

Chapter 13

Characteristics of a Process for Subjective Probability Elicitation



John Quigley and Lesley Walls

Abstract The elicitation of subjective probabilities from experts can be critical in determining a course of action when making decisions under uncertainty. A sound process to elicit probabilistic judgement is necessary to ensure that good quality data are used to inform the decision-making, as well as to provide protection to those accountable for the consequences of the determined actions. We synthesise the characteristics of a good elicitation process by critically reviewing those advocated and applied. We compare the processes inherent in the guidance produced by two professional bodies to exemplify the manner in which the characteristics manifest themselves in practice. We examine whether standardisation is meaningful given the maturity of processes for the elicitation of subjective probability.

13.1 Introduction

Big data and the digital age have not removed the need for nor diminished the importance of expert judgement; observed data are history and expert judgement is the future. We still require expert judgement to support decisions where observed data are few or non-existent. Also we can require expert judgement in situations where observed data are abundant since the relevance of the past to the future can be assessed only with expertise (Hora 2007; Quigley and Walls 2018). This is not likely to change as more observed data are collected. Further, for situations with little or no observed data, we believe that concepts like *black swans* (Taleb 2007), *perfect storms* (Junger 1997) or *deep uncertainty* (Cox 2012) should not be an excuse for superficial thinking about possible future events (Dias et al. 2018). We encounter many problems where there exists relevant expertise and for which the problem characteristics are

J. Quigley (✉) · L. Walls
University of Strathclyde, 130 Rottenrow, Glasgow G4 0GE, Scotland
e-mail: j.quigley@strath.ac.uk

L. Walls
e-mail: lesley.walls@strath.ac.uk

© Springer Nature Switzerland AG 2021
A. M. Hanea et al. (eds.), *Expert Judgement in Risk and Decision Analysis*,
International Series in Operations Research & Management Science 293,
https://doi.org/10.1007/978-3-030-46474-5_13

287

measurable in theory but not in practice; these conditions are ideally suited for expert judgement (Cooke and Goossens 2008).

We are concerned with the elicitation of quantitative subjective judgement, specifically the expression by experts of their beliefs in the form of subjective probability distributions. Such measures do not come naturally to people and so we require a process to facilitate the expression. Research has indicated that there is a need for formal elicitation to extract and quantify judgements since people, even experts, are unable to provide accurate data simply on request; see, for example, Cooke (1991), Meyer and Booker (2001).

Since the work of Tversky and Kahneman (1974), there has been awareness of the biases and heuristics people apply in decision-making under uncertainty that can result in poor probability assessments. Examples include contextual biases and heuristics such as anchoring, availability and representativeness. Other challenges associated with assessments made by people include issues such as groupthink (Janis 1971), group polarisation (Myers and Lamm 1976), overconfidence (Soll and Klayman 2004) and difficulties associated with communicating knowledge in numbers and probabilities (Gigerenzer and Edwards 2003). Inappropriate and ill-informed elicitation can amplify biases by relying on subjective and unreliable methods for selecting experts (Shanteau et al. 2002), asking poorly specified questions (Wallsten et al. 1986), ignoring protocols to counteract negative group interactions (Janis 1971) or applying subjective or biasing aggregation methods (Aspinall and Cooke 2013; Lorenz et al. 2011).

An elicitation process design should address these known issues. However, according to Burgman (2004), Krueger et al. (2012), Kuhnert et al. (2010) and Regan et al. (2005), amongst others, informal methods for elicitation persist. French (2012), Choy et al. (2009) and Krueger et al. (2012) report that few elicitations provide sufficient detail to enable review and critical appraisal. The consequences of poor judgement are misinformed decision-makers as illustrated by Wilson (2017) who reported a 52% hit rate in 95% intervals from his investigation of selected expert judgement studies.

Reported expert probabilistic assessments have been conducted for almost 50 years. Early studies include WASH-1400 concerning nuclear reactor safety (United States Nuclear Regulatory Commission 1975) that applied methods further developed into NUREG-1150 (United States Nuclear Regulatory Commission and others 1990). Approaches to elicitation continue to be developed and expert probability assessments remain a key input to policy and decision-making today. Examples include the following: Determination of volcanic eruption-related field-work risks (Christophersen et al. 2018); pollination uncertainty to inform policy-making for ecosystems (Barons et al. 2018); the combined effect of the meteorology and oceanography (also known as metocean) in offshore engineering (Astfalck et al. 2018); assessment of technology uncertainty during aerospace product design (Hodge et al. 2001); role of technical expert panels in probabilistic seismic risk analysis (Budnitz et al. 1998); expert judgement underpinning influential global environmental policies (Hemming et al. 2018) such as International Union of Conservation

Nature (IUCN) Red List (IUCN, 2012) and Inter-governmental Panel on Climate Change (IPCC) Assessments (Mastrandrea et al. 2010).

A sound process to elicit judgement for such problems is necessary to inform the decision-making. In addition, a sound process can also provide protection to those accountable for the consequences of the determined actions. Consider the 2009 L'Aquila earthquake tragedy in Italy where 309 lives were lost (for details, see Nature 2011 and Science 2012, 2014). This case highlights the need for transparent, rigorous and widely accepted processes for assessing uncertain events. On appeal in November 2014, only the government official continued his prison sentence as the responsible person for the risk communication, while the six scientists who provided expert advice were acquitted. Nevertheless, in the original trial in October 2012, the six scientists as well as the government official, who participated in Italy's National Commission for the Forecast and Prevention of Major Risks six days prior to the earthquake, were sentenced to six years in prison for manslaughter. The prosecution argued that the expert advice from the Commission resulted in 30 people deciding to stay indoors contributing to their death; the scientists were brought to trial originally because of poor practice and the presiding judge ruled their analysis superficial. Additional criticism of this risk assessment has been made by the President and General Secretary of the International Seismic Safety Organisation (ISSO) concerning the lack of independence amongst expert judgements (Martelli and Mualchin 2012).

Our contribution is to characterise more generally what makes a good elicitation process by critically reviewing relevant literature and reported applications. Our intent is to inform others responsible for developing future elicitation processes for specific purposes and contexts of the characteristics of a good elicitation process. Section 13.2 describes the seminal work of the Stanford Research Institute (SRI) in constructing an interview process to address a variety of known biases commonly encountered with the elicitation of subjective probabilities from experts. Section 13.3 extends our discussion of the issues surfaced in the previous section through a wider review of subsequent work and organises these issues into the emergent characteristics of a good elicitation process. Section 13.4 compares the elicitation guidance documents from two professional societies to illustrate and assess how the general characteristics manifest themselves for different purposes. Section 13.5 explores whether standardisation of an elicitation process for subjective probability is useful given the maturity of practice. Section 13.6 presents our concluding discussion.

13.2 Stanford Research Institute Elicitation Process—The Genesis

Although the RAND Corporation developed formal approaches for the elicitation of expert judgement in the 1960s, these were not for probabilistic judgements (Dalkey and Helmer 1963; Dalkey 1967, 1969). Spetzler and Stael von Holstein (1975) were the first to report an elicitation *process* for subjective probabilities grounded in

practice by the Decision Analysis Group at the Stanford Research Institute (SRI). Previously, research had focused upon methods to encode probability assessments that concentrated more narrowly on the quantification of expert uncertainty rather than wider processual considerations; see, for example, Hampton et al. (1973) and Winkler (1967). By broadening the scope to position encoding methods within an elicitation process, the so-called SRI five-stage approach seeks to identify potential biases and minimise their impact on the quantitative assessment. Since the SRI process is concerned with structuring an interview with the expert, some important aspects, such as expert selection, are not considered. We describe the five stages—namely, *motivating*, *structuring*, *conditioning*, *encoding* and *verifying*—since these provide a common basis that has informed many subsequent elicitation processes.

13.2.1 *Motivating Stage*

The SRI advocates that the process design should address motivational biases, such as management and expert bias. Management bias occurs when an expert provides goals rather than judgement. For example, an expert states the aspiration that there will be no weaknesses in a system by time of manufacture, rather than providing an assessment of their beliefs about the likely occurrence of weaknesses. Expert bias occurs when a person becomes overconfident merely because they are called an *expert*. During this stage, the intent is to determine if there are motivations for the expert to, consciously or unconsciously, adjust probability assignments based on perceived rewards.

Motivational biases can be identified through discussion, where the interviewer develops a rapport with the expert and discusses openly any payoffs that might be associated with the probability assignment as well as possible misuses of the information; for example, single-number predictions are often interpreted as commitments. The interviewer should make clear to the expert that no commitment is inherent in a probability distribution, and that complete judgement from the expert is sought. Additionally, the interviewer introduces the encoding task to the expert by explaining both the importance and purpose of probability encoding in relation to the decision, and clarifying the difference between probabilistic and deterministic predictions.

13.2.2 *Structuring Stage*

Structuring involves defining the event under consideration to minimise ambiguity in the questions and to explore how an expert thinks about the quantity for which probability judgement is to be elicited. The aim is to manage possible cognitive bias by simplifying the complex task of assigning probabilities by disaggregating the quantity of interest into more elementary variables (Armstrong et al. 1975). However, the unpacking principle (O'Hagan et al. 2006), also known as subadditivity (Tversky and

Koehler 1994), may be the consequence. This refers to the situation where the more detailed the description of the event, the greater the likelihood assigned to it. For example, an expert may provide an assessment for the probability of a component failing and subsequently during the elicitation process provide probabilities associated with causes of that failure that may result in a component probability exceeding the initial assessment.

The quantity of interest needs to be specified so that a measurement scale can be determined. There is a need for precise thinking about how the quantity of interest will be realised. For example, if the exchange rate between two currencies next year was to be assessed, then it would be necessary to specify the exact time next year for the measurement as well as where the currencies would be exchanged, as banks, stock exchanges and tourist agencies all buy and sell currencies and offer different rates.

