probability that the person has the disease? In a study by Plous (1993), doctors said it was about 75%. In fact, the value we seek is P(D+|T+) and that is 800/10700 or 7.5%! The inverse, P(T+|D+) = 800/1000 = 80%. Clearly, the doctors had inverted the conditional probability, another form of the representativeness bias that can seriously distort probability estimates.

14.7.3.5 Confounding Variables

When trying to predict one value based on other values known to be related, we rarely give sufficient consideration to confounding variables. Most people recognize a relationship between height and weight. There is a tendency to make simple translations from one variable to the next. For example, if a person weighs 10% more than average, there is a tendency to expect them to be 10% taller than average. This translation from weight to height simply does not bear up in reality. There are many other variables that can affect or confound the relationship between height and weight, rendering it far from the representative translation our minds tend to prefer. When we estimate uncertainties based on imperfect relationships, we can over- or underestimate them because we neglect to consider the effect of confounding variables.

TABLE 14.2
Contingency Table for Disease Condition and Test Result

	Test+	Test−	Total
Disease+	800	200	1,000
Disease−	9,900	89,100	99,000
Total	10,700	89,300	100,000

There are, as you see, several forms of representativeness bias. They can be subtle and often difficult to counteract in a probability elicitation.

14.7.4 Anchoring-and-Adjustment

Many uncertainty judgments require us to produce numerical estimates of things that are not known or that do not have a correct answer. One strategy for doing this is to start with information one knows, an anchor, and then adjust up and down from it until an acceptable value is reached, using what Tversky and Kahneman (1974) called the anchoring-and-adjustment heuristic. The problems begin during decision making, when we humans tend to anchor, or overrely, on specific information or a specific value. When asked to estimate an uncertain quantity, we tend to come up with an initial estimate that is often our estimate of a most likely value like a mean, median, or mode. It comes, often unbidden, from wherever it comes and we do not always or often scrutinize it. Call this value an anchor. We then tend to make adjustments up or down to this value based on new circumstances. Once the anchor is set, there is usually a bias toward that value. We may identify the anchor, or an anchor may be provided for us. The major problem with this bias is that we rarely question the origin of the anchor and we rarely make adjustments to it sufficient to include extreme values.

Sometimes the anchor is provided to us. Imagine arriving at a restaurant where you are told the wait is 15 minutes. At 25 minutes you are feeling very frustrated and ready to leave. Now, imagine you were told the wait was 30 minutes. At 25 minutes your hope is rising. If both parties are seated at 25 minutes one is frustrated to be seated so late and the other is happy to be seated sooner than expected. When we are provided a piece of information we anchor on it and overrely on it.

The adjustment heuristic cuts down on mental processing time. We are better at relative thinking than absolute thinking. It avoids the need to reprocess information and recalculate values. The problem is that adjustments are anchor-centric, and we tend to stay too close to the anchor. A poorly chosen anchor can lead to poor adjustments. The anchors we come up with are often arrived at rather spontaneously, often with little analytical thought. The anchors we are provided are of unknown accuracy. Unless our anchors are challenged, we often do not understand the role they play in subsequent judgments.

Don't use the Internet to answer the following question; use your expertise. What is the driving distance from Sacramento, California, to Little Rock, Arkansas? I do not expect you to be able to guess such a quantity precisely, so give me the minimum mileage it could be and the maximum mileage it could be so that you are 90% sure you have estimated correctly. In other words, estimate a realistic interval for the distance, not one somewhere between zero and a million miles!

If you are like most people, when the original question was asked, a number popped into your mind. You somehow came up with a most likely or best-guess number. That is your anchor. When asked for a minimum and a maximum, most people tend to subtract a number from the anchor and then add roughly the same number to the anchor to get their interval. Did you? Did you make enough of an adjustment to capture the true distance?* Even experts rarely do so.

---

* Google Maps reports the distance as 1,960 miles by car.

Assessments of probabilities as well as other quantities are subject to the anchoring-and-adjustment bias. Many elicitation questions come with anchors embedded in them. For example, we may ask the towboat captain the probability it takes more than two hours to get from point A on the waterway to point B. Two hours has now been established as an anchor. Alternatively, the question could be asked as, "State the time at which there is a 60% chance you could travel from point A to point B on the waterway." The anchor is now 60%. These values could affect the response to all subsequent questions.