It is important to choose a scale that is meaningful to the expert. One important consideration when selecting the quantity of interest is feedback to the expert. This is considered crucial for calibrating the expert and should be event-specific (Fischhoff 1989; Bolger and Wright 1992; Ferrell 1994). In other words, the feedback must be with respect to assigning probabilities to particular classes of relevant events and not only feedback on the ability of the expert to assign probabilities to any situation. To increase the effectiveness of feedback in terms of learning, conditions that influence the event should re-occur as often as possible (Fischhoff 1989; Kadane and Wolfson 1998). Therefore, the factors on which the measure is conditioned should be as few and general as possible. The structure of the quantity of interest may need to be expanded so that the expert does not have to model the problem further before making each judgement.

Structuring should encourage the expert to think carefully about the event before the actual encoding session begins by probing and clarifying issues concerning relevant and irrelevant background information.

13.2.3 Conditioning Stage

Information relevant to assessing the probabilities is discussed to address issues such as availability bias and anchoring and adjustment. Availability bias refers to the influence that easily recalling examples can have on the assessment of probability, such as overestimating the likelihood of a disaster because it has devastating consequences unrelated to its frequency. Anchoring and adjustment is a heuristic, where people base their judgement on a piece of information (i.e. the anchor) and adjust for the assessment. For example, an expert making a series of assessments will provide an initial assessment for the first quantity of interest, and all subsequent assessment will be adjustments; anchoring and adjustment can lead to overconfidence and other judgement errors (Kahneman et al. 1982). Such discussions can also form part of the structuring stage for these and other possible biases such as the conjunction fallacy, whereby people guess that the odds of two or more events co-occurring are greater

than the odds of any one of the events occurring alone because the co-occurrence appears more representative (Tversky and Kahneman 1983).

The conditioning stage aims to encourage the expert to think fundamentally about their judgement, understand how they make probability judgements and through revealing the information that seems most available, what (if any) anchors are used and what unstated assumptions are being made. Experts can be asked to specify the most important bases for their judgement to identify anchors as well as exploring more extreme situations.

13.2.4 Encoding Stage

This stage refers to the actual method for elicitation of the probability distribution for the quantity of interest. A popular encoding procedure for distributions is the fractile method where the expert assesses the median value of their subjective probability distribution along with, say, the (25th,75th) and the (5th, 95th) percentiles. The order in which these quantities are elicited should start with the extreme values first and progress towards the central values to avoid a central bias (Seaver et al. 1978).

After percentiles of the distribution have been assessed, graphical techniques can be applied to enhance the quality of the distribution (Chaloner et al. 1993). Once these probability values have been elicited, then a parametric distribution might be sought to maximise fit.

Spetzler and Stael von Holstein (1975) provide the following suggested steps in encoding.

- (a) Ask for extreme values—deliberate use of availability to counteract central bias.
- (b) Ask for scenarios that lead to realisations beyond extreme values—makes outcomes more available to an expert so more likely to assign higher values to extreme outcomes to address central bias.
- (c) Assign probabilities to scenarios—increases variability in overall distribution.
- (d) Choose values and assign probabilities—do not choose a significant value for assessment as this will lead to anchoring, but choose values experts will be comfortable with assessing.
- (e) Construct Cumulative Distribution Function (CDF).
- (f) Fit curve.

If the expert is to assess multiple quantities of interest, then in step (d) it is recommended the probabilities (i.e. percentile) are fixed and the values elicited so that numbers are not provided upon which an expert might anchor.

13.2.5 Verifying Stage

Since a subjective probability distribution has been elicited, the interviewer now guides the expert through a review of the distribution to ensure it reflects his/her

expressed belief. If it does not, then additional elicitation is required. Verification is accomplished by showing the expert the implications of the interviewer's interpretation of their response.

Two common activities performed at this stage include visualising the probability density function and comparing equi-probable intervals from the CDF. Asking which interval the expert would bet on supports verification because the expert should be indifferent to betting between intervals if the CDF reflects their belief. It is suggested this activity should be repeated three to five times.

Verification is required to ensure that an expert has provided a reflection of his/her true beliefs. If problems are encountered, then the previous stages are to be repeated.

13.2.6 Extensions of the SRI Process: Aggregation and Discretisation

Miley Merkhofer, manager of the Decision Analysis Research Program at the SRI between 1975 and 1983, reported an extended SRI process that included a sixth and seventh stage, namely, Aggregation and Discretisation, respectively (Merkhofer 1987).

For situations where multiple experts are assessing the same quantity of interest, then individual probability distributions may need to be aggregated; evidence suggests that combined judgement can improve assessment quality (Ashton and Ashton 1985). There are two approaches to aggregation—mathematical and behavioural. The former implies the experts should not influence each other's decisions (Ferrell 1985). The latter requires experts to share their judgement and re-assess their distributions and includes techniques such as Delphi (Ferrell 1985) and Nominal Group Technique (Moore 1987), see Gosling (2018). There are several mathematical approaches to aggregation most of which aim to evaluate a weighted average across the experts. See Cooke (1991) for a fuller discussion as well as more recently developed methods such as Wisse et al. (2008).

Rather than encoding using, say, the fractile method, it can be necessary to treat continuous random variables as discrete. Discretising refers to techniques for fitting continuous distributions to the elicited data while preserving important moments. This is accomplished by dividing the range of all possible values for the uncertain variable into intervals, selecting a representative point from each interval and assigning that point the probability that the actual value will fall within the corresponding interval. The moments can be preserved through, for example, Gaussian quadrature techniques (Miller III and Rice 1983).

13.2.7 Managing Bias

Throughout the SRI process, the interviewer explores the potential for bias with the expert and takes steps to manage bias by careful consideration of issues as discussed above.

At the time of development of the SRI in the 1970s and 1980s, such consideration of expert bias was the only approach. During the 1990s, Roger Cooke introduced a more formal means of measuring bias where seed variables are used, and experts assess quantities that are unknown to them but known to the analyst; see, for example, Cooke (1991), Quigley et al. (2018).

13.3 Characteristics of an Elicitation Process

We now synthesise the characteristics of good elicitation processes based on a review of two classes of publications. First, proposals from the scientific literature, which describe largely positive attributes we seek an elicitation process to possess. Secondly, insights gained from published criticisms of practical applications, which are indicative of pitfalls to avoid when designing and implementing an elicitation process.

A variety of literature sources have been drawn upon including books, studies, critical reviews as well as journal articles. Books focusing on expert judgement include Cooke (1991), Meyer and Booker (2001), Ayyub (2001) and O'Hagan et al. (2006). The U.S. Nuclear Regulatory Commission has published several relevant documents. These include NUREG-1150 (United States Nuclear Regulatory Commission and others 1990) which reports how to estimate the uncertainties and consequences of severe core damage accidents in selected nuclear power plants for which Keeney and Von Winterfeldt (1991) provide a critical appraisal. NUREG/CR-6372 (US Nuclear Regulatory Commission 1997) provides guidance on the use of expert judgement for seismic hazard analysis and is accompanied with practical guidance from NUREG-2117 (Kammerer and Ake 2012); lessons from more recent application of these guidelines are discussed in Siu et al. (2015). Cooke and Goossens (2008) review various elicitation applications, while Shephard and Kirkwood (1994) provide an in-depth description of an elicitation case study. Walls and Quigley (2001) describe how the SRI model informed an elicitation process for assessing uncertainty in product development, for which Hodge et al. (2001) reflect upon the lessons learnt from the perspective of multiple participant roles. Additional references are given to the literature in relation to specific issues discussed below.

We acknowledge that a specific situation will require a particular elicitation process. Figure 13.1 summarises some questions likely to be asked by any analyst approaching elicitation of subjective probabilities and indeed is based on the questions the authors themselves posed in such a situation. Within the diagram, the specific questions are grouped around the who, what and how of approaching an elicitation process. Here, we seek only to discuss general characteristics of good practice that will be transferable across a variety of problem domains where subjective probability assessment of some quantity of interest is required.

Figure 13.2 illustrates our structured collation of the issues emerging from the literature. We show that the *process* should be grounded in scientific *principles*, while taking account of the *purpose* of elicitation that determines the quantities of

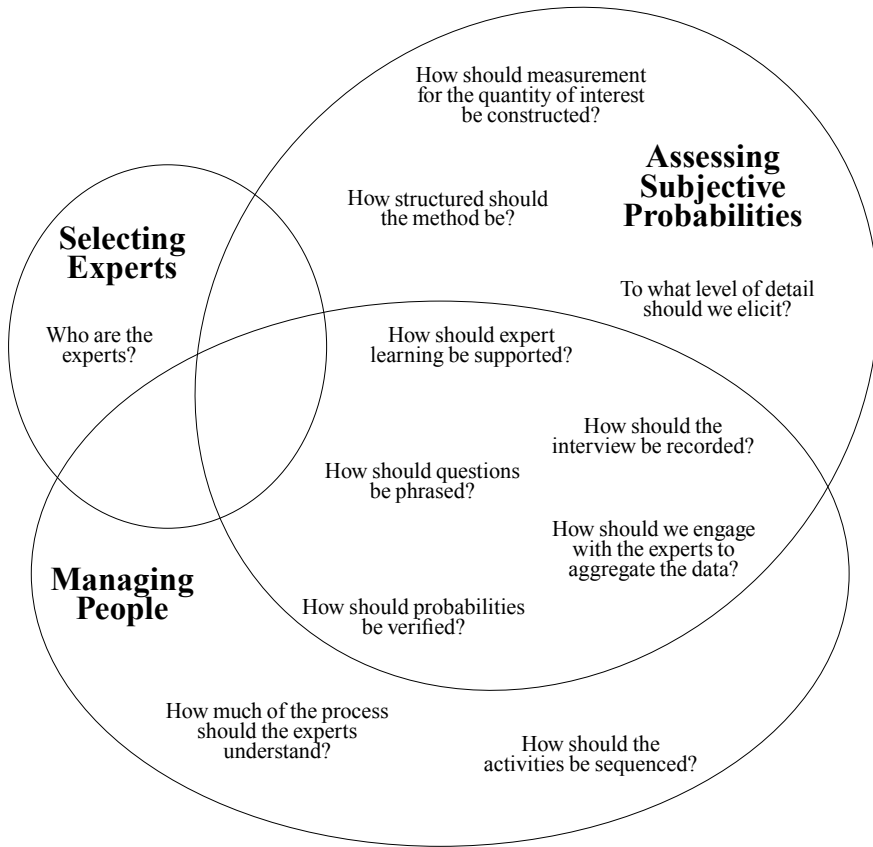


Fig. 13.1 Example questions posed by an analyst when designing an elicitation process

interest for which *probabilities* are to be elicited. The *people* whose assessments will be elicited will be sourced from the *purpose* context, and their expertise will depend on the quantities for which *probabilities* are required. The inner design, plan and implement loop reflects that an elicitation is not necessarily a linear process. Our following discussion reflects the inter-dependency between issues to be considered when developing a good elicitation process.

13.3.1 Principles

Cooke (1991) proposes that expert judgement processes should be subject to the following principles.

Scrutability/accountability: All data, including experts’ names and assessments, and all processing tools should be open to peer review and the results must be reproducible

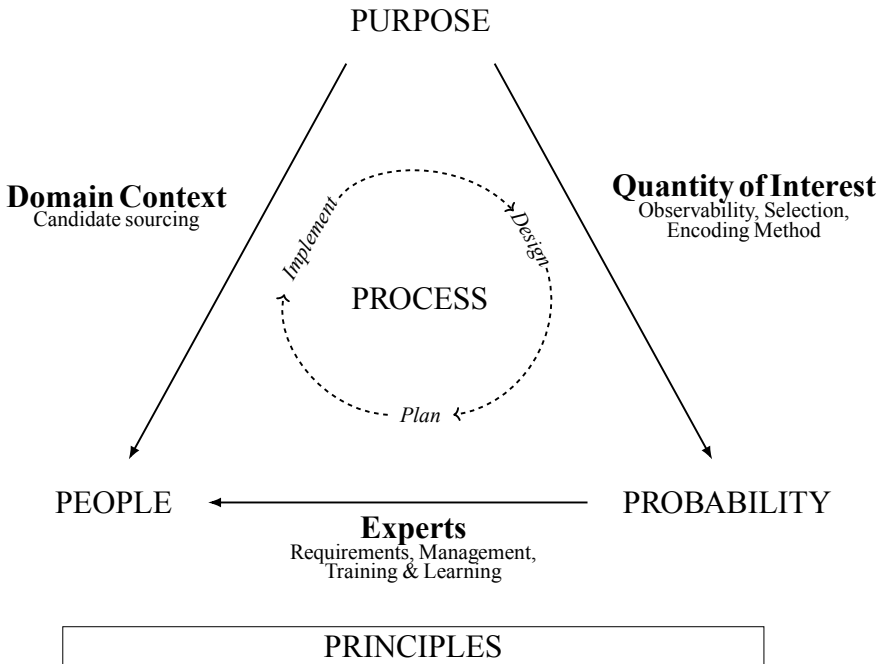


Fig. 13.2 Key characteristics of a good elicitation process emergent from the literature

by competent reviewers. It is not sufficient to present synthesised summary measures of expert assessments only and all subjective probabilities should be traceable back to the individual expert. Cooke (1991) argues strongly for publishing the names of experts for public decision-making but acknowledges the potential disadvantage in relation to conflict of interest for private firms.

Empirical control: Expert assessments should be susceptible in principle to empirical control. Scientific statements should be falsifiable in principle and, while such a test may not be feasible, it should be possible. Essentially this principle is guarding against an expert being free to say anything, and inferring one subjective probability is as good as another.

Neutrality: The method of elicitation should encourage experts to state their true opinions. Cooke (1991) suggests the Delphi technique, as reported by Sackman (1975), is an example where experts are encouraged not to deviate too far from the median of the group, as well as a providing a process where experts are required to self-assess their judgement implying there is no incentive for honesty (see Brockhoff 2002).

Fairness: All experts should be treated equally a priori. Note that this does not prevent unequal treatment of experts a posteriori. In contrast to the fairness principle, some

Bayesian methods for combining expert judgement require the decision-maker to assess the reliability of an expert prior to the elicitation. Further, there is a lack of guidance upon what basis such an assessment can be made. Even if such guidance did exist, then the fairness principle would be violated.

13.3.2 Purpose

According to Kammerer and Ake (2012) and the US Nuclear Regulatory Commission (1997), the purpose of an elicitation study is to provide a representation of the centre, body and range of views of the informed expert community regarding the quantity of interest. The output of an elicitation process involving multiple experts is not consensus but integration, since there is no one correct answer. The elicitation leads to the construction of what has been termed by Siu et al. (2015), as a community probability distribution.

The purpose of a process is more than just facilitating the elicitation of structured expert judgement while minimising biases in the resulting assessments. Importantly and additionally, the process should enable judgements to be subject to review and critical appraisal (French 2012). This need for a process to be transparent and repeatable means that documentation is a key enabling activity. It is the recording of the goal of the exercise, the design of the process and the judgements obtained to an appropriate level of detail and clarity that enables the process to be repeated (Cooke 1991).

Documentation has a variety of purposes (Bonano et al. 1990), including improving decision-making, enhancing communication, facilitating peer review, avoiding biases in judgement, unambiguous identification of the current state of knowledge and providing a basis for updating. A key strength of formal expert elicitation is in the documentation of the complete process as well as of the elicitation results and reasoning (Keeney and Von Winterfeldt 1991).

13.3.3 Probability Assessment of the Quantity of Interest

We consider a set of issues relating to the definition of meaningful quantities of interest for which probability assessments are required and the nature of modelling choices to be made in relation to how such assessments are obtained.

13.3.3.1 Observable Quantity of Interest

Many models contain parameters that are both unobservable and uncertain. The variable to be elicited should be related to an observable quantity, at least in principle. Cooke and Jager (1998) and Frijters et al. (1999) examine how to accomplish this.

For a relative frequency, for example, a large virtual population could be imagined and appropriate random selections considered. This should assist in ensuring use of a consistent definition for the quantities of interest (Keeney and Von Winterfeldt 1991).

13.3.3.2 Selection of Quantity of Interest

While an audit of available data may preclude the necessity of quantifying some variables, care is needed when assessing the relevance of the data (Siu et al. 2015). It can be useful to remind the expert in situations where empirical data are available, say test data, that judgement is needed to assess the relevance of that data to the practical field situations of interest. Mechanical processing of empirical data without consideration of its relevance towards the specific conditions under consideration may not result in appropriate representations of uncertainties.

Quantities of interest can be organised into similar groups to help reduce the cognitive burden by minimising the number of mental models required in assessment (Quigley and Walls 2010). However, this could introduce the bias of anchoring, so care should be taken.

Resources will be constrained and elicitation can be time-consuming; hence, better planning can result in better data. Careful expression of judgement can be a fatiguing process for an expert, especially when relevant data are sparse (Siu et al. 2015). The number of quantities of interest that an expert can meaningfully quantify in a session is limited (Keeney and Von Winterfeldt 1991); hence, it might not be possible to quantify each, and so screening the variables may be necessary (Bonano et al. 1990). Consideration should be given to the number of parameters that can be realistically assessed so that they can be prioritised for judgement. Alternative strategies will need to be considered for the remaining variables, informed by their importance towards the decision-making and the associated range of uncertainty. Bonano et al. (1990) suggest three experts being involved in parameter selection: specialists with subject matter knowledge, generalists with expertise in modelling and experts in sensitivity analysis.

13.3.3.3 Method for Encoding Probability

Requesting experts to provide their estimates in the form of a set of pre-designated quantiles can protect the expert against anchoring to the provided values as well as creating some consistency across questions that may lead to efficiencies in assessments. However, providing such judgements requires a degree of introspection and many experts do not think naturally in terms of the quantiles (Siu et al. 2015). Processes should remain flexible in accepting expert input in the form the expert feels most representative of their beliefs. When there are multiple experts, then compromise might be required.

There are important advantages in using parametric probability distributions to represent expert judgement. Three advantages, in particular, are intrinsic smoothing of the expert assessments, interpolation between assessments and extrapolation beyond the assessments. However, care is needed to avoid force-fitting a parametric model to expert assessments (Siu et al. 2015). A parametric probability distribution, such as the Normal distribution, provides an infinite range of probabilistic assessments and, typically, an expert only provides a few assessments. Given that the implications of a parametric model choice may not be apparent to an expert, then it is important that the elicitation process should include activities to check model adequacy for the expert assessment. It is particularly important that the analyst checks that the probability distribution elicited is a good fit in the part of the function (e.g. the tails) that will drive the decisions.

Considering only a limited number of parametric models to represent an expert's belief should be avoided. For example, situations where multi-modal distributions (representing the possibility of distinct, competing "models of the world") accurately represent the state of knowledge can be missed if the elicitation process focuses only upon a limited number of parametric models, especially the common uni-modal distributions. Therefore, the elicitation process should ensure that any mathematical representations of probability assessments do not unduly distort an expert's beliefs for the sake of convenience.

13.3.4 Managing the People Participating in the Elicitation

We now characterise the different roles of participants in the elicitation process and, in particular, discuss issues relating to the management and training of experts.

13.3.4.1 Classes of Participants

Each participant involved in the elicitation process should be clear about his/her role, the aims of the exercise and should be made aware of how his/her judgements will be used, i.e. how their data will inform particular decisions (Siu et al. 2015).