The anchoring-and-adjustment heuristic has important implications for the order and nature of questions posed in an elicitation. It can be reduced by avoiding anchors, that is, do not embed them in the questions. You may be able to help prevent experts from anchoring by exploring the upper and lower limits the quantity can take before examining central tendencies. Make sure people have time to think rationally about a quantity; pressure makes people default to anchors. More personally, observe yourself thinking, look for instances where you are adjusting a number instead of thinking critically. Identify the factors that could push you toward a larger value and think of their effects. Then do the same for a smaller value. Strive to replace your relative thinking with absolute thinking.

14.7.5 Motivational Bias

Bias is hard to avoid because we are often completely oblivious to our own biases. In addition to the cognitive biases identified here, there is the possibility of motivational bias. People may want, consciously or otherwise, to influence a decision; they may perceive that they will be evaluated based on the outcome; they may suppress their uncertainty to appear more knowledgeable; or they may have a strong position on a question (Morgan and Henrion 1990). Some people may just be cockeyed optimists while others are perpetual pessimists. Any of these outlooks can bias an estimate.

People, including experts, sometimes find an incentive to offer information that does not reflect their expert knowledge. People who have figured out what kinds of answers are most likely to secure more desirable outcomes may be motivated to provide those answers regardless of their true beliefs. In some situations, the rewards and punishments associated with over- or underestimating quantities vary significantly. Underestimating a human health risk has a greater risk to the expert than overestimating that risk, for example. Criminal lawyers will rarely tell you things will be fine. They are inclined to emphasize the seriousness of the defendant's situation. It makes higher fees easier to swallow. Estimates of uncertain quantities may be affected by implicit incentives as well as explicit ones.

Intentional motivational bias may be impossible to prevent. Unintentional motivational biases may be reduced if you can identify some of the biases of the individual or their organization. So, for example, if you have an expert from an environmental group known to oppose the activity that has given rise to your expert elicitation, it might be helpful to ask the expert to explore other perspectives. Supposing the navigation industry is known to support the activity, one might ask what the expert thinks they would say, then ask if there is a rational explanation for the differences. Adopting different perspectives can be an effective strategy for addressing many of these heuristics.

14.7.6 Confirmation Bias

Confirmation bias is a tendency to search for or interpret information in a way that confirms one's preconceptions. The problem is people, including experts, have been shown to actively seek out and assign more weight to evidence that confirms their hypothesis, and ignore or underweigh evidence that could disconfirm their hypothesis. Confirmation bias not only affects how people gather information, but also how they interpret and recall information. We need look no further than politics for a plethora of examples. If you like the incumbent president of your country, you likely get your news from sources that share your view and you see alternative sources as biased. If you dislike the incumbent you likely do the opposite. Furthermore, every action the incumbent takes is filtered through your like/dislike lens. We tend to seek information that supports our existing beliefs. This type of bias can prevent us from looking at a situation objectively. It can also influence the decisions we make and can lead to poor or faulty choices made when we fail to see the world as it truly is.

> Warren Buffett said, "What the human being is best at doing is interpreting all new information so that their prior conclusions remain intact."

To neutralize confirmation bias we must look for ways to challenge what we think we see. It can be difficult to motivate people to do that. Seeking information from a range of sources that provide multiple perspectives is a good place to start. These data sources can be included in an elicitation read ahead or it can be made part of the discussions that precede the elicitation. A free and open discussion with the intention of revealing information that challenges your view on a subject can be helpful. Assigning the role of "devil's advocate" to a person not on the expert panel can be an effective way to introduce evidence that is otherwise overlooked. Tasking experts with examining evidence that conflicts with their beliefs may help.

14.7.7 Framing Bias

One of the critical steps in an elicitation is to frame the questions; this is where things can often go wrong. The way a problem is framed or question is posed can profoundly influence the choices we make and the answers we give. In an elicitation, people tend to accept the frame they are given. It is rare that they would pause to reframe it in their own words.

Framing bias suggests that how something is presented, the "frame," influences the choices people make. This can interfere with the notion that rational choosers should always make the same decision when given the same data. Tversky and Kahneman (1982), in their explorations of framing theory, found otherwise. In their experiment, they gave some students a decision phrased in positive terms as a choice between a sure gain and an uncertain gamble. Most students exhibited risk aversion and chose the sure gain option. Other students were given the same choices phrased in negative terms as a choice between a sure-loss option and the risky gamble. The majority exhibited risk seeking tendencies and chose the risky gamble. This showed the frame could affect the choice people made.