Who is an expert? Multiple definitions exist, with many focusing upon expertise, typically gained through experience in a particular field. For example, Ferrell (1994) defines an expert to be

a person with substantive knowledge about the events whose uncertainty is to be assessed

while Meyer and Booker (2001) define an expert as

a person who has a background in the subject matter at the desired level of detail and who is recognised by his/her peers or those conducting the study as being qualified to solve the questions.

These definitions implicitly assume experts can be accessed if required. When gathering experts from a constrained pool, then the definition by O'Hagan et al. (2006) might suffice since

an expert may, in principle, just mean the person whose judgements are to be elicited.

In addition to having expertise in the domain problem, we require an expert who can express their uncertainty accurately as a subjective probability. Being able to accurately assess uncertainty is not the same as being a subject matter expert. One may know less, but be more capable of expressing this degree of uncertainty quantitatively. Hora and Von Winterfeldt (1997) suggest six criteria for identifying experts: tangible evidence of expertise; reputation; availability and willingness to participate; understanding of the general problem area; impartiality; and lack of an economic or personal stake in the potential findings.

In order to elicit a wide spectrum of judgement, we may use a group of experts with diverse knowledge that encompasses all facets of scientific thought on a particular problem. This should help to identify areas of interest that may be missed with a small group of experts or with a group of experts from a specific school of thought. See, for example, Hogarth (1978), Clemen and Winkler (1986), Broomell and Budescu (2009), Larrick et al. (2011).

For further details, see Bolger (2018) who provides a detailed consideration of experts and their selection.

Beyond the expert, there are other participants in the elicitation process. Bedford et al. (2006) identify two additional roles of decision-maker and analyst. The decision-maker is the problem owner and the one who is ultimately responsible for any decision and wishes to be informed of the uncertainties that exist. The analyst is the person responsible for identifying the necessary experts, the events of interest and developing the elicitation protocol. Others describe similar additional roles. O'Hagan et al. (2006) make a distinction within the analyst role between that of a facilitator and a statistician. The facilitator manages the interaction with the experts and should be an expert in facilitation, while the statistician is an expert in probability and gives training to the experts, validates the results and provides feedback. However, O'Hagan et al. (2006) state that these roles can be merged. Booker and McNamara (2002) also identify the role of advisor-expert. An advisor-expert is someone who helps to support experts by offering technical support. This support may be in the area of identifying the appropriate experts or areas of interest that we wish to elicit judgement about. Within professional guidance, NUREG/CR-6372 identifies three different expert roles. A resource expert to present data, models and methods in an impartial manner. A proponent expert as an advocate for a specific model, method or parameter. An evaluator expert who will objectively examine available data and models, challenge technical bases, underlying assumptions and, where possible, test the models against observations.

13.3.4.2 Managed Experts

Assessments based on the aggregation of multiple experts' judgements are reported to be more accurate than predictions based on an individual's judgement (e.g. Page 2007; Soll and Larrick 2009). Therefore, we need to consider how experts should be managed, particularly their interaction. That is, should experts communicate and, if so, how?

The experts (as representatives of the informed technical community) will evaluate the available evidence (e.g. numeric data, models, theories and scientifically accountable positions) to inform their judgements. The selection of experts is to ensure a breadth of the collective state-of-knowledge. The extent to which experts should discuss the assessment of a quantity of interest varies by approach. Behavioural approaches such as Gosling (2018) advocate a facilitated discussion to arrive at the community probability distribution, while performance-based approaches as described in Quigley et al. (2018) propose experts form their assessments independently. Additionally, others propose hybrid approaches (Hanea et al. 2018; Hemming et al. 2018).

There is evidence in some contexts (Siu et al. 2015) that experts can be reluctant to quantify their beliefs as well as to share with fellow experts. Therefore, a process needs to consider the socio-technical nature of elicitation to help put experts at ease, thereby encouraging them to openly share their point of view, even if it is not shared by others. Challenges to proponent positions are important to enhance a group's understanding but need to be managed carefully.

When managing experts, there are three "i's" to consider when designing the process activities (Siu et al. 2015).

- (a) *Independence*—judgement should be based upon an individual's expertise; judgement should not be influenced by the organisation that the expert represents.
- (b) *Interaction*—if a behavioural aggregation approach is undertaken, then the process of evaluation, elicitation and integration is achieved through interaction amongst experts.
- (c) *Integration*—the process should emphasise integration (rather than consensus) of individuals' interpretations or judgement.

The advantage of a performance-based approach is the ability to discriminate between the quality of the experts' quantitative judgement through testing during the interview; this is not possible in a behavioural elicitation workshop.

13.3.4.3 Training and Learning

It is acknowledged by, amongst others, Bonano et al. (1990), Keeney and Von Winterfeldt (1991), US Nuclear Regulatory Commission (1997) that both an expert's willingness to provide numerical estimates and the quality of their assessments can be improved through training. Such training should explain the meaning of subjective

probability, raise awareness of well-known sources of bias and provide *meaningful* exercises on which to practice. To be meaningful to engage the experts, such exercises should align with the problem under investigation.

Bonano et al. (1990) suggest three key tasks be conducted during the training session: First to familiarise the expert with the process and motivate them to provide formal judgements. Second to provide experts with practice at expressing their judgement. Third to educate the experts on potentials for bias.

The quality of subjective probabilities from experts is dependent on both the expert's experience and the method of elicitation. If the expert lacks experience, the prior distributions will be uninformative or misleading, regardless of the elicitation approach employed. Poorly designed elicitation techniques may degrade the quality of information provided by experts. Fischhoff (1989) proposes the following four necessary conditions to support improving judgement skills.

- (a) *Abundant practice with a set of reasonably homogeneous tasks*—to assist the expert in developing their judgemental skills on the relevant task.
- (b) *Clear-cut criterion events for outcome feedback*—learning requires feedback to the expert, but this can be challenging to evaluate if the judgements are components of complex systems (natural, social or biological).
- (c) *Task-specific reinforcement*—performance should be based on the wisdom of their judgement; be aware if there are implicit rewards for the experts. e.g. did they bring good news? Did they disrupt plans?
- (d) *Explicit admission of the need for learning*—using titles such as expert can result inhibit learning.

Fischhoff (1989) also points out that often judgements concern events that are not realised for years, which provide little opportunity to learn about the quality of such judgements.

13.3.5 *Process Considerations*

The elicitation process is more than the means by which the method to obtain the probability assessments is implemented with the selected experts, say by an interview or some other means. We examine issues that are important in creating a coherent process that allows design choices about the probability and people aspects, as discussed in the previous sections, to be meaningfully planned and implemented.

13.3.5.1 *Core Activities*

The process should account for key activities that add value to the quality of the data collected. Such activities include the recruitment of experts, the framing of questions, the elicitation and aggregation of their judgements, using procedures that have been

tested and clearly demonstrated to improve judgements (e.g. Cooke 1991; Mellers et al. 2014).

In particular, the following activities are core to the process.

- (a) *Preparation*—this will entail the development of the following: problem statement, project plan, expert panel, reading material, package of data available and elicitation procedures.
- (b) *Pilot study/Expert training*—it is essential that all experts share a common understanding of the problem and the specific quantities to be estimated, as well as being trained in using probability. Moreover, the intended use of the outcomes, the elicitation process and the participants' roles need to be explained.
- (c) *Expert elicitation*—depending on the approach undertaken, this could be in the form of a group workshop or individual interviews.
- (d) *Combine judgements*—depending on the approach undertaken, this could be during the group workshop through interaction or by the analyst following all interviews.
- (e) *Feedback*—to all experts.
- (f) *Documentation*—participation needs to be appropriately documented, specifically which experts were involved in assessing which quantities as it would be misleading to identify a panel of experts and the resulting assessment only.

13.3.5.2 Tactics for Sound Process Management

Providing guidance on the underpinning reasons for each activity in the process allows the analyst to make better informed design decisions. Hence, explication of the process logic and the role of each activity is important because, if not, then users might approach the process rather superficially through lack of detailed understanding and so inadvertently introduce substantial variations in the elicitation outcomes.

Elicitation processes are lengthy and require the expert to concentrate for a considerable length of time, which can result in compromising the level of accuracy in the elicited probability (Shephard and Kirkwood 1994). The process design should manage experts so that they spend a greater fraction of time on issues of greatest uncertainty. This will avoid a common tendency of spending time on aspects of the problem where data exist and the problem is well understood. Having experts document and bring their written rationales to the elicitation will facilitate the clarification of substantive issues and reduce time (Cooke and Goossens 1999).

13.3.5.3 Checking

The analyst should perform credibility checks to ensure that the probability assessments provided are consistent with an expert's beliefs. This means not only the elicited values, but also the implications of how the analyst is interpreting the judge-

ments by having the expert reflect both the underlying quantity of interest and also the data that will be realised (Keeney and Von Winterfeldt 1991).

When assessing multiple quantities of interest, it is important to check for trends across the values for each to determine if there are any indicators of anchoring and adjustment bias (Siu et al. 2015).

It is also possible to include checks within and beyond the time frame of the elicitation to estimate the predictive accuracy of judgemental probability assessments of uncertainties. For example, “test” quantities of interest for which realisations will be obtained within the time frame of the elicitation provide a means to understand the degree to which an expert is calibrated (Anderson et al. 2015), while having a forward-looking activity to monitor and record any realisations of the quantities of interest enables empirical control, even if only in principle.

13.4 Comparison of Two Elicitation Processes

There are several guidance documents for elicitation processes from a variety of professional or academic sources available in the public domain. We consider the guidance on elicitation processes produced by the European Food Safety Authority (EFSA) in 2014 and the Institute and Faculty of Actuaries (IFoA) in 2015. We select these because they are examples of practice that allow us to illustrate the diversity in application domain as well as the variation in the scope, repetitiveness and level of process prescription. After summarising the salient elements of the guidance for these two processes, to the level expressed by the respective documents, we compare their characteristics in relation to those discussed in Sect. 13.3.

13.4.1 *European Food Safety Authority (EFSA) Guidance*

The European Food Safety Authority (2014) has developed a detailed process that also includes procedures for expert judgement elicitation within a project, as shown in Fig. 13.3. EFSA is responsible for food safety risk assessment in Europe and operates independently of European legislative, executive institutions and EU Member States. Hence, it is separate from risk management or policymaking. EFSA is a regulator and so deals with expert problems or, occasionally, textbook problems (Hartley and French 2018).

The EFSA process comprises three main phases—initiation, pre-elicitation and elicitation—which are each managed by a different group—working, steering and elicitation group, respectively.

The Working Group defines the problem and justifies the need for an elicitation. This first step requires consideration of all of the relevant model parameters, and to determine which require expert elicitation and which do not. Thus the Working Group prepares a document of the background information.

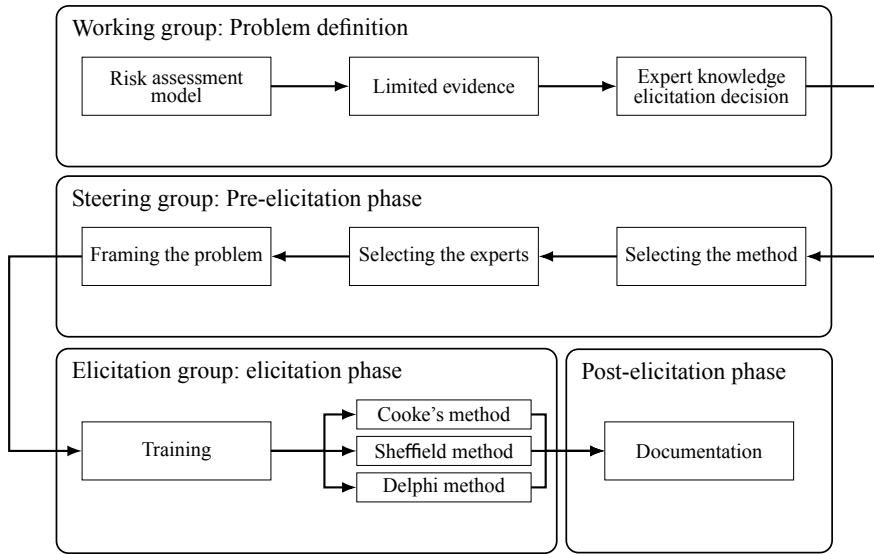


Fig. 13.3 Key phases of the EFSA elicitation process. (Adapted from European Food Safety Authority 2014)

The Steering Group can be a subset of the Working Group and will comprise scientists, experts on elicitation and administrative staff. Their remit is to plan the elicitation process by designing the elicitation protocol. This group specifies the questions suitable for expert elicitation, defines expert profiles and selects the experts and elicitation method as well as the Elicitation Group. Procedures are given for three elicitation methods: the Classical Model, which EFSA calls Cooke’s method, Sheffield and Delphi methods which we outline below.

The Elicitation Group typically comprises one or two elicitors with additional administrative support who are familiar and experienced with the selected elicitation protocol. All direct contacts with the experts are made by the Elicitation Group, so members should have a neutral position on the elicitation question. To enhance trust and guarantee confidentiality in ambiguous or conflictive situations, the Elicitation Group should be independent of all parties involved. This group is responsible for executing the elicitation method as well as providing training for the experts.

The evidence dossier is a key part of the guidance to capture the evidence regarding each quantity of interest to be elicited. Expert judgement should not differ because experts have access to different data; difference in opinion should be due to different expertise and interpretation of data. Therefore, data to which the experts have access should be documented and shared. Such documentation should not be too large since it challenges the experts in assimilating all the evidence as well as pointing out the weakness (e.g. small sample sizes), and it can also make the expert anchor on the provided evidence and fail to consider counter facts. The documentation should include any new evidence submitted by experts prior to the elicitation.

Documentation is made public since EFSA upholds the three principles of repeatability, transparency and confidentiality. Three types of report are produced. The result report to summarise the findings; a technical support document to detail a full description of the process and its execution; and expert feedback which is a confidential summarising the input from each expert.

Disclosing personal data that might identify individual experts with their judgements is neither an objective of the process nor necessary to fulfil transparency requirements and may discourage experts from taking part in the process or influence their responses. Participating experts are assured on the confidential treatment of their individual answers, where reports will include who took part, what was said but not who said it.

The Sheffield method is a behavioural aggregation method, where experts participate in a facilitated workshop to create a subjective probability distribution for each quantity of interest. Once the training session has been conducted, the workshops progress through four stages for each quantity of interest. An initial review of evidence is followed by each expert individually assessing their judgement on the quantity. These individual judgements are shared and discussed amongst the group. Aspects of individuals' distributions which are different are discussed within the group and rationales elicited. Then the group judgement is formed as one distribution to represent the view of the rational observer. See Gosling (2018) for a detailed description.

The version of the Delphi method included in the EFSA guidance uses pools of experts but, to minimise adverse group effects, it restricts interpersonal interaction by controlling the flow of information. Experts do not meet, instead they exchange their beliefs and assessments through the facilitator. The facilitator summarises the group's views to the experts and invites each to revise their judgements. See European Food Safety Authority (2014) for a detailed description.

The Classical Model is a performance-based method, where experts work with the analyst independently and without interaction with other experts to assess the uncertainty in the unknown quantity of interest as well as for other variables for which the answer is known to the analyst but not the expert, known as seed questions. Seed questions provide an opportunity of assessing the quality of the responses provided by the experts. See Cooke (1991), Quigley et al. (2018) for more details.

13.4.2 Institute and Faculty of Actuaries (IFoA) Working Paper

The IFoA is the only UK chartered professional body dedicated to educating, developing and regulating actuaries based both in the UK and internationally. Actuaries serve the public interest by conducting analysis where there is uncertainty of future financial outcomes. Solvency II is an EU Directive that came into effect on 1 January 2016 and primarily concerns the risk of insolvency of EU insurance compa-

nies. The associated judgement by actuaries in applying the principles of Solvency II prompted a working party for the IFoA to present a paper providing a practical framework regarding expert judgement processes, including their validation, for repeated assessment of risk (Ashcroft et al. 2016). The views expressed in the publication are not necessarily those of the IFoA.

A key motivation for the paper was a lack of transparency on the use of expert judgement within the profession, which is one where judgements have significant impacts on risk assessments and subsequent decisions taken. The authors consider knowledge to be socially constructed so that common judgement can be created through mediation of experiences and ideas. As such, their process is designed to facilitate the pooling of experience and ideas, and not necessarily in consensus.

What we shall label as the IFoA process has five key stages, as shown in Fig. 13.4, and which we discuss below.

First, there is a preliminary assessment to determine whether a formal expert judgement process is relevant. This involves considering whether the nature of the judgement is within the scope of an expert judgement process.

Second, the problem is defined. The problem is articulated and its scope is defined. The current level of understanding of the problem is determined to develop an expert brief. Terminology should be made clear to ensure a consistent interpretation of the problem, which is especially important if using external experts. Potential experts

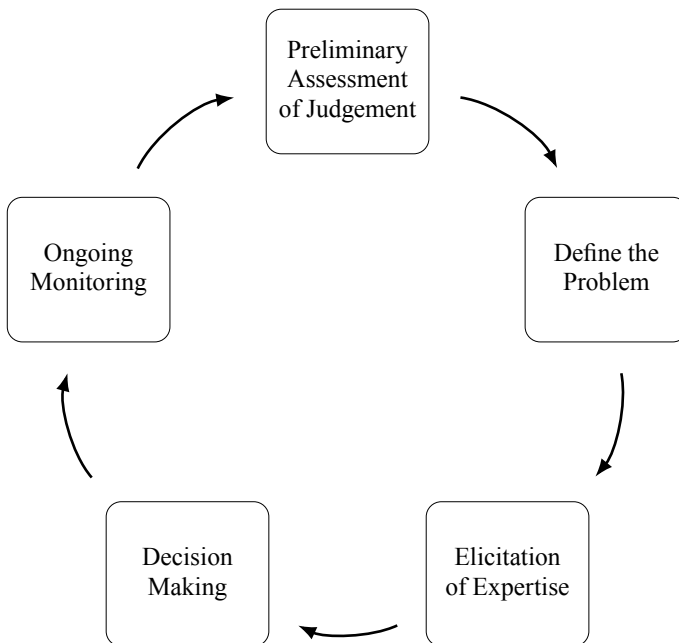


Fig. 13.4 Key stages of IFoA elicitation process. (Adapted from Ashcroft et al. 2016)

are identified, and an initial plausible range of the values of the quantity of interest are assessed.

Third, elicitation of expertise is designed and conducted. The method for elicitation is chosen and will depend upon the nature and importance of the problem; this will include the methods for both encoding the quantities of interest and for combining the views across experts. Documentation is required to describe the available data, assumptions, principles, methodologies and models applied in arriving at a recommendation and on any potential limitations.

Fourth is the decision-making. The decision-makers should review all information (which might include confidential data not known to the experts) and expert judgements to ensure consistency. Decision-makers should set out their thought process clearly, explicating how they are making use of the expert judgement, making clear how they weight the relative importance of information and identify triggers for non-scheduled reviews. This practice is intended to help facilitate a multi-layer governance structure through transparency. A final decision on the judgemental assessment of the quantity of interest is recommended; an overall plausible range, and a summary of the rationale for this, should be communicated back to the experts. This provides an opportunity to flag any serious concerns they may have and which can then be fed back to the decision-makers.