A "framing effect" occurs when factually equivalent descriptions of a decision scenario lead to systematically different decisions depending on how they are phrased. Advertisers and politicians have mastered the art of framing. Meat is always 80% lean and never 20% fat. Framing effects can also affect our memories and reinforce other heuristics. What politician uses cost, expense, pay, or tax when touting the merits of their latest initiative?

To counteract this bias, try posing problems in a neutral way that combines gains and losses or embraces different reference points. Choose a frame that captures all of what's important. Encourage experts to make sure they are understanding the real problem. Help them look at the problem from other perspectives. For example, reverse the context. If you are creating the risk, how would you see things if you were bearing the risk? In some instances, it may help to choose a high-level perspective for framing. For example, looking only at project-by-project risk may result in a portfolio of overly conservative projects.

Clemen and Reilly (2001) offer a general suggestion for improving estimates based on knowledge of the many heuristics summarized here. Awareness of the heuristics can help people learn to become better assessors of probabilities. Knowing how not to think does not guarantee one will not think in these ways, but it does give the conscientious assessor a fighting chance. If assessors can recognize the effects as they occur, they may be able to avoid or countervail them.

O'Hagan et al.'s review of the literature finds that many of the biases mentioned here are ephemeral and can be eliminated entirely in some instances. The literature also suggests some people are more prone to these biases than others. More intelligent people are more likely to be able to reason through the biases and avoid them. The bottom line, however, remains the same. A careful elicitation protocol is the best way to address these problems.

14.8 CALIBRATION

Morgan and Henrion (1990) reported, "The most unequivocal results of experimental studies of probability encoding has been that assessors are poorly calibrated." If we go to all the trouble of a formal subjective probability elicitation, it would be nice if we could have some confidence in its results. The tricky part is if we say there is a 0.9 chance an oil spill will be less than 1 million gallons and a 2-million-gallon spill occurs, does that mean we were wrong? The answer, of course, is no. Only statements of absolute certainty, that is, $p = 0$ or $p = 1$, can be proven right or wrong by subsequent observations.

That means the strength or weakness of our elicitation more or less rises and falls with our process and our experts. Experts who are overconfident overestimate the actual frequency of events in their subjective probabilities. They might estimate a spill size that occurs with a 0.4 probability as having a 0.7 probability. Underconfident experts do the opposite. They might estimate the probability of that spill as 0.3.

Overextremity is another common pattern of estimation by experts. This is neither over- nor underconfident; rather it is a mix of both. These folks tend to overestimate the likelihood of low-probability events and underestimate the likelihood of high-probability events.

If it is important enough, your experts can be calibrated so you and they can learn how well they tend to recognize and evaluate their uncertainty. One common way to calibrate an expert is to ask a series of questions with factual answers that are unlikely to be known by the expert. Ideally, these questions will be from the domain of knowledge most relevant to the elicitation.

There are three general styles of calibration. On relies on asking a large number of a variety of questions as well as the expert's confidence in the correctness of the answers. The questions might include such things as how many gallons in an acre-foot of water, what is the capital of Zambia, and what is the currency used in Thailand? Experts are asked to express the likelihood that their answer is correct. With enough such questions, one would expect a well-calibrated expert to get about 60% of the questions they said they were 60% confident in to be correct, 70% of the questions they are 70% confident about correct, and so on.

Binary calibrations are a second common technique. This consists of a series of true/false questions such as the state of Idaho borders Iowa, the Baltimore Colts won the 1973 Super Bowl, blue light has a shorter wavelength than green light, and so on. The expert answers each question and then identifies the confidence that the answer is correct as 50%, 60%, 70%, 80%, 90%, or 100%. You total all of their percentages and convert it to a decimal to identify the expected number of correct answers. Suppose there are 100 questions and the sum of all confidence ratings is 6240%. As a decimal that is 62.4, so that is the expected number of correct answers. An expert with fewer than 62 correct answers is overconfident. One with more than 62 correct answers is underconfident. The expert with close to 62 correct answers is well calibrated.

A third technique requires experts to estimate quantitative questions with 90% confidence intervals. They might be asked the distance from Kiev to Paris, the height of Mount Denali, the mean weight of an adult male boxer dog, and so on. A well-calibrated expert will get 90% of the answers within their interval estimates. Those with fewer than 90% correct are overconfident, and those with more than 90% are underconfident. Other more sophisticated techniques and scoring rules, like the Brier Score (Brier 1950), have also been used.