Fifth, there is on-going monitoring. A robust system should be created to monitor the validity of the probability assessment, reflecting on scope of its application, appropriateness of assumptions and triggers for review.

13.4.3 Comparison

The guidance provided by EFSA and the paper from the IFoA naturally differs due to the distinct nature of the problems addressed by the two organisations. EFSA has developed more detailed guidelines aligned with their own organisational need, while IFoA provides higher level guidance to be used by various insurance companies. Conceptually, the processes advocated by both organisations have elements in common, such as problem structuring, an initial evaluation and probability assessment. Since the nature of the problem addressed by the IFoA is on-going, it explicitly continues monitoring after the initial probability assessment, unlike EFSA which assumes a one-off project.

We now compare the two documents in relation to the characteristics of an elicitation process identified in Sect. 13.3.

Principles: Neither process explicitly contradicts any of the principles, but each document supports the principles in varying degrees. Both processes include detailed discussion on documentation and governance. Both allow for processes where expert assessment could be falsifiable with further data; this is either explicitly stated as a goal or implicit in the construction of the elicitation question. The EFSA guidance explicitly states neutrality as a required feature of an elicitation process but how

this is ensured is not stated, while the IFoA only mentions the need to manage bias. Fairness is implicit in both processes.

Purpose: Both documents provide a clear statement of purpose. The IFoA has developed guidance for a specific purpose, while EFSA has developed guidance for use in a variety of projects within their remit. EFSA has more clearly identified groups with associated responsibilities of the process. The IFoA acknowledges that multi-layer governance structure may vary by institution. Both emphasise the needs for documentation of the elicitation process.

Observable quantity of interest: The EFSA guidance explicitly requires quantities of interest to be observable in principle, whereas the IFoA does not mention this.

Selection of quantity of interest: Both processes advise on the use of data as well as consider an initial assessment of the uncertainty associated with the quantities of interest. These can then be used in a sensitivity analysis prior to conducting the subjective probability assessment by informing prioritisation of the variables to quantify.

Method of encoding: Only the EFSA process explicitly provides guidance on probability distribution fitting with appropriate checks in place and that non-parametric approaches are also available.

Classes of participants: While the EFSA elicitation of the expert's judgement is led by what they term an elicitor, the IFoA considers various formal roles that need to be fulfilled such as the decision-maker, coordinator and validator in addition to the expert. They differentiate between internal and external experts depending upon whether the expert works for the organisation making the risk assessment.

Managed experts: The IFoA process does not provide detailed guidance on the approach to managing groups of experts to the same extent as the EFSA process. The IFoA suggests one may use Delphi or Nominal Group Technique but provides no guidance on the management of interaction. In contrast, the EFSA guidance provides choice depending on whether or not interaction is desired.

Expert training and learning: Both processes acknowledge that some experts will require training in expressing subjective probabilities as well as explicitly requiring that expert feedback is given. However, the nature of the assessments being made implies that meaningful feedback on the predicted assessments is not considered.

Core process activities: Only the EFSA process provides guidance at the level of detail described in Sect. 13.3.5.1.

Tactics for process management: EFSA provides guidance on identifying and managing elicitation fatigue by experts, while the IFoA advises on efficient structuring the elicitation questions to address this issue.

Checking: Both processes provide guidance on feedback to experts as well as validation of their probability assessments. The IFoA process only requires that this be documented but it does not provide guidance on how to validate expert judgement,

whereas the EFSA process states that “validation requires eliciting uncertainty on variables whose true values will be known within the time frame of the study” (p. 159 European Food Safety Authority 2014).

13.5 The Value of an Elicitation Process Standard?

Following our comparison of practical guidance, and in light of our abstraction of issues emergent from the literature, we now explore whether it is meaningful to characterise a standard process for elicitation.

Standards represent a voluntary acceptance of the rules. Interestingly, the creation of international standardisation bodies, such as ISO, is grounded in the need to answer the question “what is the best way of doing this?”¹

The UK national standard body, BSI, identifies three important general drivers for the creation of standards: First, that a standard represents “an agreed way of doing something”. Second, that a standard is “the distilled wisdom of people with expertise in their subject matter and who know the needs of the organisations they represent”. Thirdly, the “point of a standard is to provide a reliable basis for people to share the same expectations about a ...” process. See BSI (2018).

We can frame such drivers as the characteristics of a process (or service or product or technology and so on) that has reached sufficient maturity to be standardised. Specifically, we ask whether elicitation processes for the assessment of uncertainty in a quantity of interest have reached such maturity that standardisation would be valuable, and if yes, then how this might be achieved?

We have compiled a set of characteristics of a good elicitation process that are recognised in the literature and correspond to features of an elicitation process that embrace, but also extend beyond, the core scientific principles of Cooke (1991). Further, we have examined the pivotal role of the SRI process in providing a genesis for later, more bespoke elicitation processes. While the latter might emphasise distinct process elements, this might partly be a function of, for example, the distinct purpose of the process in the wider modelling context, the disciplinary bias of the process or method creator and the problem domain in which the process might be applied. By tracing the relationships between features of the leading modern elicitation processes and the SRI in relation to the characteristics of a good elicitation process, we have shown that there is indeed considerable agreement in the way to approach the development of a good elicitation process.

There already exist many guidance documents for elicitation processes and procedures. We have only examined two. Both the EFSA and the IFoA documents are grounded in the wisdom of people with expertise in designing and implementing elicitation exercises as well as experience in understanding the needs of the organisation(s) who will use the elicitation in context. The coverage of such guidance is a function of the scope of the elicitation and the selection of people who have

¹<https://www.iso.org/benefits-of-standards.html> [accessed on 20 December 2018].

contributed authorship. The process of creating the guidance documents, of course, will influence the content. Given that there is no guidance yet created by professional standardisation bodies, with all the balances and checks that they deploy in recruiting experts and consensus-forming practices, any elicitation guidance currently in the public domain is subject to the manner in which the commissioning body has procured the guidance, although, having said that, in commissioning guidance there is an implicit intent to provide a reliable basis for people to share the same expectations about the elicitation process.

Following this line of argument, it appears that elicitation processes have reached a state of maturity generally associated with standard creation. Specifically, following Swann and Lambert (2017), we class an elicitation standard as primarily *informative* because they codify process knowledge. This is in contrast to standards that might be classed as primarily *constraining*, such as health and safety. But, even if the intent is to codify and share knowledge to enhance best practice, what are the pros and cons of elicitation process standards? Table 13.1 summarises some key points which we believe are important.

There exist other established standards which achieve the same goals of guiding users in developing, implementing and documenting processes that we might seek to achieve with an elicitation standard. If required by regulation or by contract then such standards can also offer protection to users. Given the recent legal consequences arising in relation to the use of expert assessments of uncertainty as discussed in our Introduction, this aspect might be particularly relevant and novel for elicitation.

There are, of course, mitigations that might help to remove or to reduce the effects of the negative aspects of a standard. For example, much research has been conducted in the relationship between standards and innovation more generally (Blind 2013) with some lessons being applicable for elicitation. Findings show that informative standards can enable, rather than inhibit, innovation within the user organisations given the sharing of codified knowledge. However, the mechanisms for maintaining standards through a formal review and revision process is required to ensure up-to-date guidance. While official standards bodies are empowered to provide such infrastructure, it is not always evident that it occurs within all specific domain bodies. A common concern with standards is that users treat process guidance as a defined procedure rather than think meaningfully about the translation of a process guidance to the specific context. Already elicitation guidance documents have been crafted

Table 13.1 Pros and cons of a standard elicitation process

Pros	Cons
Enhance the craft for those process elements that can be identified a priori	Constrain process innovation to embrace new elicitation knowledge
Increase rigour of process implementation and reduce susceptibility to poor practice	Limit responsibility of the user to think deeply about the specific elicitation
Provide protection to process participants	Make process accessible to poor facilitators

to support better thinking rather than to supplant thinking. However, crafting such guidance is challenging especially if approached at a more general level. As for other process standards, providing guidance on making choices about key activities, such as the selection and definition of the quantities of interest, can be more difficult than giving advice on standard components of documentation, simply because the former is so contingent on the complexity of the modelling problem while the latter is relatively transferrable between applications.

At present, the prevalence of domain elicitation guidance implies that choice of process facilitator is within the control of the commissioning organisations. If there is an elicitation standard, then there may be a growth in the facilitator market meaning less reliance on a smaller pool of knowledgeable facilitators who have earned trust. Creating some form of elicitation facilitator certification might mitigate this risk.

The suggested mitigations tend to rely upon the formalities of a recognised body with responsibility for producing standards as documents established by consensual process. Such bodies already provide standards in other areas of data collection scoped to interface with user needs. There are reduced degrees of standardisation in that recognised bodies also provide technical reports which do allow for sharing of codified knowledge that is informative only. An official standard will contain normative as well as informative text. There is increasing attention to open standards (Maxwell 2006) which are a means to give users permission to use “technology” freely without the involvement of a recognised body in the creation of the standard.

We have established that the practice of elicitation process design and implementation has reached a degree of maturity that allows standard codification of knowledge and we have explored some options regarding creation of a standard for a process for eliciting subjective probability assessments. However, we leave it to the reader to decide whether creating such a standard for an elicitation process would be a valuable endeavour and if, so, in what form.

13.6 Concluding Discussion

We have examined the characteristics of a sound process to elicit judgement to ensure good quality of data to inform decision-making under uncertainty. Even in the contemporary digital world, there is continued need for subjective probability assessments for problems where observed data are non-existent or limited, as well as in situations where observed data are abundant since the relevance of the past to the future need to be assessed with expertise. By exploring the evolution of elicitation processes temporally and across a variety of distinctive problem domains, we have synthesised the characteristics underpinning a good elicitation process—these encompass the probabilistic as well as people aspects of developing a process that aligns with the problem purpose. Such characteristics, and their illustration for elicitation guides produced by two professional organisations, provide a collection of attributes to which a good elicitation should aspire as well as some practical pitfalls

to beware. Our goal has been to highlight elicitation process characteristics that are sufficiently general to be widely applicable.

A defensible elicitation process can provide protection to those accountable for the consequences of the determined actions. Appropriate levels of accountability can increase trust in risk information (Frewer et al. 1996). The Cambridge dictionary defines accountability as

a situation in which someone is responsible for things that happen and can give a satisfactory reason for them

and responsible as

to have control and authority over something or someone and the duty of taking care of it, him, or her.

As such, a sound elicitation process should produce a satisfactory reason for its results for the person with the duty of care. How satisfactory the reason provided for the subjective probability judgements will depend upon the problem and the associated stakeholders, so guidance will vary in detail across domains. Puig et al. (2018) highlight that a lack of accountability mechanisms are in place to ensure national governments rely on scientifically sound processes for the appropriateness of their forecasting. However, while important, the responsibility does not just rest with the end user. Since some elicitation processes will involve assessing multiple uncertainties by various experts, ensuring each participant is clear about their role in the process should produce accountability; each participant has a responsibility for their contribution. The L'Aquila tragedy, discussed in the Introduction, led to experts being initially held accountable for poor practice and superficial analysis. Of course, a sound process does not guarantee immunity from criticism as other factors will play a role in this social process. For example, Pidgeon (1997) argues that

despite the inherent complexity and ambiguity of the environments within which large-scale hazards arise and the systemic nature of breakdowns in safety, cultural myths of control over affairs ensure that a culprit must be found after a disaster or crisis has unfolded.

So the responsible should have their reason prepared.

Acknowledgments We would like to thank David Purves and Ian Belton, as well as the editors, for their critical feedback and suggestions.

References

- Anderson, G., Walls, L., Revie, M., Fenelon, E., & Storie, C. (2015). Quantifying intra-organisational risks: An analysis of practice-theory tensions in probability elicitation to improve technical risk management in an energy utility. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 229(3), 171–180.
- Armstrong, J. S., Denniston, W. B, Jr., & Gordon, M. M. (1975). The use of the decomposition principle in making judgments. *Organizational Behavior and Human Performance*, 14(2), 257–263.

- Ashcroft, M., Austin, R., Barnes, K., MacDonald, D., Makin, S., Morgan, S., et al. (2016). Expert judgement. *British Actuarial Journal*, 21(2), 314–363.
- Ashton, A. H., & Ashton, R. H. (1985). Aggregating subjective forecasts: Some empirical results. *Management Science*, 31(12), 1499–1508.
- Aspinall, W., & Cooke, R. (2013). *Quantifying scientific uncertainty from expert judgement elicitation*. In: *Risk and uncertainty assessment for natural hazards*. Cambridge University Press Cambridge, UK, p 64.
- Astfalck, L., Cripps, E., Gosling, J., Hodkiewicz, M., & Milne, I. (2018). Expert elicitation of directional metocean parameters. *Ocean Engineering*, 161, 268–276.
- Ayyub, B. M. (2001). *Elicitation of expert opinions for uncertainty and risks*. CRC Press.
- Barons, M. J., Hanea, A. M., Wright, S. K., Baldock, K. C., Wilfert, L., Chandler, D., et al. (2018). Assessment of the response of pollinator abundance to environmental pressures using structured expert elicitation. *Journal of Apicultural Research*, 57(5), 593–604.
- Bedford, T., Quigley, J., & Walls, L. (2006). Expert elicitation for reliable system design. *Statistical Science*, 428–450
- Blind, K. (2013). The Impact of Standards on Innovation, NESTA Report. Retrieved 25 Jan, 2019, from http://www.innovation-policy.org.uk/share/14_The%20Impact%20of%20Standardization%20and%20Standards%20on%20Innovation.pdf.
- Bolger, F. (2018). *The selection of experts for (probabilistic) expert knowledge elicitation*. In *Elicitation*, Springer, pp. 393–443.
- Bolger, F., & Wright, G. (1992). *Expertise and decision support*. Springer Science & Business Media.
- Bonano, E. J., Hora, S., Keeney, R., & Von Winterfeldt, D. (1990). Elicitation and use of expert judgment in performance assessment for high-level radioactive waste repositories. Tech. rep., Nuclear Regulatory Commission, Washington, DC (USA). Div. of High-Level.
- Booker, J., & McNamara, L. (2002). Expertise and expert judgment in reliability characterization: A rigorous approach to eliciting, documenting and analyzing expert knowledge.
- Brockhoff, K. (2002). The performance of forecasting groups in computer dialogue and face-to-face discussion. In: *The Delphi method: Techniques and applications*, <http://www.is.njit.edu/pubs/delphibook/delphibook.pdf>, pp 285–311.
- Broomell, S., & Budescu, D. (2009). Why are experts correlated? decomposing correlations between judges. *Psychometrika*, 74(3), 531–553.
- BSI. (2018). Information about standards. Retrieved 20, Dec 2018, from <https://www.bsigroup.com/en-GB/standards/Information-about-standards>.
- Budnitz, R. J., Apostolakis, G., Boore, D. M., Cluff, L. S., Coppersmith, K. J., Cornell, C. A., et al. (1998). Use of technical expert panels: applications to probabilistic seismic hazard analysis. *Risk Analysis*, 18(4), 463–469.
- Burgman, M. (2004). Expert frailties in conservation risk assessment and listing decisions. In: P. A. Hutchings, D. Lunney, C. R. Dickman (Eds.) *Threatened Species Legislation: is it just an act?*, pp. 20–29.
- Chaloner, K., Church, T., Louis, T. A., & Matts, J. P. (1993) Graphical elicitation of a prior distribution for a clinical trial. *The Statistician*. pp. 341–353.
- Choy, S. L., O’Leary, R., & Mengersen, K. (2009). Elicitation by design in ecology: using expert opinion to inform priors for Bayesian statistical models. *Ecology*, 90(1), 265–277.
- Christophersen, A., Deligne, N. I., Hanea, A. M., Chardot, L., Fournier, N., & Aspinall, W. P. (2018). Bayesian Network modeling and expert elicitation for probabilistic eruption forecasting: Pilot study for Whakaari/White Island New Zealand. *Frontiers in Earth Science*, 6, 211.
- Clemen, R., & Winkler, R. (1986). Combining economic forecasts. *Journal of Business and Economic Statistics*, 41(1), 39–46.
- Cooke, R. (1991). *Experts in uncertainty: Opinion and subjective probability in science*. Oxford University Press on Demand.
- Cooke, R., & Jager, E. (1998). Failure frequency of underground gas pipelines: methods for assessment with structured expert judgment. *Risk Analysis*, 18(4), 511–527.

- Cooke, R. M., & Goossens, L. (1999). *Procedures guide for structured expert judgment* (p. 18820). EUR: Project Report to the European Commission.
- Cooke, R. M., & Goossens, L. L. (2008). TU delft expert judgment data base. *Reliability Engineering & System Safety*, 93(5), 657–674.
- Cox, L. A., Jr. (2012). Confronting deep uncertainties in risk analysis. *Risk Analysis: An International Journal*, 32(10), 1607–1629.
- Dalkey, N. (1967). *Delphi (Report P-3704)*. Santa Monica, CA: Rand Corporation.
- Dalkey, N., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management Science*, 9(3), 458–467.
- Dalkey, N. C. (1969). The Delphi method: An experimental study of group opinion. Tech. rep., The RAND Corporation, Santa Monica, CA (No. RM-5888-PR).
- Dias, L. C., Morton, A., & Quigley, J. (2018). Elicitation: State of the Art and Science. In A. Morton & J. Quigley (Eds.), *Dias LC* (pp. 1–14). Springer: Elicitation.
- European Food Safety Authority. (2014). Guidance on expert knowledge elicitation in food and feed safety risk assessment. *EFSA Journal*, 12(6), 3734.
- Ferrell, W. R. (1985). *Combining individual judgments*. In: Behavioral decision making, Springer, pp. 111–145.
- Ferrell, W. R. (1994). *Discrete subjective probabilities and decision analysis: Elicitation*. Calibration and Combination: Wiley.
- Fischhoff, B. (1989). Eliciting knowledge for analytical representation. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(3), 448–461.
- French, S. (2012). Expert judgment, meta-analysis, and participatory risk analysis. *Decision Analysis*, 9(2), 119–127.
- Frewer, L. J., Howard, C., Hedderley, D., & Shepherd, R. (1996). What determines trust in information about food-related risks? Underlying psychological constructs. *Risk Analysis*, 16(4), 473–486.
- Frijters, M., Cooke, R., Slijkuis, K., & van Noortwijk, J. (1999). *Expert judgment uncertainty analysis for inundation probability*. Utrecht: Ministry of Water Management, Bouwdienst, Rijkswaterstaat.
- Gigerenzer, G., & Edwards, A. (2003). Simple tools for understanding risks: From innuery to insight. *BMJ*, 327(7417), 741–744.
- Gosling, J. P. (2018). SHELF: the Sheffield elicitation framework. In A. Morton & J. Quigley (Eds.), *Dias LC* (pp. 61–93). Springer: Elicitation.
- Hampton, J., Moore, P., & Thomas, H. (1973). Subjective probability and its measurement. *Journal of the Royal Statistical Society Series A (General)*, 21–42
- Hanea, A. M., Burgman, M., & Hemming, V. (2018). IDEA for uncertainty quantification. In A. Morton & J. Quigley (Eds.), *Dias LC* (pp. 95–117). Springer: Elicitation.
- Hartley, D., & French, S. (2018). Elicitation and calibration: A Bayesian perspective. In A. Morton & J. Quigley (Eds.), *Dias LC* (pp. 119–140). Springer: Elicitation.
- Hemming, V., Burgman, M. A., Hanea, A. M., McBride, M. F., & Wintle, B. C. (2018). A practical guide to structured expert elicitation using the IDEA protocol. *Methods in Ecology and Evolution*, 9(1), 169–180.
- Hodge, R., Evans, M., Marshall, J., Quigley, J., & Walls, L. (2001). Eliciting engineering knowledge about reliability during design-lessons learnt from implementation. *Quality and Reliability Engineering International*, 17(3), 169–179.
- Hogarth, R. (1978). A note on aggregating opinions. *Organizational Behavior and Human Performance*, 21, 40–46.
- Hora, S. C. (2007). Eliciting probabilities from experts. *Advances in decision analysis: From foundations to applications* 129.
- Hora, S. C., & Von Winterfeldt, D. (1997). Nuclear waste and future societies: A look into the deep future. *Technological Forecasting and Social Change*, 56(2), 155–170.
- Janis, I. L. (1971). Groupthink. *Psychology Today*, 5(6), 43–46.
- Junger, S. (1997). *The perfect storm: A true story of men against the sea*. WW Norton & Company.