The calibration process, prior to an elicitation, is one of the most effective ways to make an expert aware of the overconfidence bias. Once aware, many educated experts are able to neutralize that bias.

14.9 MULTIPLE EXPERTS

Multiple experts will frequently be used when an elicitation is needed. Assuming the task is to elicit a probability distribution, the group can be used to create a single distribution (behavioral aggregation), or each individual expert may provide his or her own distribution. In this latter case, it may be desirable or necessary to eventually combine them all into a single distribution (mathematical aggregation) sometimes called the composite expert or the synthetic decision maker.

The literature on strategies to combine experts' probabilistic judgments is extensive. Clemen and Winkler (1999, 2007) provide excellent overviews of this literature. Combining the judgments of different experts may be a sensible thing to do at times. However, the purpose of an expert elicitation is to quantitatively describe the uncertainty being characterized. If the experts' judgments about the relevant science are noticeably different, then it may be best not to combine the separate judgments. Instead it may be prudent to run separate analyses using the inputs provided by the individual experts to explore how the different opinions affect the outcomes of interest. It is wise to consider the spread of expert assessments as valuable information that should always be part of the reporting. There is no reason to combine expert assessments if you do not need to (Cooke and Probst 2006).

Faced with uncertainty, decision makers invariably prefer agreement or an unambiguous consensus from experts. It is not reasonable to expect total consensus when dealing with difficult uncertainties. When scientists disagree, any attempt to impose agreement can cause people to confuse consensus with certainty. The goal should be to quantify uncertainty, not to disguise it with consensus (Aspinall 2010). Thus, multiple judgments are to be preferred to a single judgment whenever practicable.

One common model for elicitation is to have individual experts develop their own distributions and then to share these results, anonymously, with the group of experts. In a plenary session, they consider their results in discussion and then allow the experts to modify their elicitations based on what they learn from the discussion that ensues. This method still produces as many distributions as there are experts. Using each of the individual distributions as a sort of sensitivity analysis, to establish whether or not the differences among the experts make much of a difference in model outputs and decisions, is a solid strategy. This method has the advantage of providing what amounts to high and low estimates of the outputs of interest, which may help bound the uncertainty about these values.

There are a number of techniques available for aggregating individual distributions when a single estimate is required or desired. Bayesian methods (Clemen and Winkler 1999), opinion pooling that assigns weights to the distributions, and Cooke's classical method (Cooke 1991) are all in common usage. O'Hagan et al. (2006) suggest that the average of the distributions provides a simple, general, and robust method for aggregating expert knowledge. Averages may be computed across probabilities, that is, averaging all x values for each selected probability, or across x values, that is, averaging all probabilities for a given x value. They report that the former is most common and less preferred by those with an interest in characterizing the uncertainty. Averaging across p values rather than x values has the disadvantage of masking the uncertainty that exists when there is a wide range of opinions.

Structured expert judgment, also known as Cooke's classic method, is a highly regarded method for developing a rational consensus among expert opinions. This process includes a calibration of the expert's proficiency using a set of domain relevant seed questions. The elicitation is then conducted using one of the standard protocols. The expert's performance is, measured based on statistical accuracy and informativeness (the degree to which the expert concentrates his or her probability mass on a small region of the number line). Weights are assigned based on each expert's performance, and a composite is produced.

Source: Cooke and Probst (2006).

Group elicitation seems to offer substantial and relatively untapped potential for modelers.

O'Hagan et al. argue that it offers better synthesis and analysis of knowledge through group interaction. The facilitator's role grows ever more important when group dynamics play such a significant role. The process should encourage sharing knowledge as opposed to opinions. It should also recognize expertise and differences in expertise as well as make effective use of feedback from the group. At the same time, the process must avoid domination by individuals or by shared knowledge. Groups can have an even stronger tendency toward overconfidence, especially if group-think sets in. The challenge of avoiding group reliance on heuristics remains with a group process. Of greatest concern is the tendency for group consensus to mask the true extent of the uncertainty.

14.10 SUMMARY AND LOOK FORWARD

There will be times when you simply do not have all the data you wished you had for your risk assessment. Some of those times you will not have the luxury of being able to commission and wait for research to fill those voids. Subjective probabilities can be useful in filling in data gaps and voids.

Not every subjective probability estimate warrants a formal elicitation. Nonetheless, assessors are well advised to follow good elicitation techniques whenever they estimate subjective probabilities. Spoken language has proven to be notoriously imprecise when trying to express our beliefs about things that are uncertain. Hence, we have come to rely on numerical estimates of probabilistic information and probabilistic estimates of numerical information. The most consistent result in all the related literature is that we are notoriously bad at estimating probabilistic information. There are several well-known heuristics we have come to rely on that short-circuit any analysis of uncertain data. Consequently, it is important to devise and use elicitation techniques that can limit the damage done by these heuristics.

Now that we have a pretty good idea how probability distributions are chosen and elicited, we turn in the next chapter to the Monte Carlo process to see how this technique is used. Probability distributions are used more often in simulations than in analytical calculations in risk assessment, and the Monte Carlo process is an extremely powerful and useful technique for propagating the uncertainty in our model inputs through a model. Chapter 15 takes a peek behind the curtain of the Monte Carlo process to see how it works so you can better understand how your risk assessment models work.

REFERENCES

Aspinall, W. 2010. A route to more tractable expert advice. *Nature* 463 (7279): 294–295.

Ayyub, B. M. 2001a. *Elicitation of Expert Opinions for Uncertainty and Risks*. Boca Raton, FL: CRC Press.

Ayyub, B. M. 2001b. *A Practical Guide on Conducting Expert-Opinion Elicitation of Probabilities and Consequences for Corps Facilities*. IWR Report 01-R-01. U.S. Army Corps of Engineers Institute for Water Resources, Alexandria, VA 22315-3868.

Brier, G. W. 1950. Verification of forecasts expressed in terms of probability. *Monthly Weather Review* 78: 1–3. http://docs.lib.noaa.gov/rescue/mwr/078/mwr-078-01-0001.pdf.

Clemen, R. T., and T. Reilly. 2001. *Making Hard Decisions with DecisionTools*. Pacific Grove, CA: Duxbury Thomson Learning.

Clemen, R. T., and R. L. Winkler. 1999. Combining probability distributions from experts in risk analysis. *Risk Analysis* 19 (2): 187–203.

Clemen, R. T., and R. L. Winkler. 2007. *Advances in Decision Analysis: From Foundations to Applications*, eds. W. Edwards, R. F. Miles Jr., and D. von Winterfeldt, 154–176. New York: Cambridge University Press.

Cooke, R. M. 1991. *Experts in Uncertainty: Opinion and Subjective Probability in Science*. Oxford, UK: Oxford University Press.

Cooke, R. M., and L. H. J. Goossens. 1999. *Procedures Guide for Structured Expert Judgment, Delft University of Technology Delft*. The Netherlands Report prepared under contract No. ETNU-CT93-0104-NL for the Commission of European Communities Directorate-general XI (Environment and Nuclear Safety) Directorate D (DG 11 CCJH) June 1999 EUR 18820, Luxembourg/Brussels.

Cooke, R. M., and K. N. Probst. 2006. Highlights of the expert judgment policy symposium and technical workshop. *Conference Summary, Resources for the Future*.

DeGroot, M. H. 1970. *Optimal Statistical Decisions*. Hoboken, NJ: Wiley Interscience, 2004.

Fiedler, K. 1988. The dependence of the conjunction fallacy on subtle linguistic factors. *Psychology Research* 50 (2): 123–129.

Gigerenzer, G., P. M. Todd, and ABC Research Group. 1999. *Simple Heuristics That Make Us Smart*. New York: Oxford University Press.

Gilovich, T., D. Griffin, and D. Kahneman, eds. 2002. *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge, U.K.: Cambridge University. http://assets.cambridge.org/97805217/92608/sample/9780521792608ws.pdf.

Hardman, D. 2009. *Judgment and Decision Making Psychological Perspectives*. West Sussex: John Wiley and Sons.

Koehler, J. J. 1996. The base rate fallacy reconsidered: Normative, descriptive and methodological challenges. *Behavioral and Brain Sciences* 19 (1): 1–53.

Lindley, D. V., A. Tversky, and R. V. Brown. 1979. On the reconciliation of probability assessments. *Journal of the Royal Statistics Society* 142 (2):146–180.

Meyer, M. A., and J. M. Booker. 2001. *Eliciting and Analyzing Expert Judgment: A Practical Guide*. Philadelphia, PA: Society for Industrial and Applied Mathematics and the American Statistical Association.

Morgan, M. G. 2014. Use (and abuse) of expert elicitation in support of decision making for public policy. *Proceedings of the National Academies of Science*. http://www.pnas.org/content/111/20/7176 (accessed April 14, 2018).