- Kadane, J., & Wolfson, L. J. (1998). Experiences in elicitation. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(1), 3–19.
- Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases* (pp. 3–20). Judgement under uncertainty: Heuristics and biases.
- Kammerer, A. M., & Ake, J. P. (2012). *Practical implementation guidelines for SSHAC Level 3 and 4 hazard studies*. Office of Nuclear Regulatory: United States Nuclear Regulatory Commission.
- Keeney, R. L., & Von Winterfeldt, D. (1991). Eliciting probabilities from experts in complex technical problems. *IEEE Transactions on Engineering Management*, 38(3), 191–201.
- Krueger, T., Page, T., Hubacek, K., Smith, L., & Hiscock, K. (2012). The role of expert opinion in environmental modelling. *Environmental Modelling & Software*, 36, 4–18.
- Kuhnert, P. M., Martin, T. G., & Griffiths, S. P. (2010). A guide to eliciting and using expert knowledge in Bayesian ecological models. *Ecology Letters*, 13(7), 900–914.
- Larrick, R. A. E. M., B. S. J. (2011). The social psychology of the wisdom of crowds. In: Ji, K. (Ed.) *Frontiers of social psychology: Social psychology and decision making*, New York: Psychology Press.
- Lorenz, J., Rauhut, H., Schweitzer, F., & Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences*, 108(22), 9020–9025.
- Martelli, A., & Mualchin, L. (2012). Indictment and conviction of members of the Italian “Commissione Grandi Rischii” (open letter to the President of Italy). Retrieved 8, June 2017, from www.angeologi.it/wp-content/uploads/2012/10/CoverletterandStatementSSO1.pdf.
- Mastrandrea, M. D., Field, C. B., Stocker, T. F., Edenhofer, O., Ebi, K. L., Frame, D. J., Held, H., Kriegler, E., Mach, K. J., & Matschoss, P. R. et al. (2010). Guidance note for lead authors of the IPCC fifth assessment report on consistent treatment of uncertainties. Intergovernmental Panel on Climate Change.
- Maxwell, E. (2006). Open standards, open source and open innovation; harnessing the benefits of openness. *Innovations: Technology, Governance, Globalization*, 1(3), 119–176.
- Mellers, B., Ungar, L., Baron, J., Ramos, J., Gurcay, B., Fincher, K., et al. (2014). Psychological strategies for winning a geopolitical forecasting tournament. *Psychological Science*, 25(5), 1106–1115.
- Merkhofer, M. W. (1987). Quantifying judgmental uncertainty: Methodology, experiences, and insights. *IEEE Transactions on Systems, Man, and Cybernetics*, 17(5), 741–752.
- Meyer, M., & Booker, J. (2001). Eliciting and analyzing expert judgment: A practical guide. *American Statistical Association and Society for Industrial and Applied Mathematics*. Philadelphia.
- Miller, A. C. I. I. I., & Rice, T. R. (1983). Discrete approximations of probability distributions. *Management Science*, 29(3), 352–362.
- Moore, C. M. (1987). *Group techniques for idea building*. Sage Publications, Inc.
- Myers, D. G., & Lamm, H. (1976). The group polarization phenomenon. *Psychological Bulletin*, 83, 602–627.
- Nature. (2011). Scientists on trial: at fault? Retrieved 8 June 2017, from www.nature.com/news/2011/110914/full/477264a.html.
- O’Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J. et al. (2006). *Uncertain judgements: eliciting experts’ probabilities*. Wiley.
- Page, S. E. (2007). Making the difference: Applying a logic of diversity. *Academy of Management Perspectives*, 21(4), 6–20.
- Pidgeon, N. (1997). The limits to safety? Culture, politics, learning and man-made disasters. *Journal of Contingencies and Crisis Management*, 5(1), 1–14.
- Puig, D., Morales-Nápoles, O., Bakhtiari, F., & Landa, G. (2018). The accountability imperative for quantifying the uncertainty of emission forecasts: Evidence from Mexico. *Climate Policy*, 18(6), 742–751.
- Quigley, J., & Walls, L. (2010). Reconciling experts opinion concerning the value of testing using Bayesian networks: A bridge too far? In *5th International ASRANet Conference*.
- Quigley, J., & Walls, L. (2018). A Methodology for Constructing Subjective Probability Distributions with Data. In A. Morton & J. Quigley (Eds.), *Dias LC* (pp. 141–170). Springer: Elicitation.

- Quigley, J., Colson, A., Aspinall, W., & Cooke, R. M. (2018). Elicitation in the classical model. In A. Morton & J. Quigley (Eds.), *Dias LC* (pp. 15–36). Springer: Elicitation.
- Regan, T. J., Burgman, M. A., McCarthy, M. A., Master, L. L., Keith, D. A., Mace, G. M., et al. (2005). The consistency of extinction risk classification protocols. *Conservation Biology*, *19*(6), 1969–1977.
- Sackman, H. (1975). Delphi critique. Expert Opinion, Forecasting. Group Process NY: Lexington Books, pp. 30–50.
- Science. (2012). Earthquake experts convicted of manslaughter. Retrieved 8 June 2017, from www.sciencemag.org/news/2012/10/earthquake-experts-convicted-manslaughter.
- Science. (2014). Updated: Appeals court overturns manslaughter convictions of six earthquake scientists. Retrieved 8 June 2017, from www.sciencemag.org/news/2014/11/updated-appeals-court-overturns-manslaughterconvictions-six-earthquake-scientists.
- Seaver, D. A., Von Winterfeldt, D., & Edwards, W. (1978). Eliciting subjective probability distributions on continuous variables. *Organizational Behavior and Human Performance*, *21*(3), 379–391.
- Shanteau, J., Weiss, D. J., Thomas, R. P., & Pounds, J. C. (2002). Performance-based assessment of expertise: How to decide if someone is an expert or not. *European Journal of Operational Research*, *136*(2), 253–263.
- Shephard, G. G., & Kirkwood, C. W. (1994). Managing the judgmental probability elicitation process: a case study of analyst/manager interaction. *IEEE Transactions on Engineering Management*, *41*(4), 414–425.
- Siu, N., Xing, J., & Taylor, G. (2015). Eliciting expert judgment—peer review observations from a recent exercise and future plans. www.nrc.gov/docs/ML1502/ML15028A183.pdf.
- Soll, J. B., & Klayman, J. (2004). Overconfidence in interval estimates. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 299.
- Soll, J. B., & Larrick, R. P. (2009). Strategies for revising judgment: How (and how well) people use others' opinions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(3), 780.
- Spetzler, C. S., & Stael von Holstein, C. A. S. (1975). Exceptional paper-Probability encoding in decision analysis. *Management Science*, *22*(3), 340–358.
- Swann, G., & Lambert, R. (2017). Standards and innovation: A brief survey of empirical evidence and transmission mechanisms. In R. Hawkins, K. Blind, R. P. (Eds.) *Handbook of innovation and standards*, Edward Elgar Publishing.
- Taleb, N. N. (2007). *The black swan: The impact of the highly improbable*, (Vol. 2). Random House.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131.
- Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgement. *Psychological Review*, *90*(4), 293–315.
- Tversky, A., & Koehler, D. (1994). Support theory: A non-extensional representation of subjective probability. *Psychological Review*, *101*(4), 547–567.
- United States Nuclear Regulatory Commission (1975) Reactor Safety Study, An Assessment of Accident Risks in U.S. Commercial Nuclear Power Plants, WASH-1400. NUREG-75014 [commonly referred to as the Rasmussen Report].
- United States Nuclear Regulatory Commission and others. (1990). Severe accident risks: An assessment for five US nuclear power plants. NUREG-1150.
- US Nuclear Regulatory Commission. (1997). (NUREG/CR-6372) US Nuclear Regulatory Commission, Guidance on Uncertainty and Use of Experts. www.nrc.gov/reading-rm/doc-collections/nuregs/contract/cr6372/.
- Walls, L., & Quigley, J. (2001). Building prior distributions to support Bayesian reliability growth modelling using expert judgement. *Reliability Engineering & System Safety*, *74*(2), 117–128.
- Wallsten, T. S., Budescu, D. V., Rapoport, A., Zwick, R., & Forsyth, B. (1986). Measuring the vague meanings of probability terms. *Journal of Experimental Psychology: General*, *115*(4), 348.

- Wilson, K. J. (2017). An investigation of dependence in expert judgement studies with multiple experts. *International Journal of Forecasting*, 33(1), 325–336.
- Winkler, R. L. (1967). The assessment of prior distributions in Bayesian analysis. *Journal of the American Statistical Association*, 62(319), 776–800.
- Wisse, B., Bedford, T., & Quigley, J. (2008). Expert judgement combination using moment methods. *Reliability Engineering & System Safety*, 93(5), 675–686.