Chapter 1
Introduction and Overview of Structured Expert Judgement

Simon French, Anca M. Hanea, Tim Bedford, and Gabriela F. Nane

1.1 Background

Although we live in a data-rich age, it is not true that we have or ever will have sufficient data to evaluate all potential future events, risks or opportunities. Some will have novel or unexplored characteristics: for instance, the medium- and long-term socio-economic effects of Brexit, the effects of plastic waste in the aqueous environment or a future pandemic. In such cases, there are simply too few data on which solely to base useful quantitative assessments of risk and we need to look to experts for guidance. Of course, any risk or decision analysis relies on expert judgement to some extent. Data necessarily refer to events or entities in the past or immediate present, so a judgement is needed that they are relevant to the prediction of any risk or opportunity in the future. Expert judgements are also needed to select appropriate models and analytic methods, to interpret the output of an analysis and to assess whether it provides sufficient guidance to implement risk management strategies or make a decision. Although such topics will be touched on in the following chapters, they are not the prime focus of this collection of readings. Rather our concern is with the use of expert judgement to provide quantitative probabilistic assessments of key uncertainties in an analysis when empirical data are unavailable, incomplete, uninformative or conflicting.

S. French (✉)
University of Warwick, Coventry, England
e-mail: simon.french.50@gmail.com

A. M. Hanea
University of Melbourne, Melbourne, Australia

T. Bedford
University of Strathclyde, Glasgow, Scotland

G. F. Nane
Delft University of Technology, Delft, Netherlands

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_1

Theoretical studies of how to coherently and systematically use expert judgement go back at least to the 1960s and arguably a century or two before that (French 1985; Cooke 1991). Practical studies really began in the second half of the 1980s: see Chap. 9. Despite such a long history and the fact that expert judgement is routinely required to inform critically important decisions across many domains, too often it is obtained by canvasing the judgement of one expert alone, or by using ill-informed and inappropriate elicitation and aggregation methods. Instead, best practice requires the use of panels of experts, structured elicitation protocols and aggregation methods that recognise the complexity of human judgements.

Some organisations, e.g. the European Food Safety Authority (EFSA), have established full protocols for drawing expert judgement into their analyses and working practices in a structured and explicit manner (EFSA 2014). More organisations need to follow EFSA's lead, but currently there is undoubted growth in the number of risk and decision studies that use such methods.

We shall emphasise the importance of expert judgement studies being *structured* and *explicit*. It is easy to ask anyone, expert or not, for their judgement and, by and large, they will give it. But it is not so easy to do this in a way that encourages a thoughtful, auditable and relevant answer that is not affected or biased in some way by the giver's psychology. Moreover, when more experts are asked, seldom will they give the same answer. So how should we combine them? Should we give them 'equal weight' in some sense or perhaps give greater importance to those with either acknowledged or assessed expertise? Should we allow interaction and provide feedback? Whatever we do, those words 'structured' and 'explicit' tell us that we should do so in a careful, auditable, fully reported way. As scientists, we are well versed in how to report data and analyses from empirical studies so that they are clear, open to peer review and allow repetition in validation studies. How do we report the process of gathering the relevant judgements of experts so that they can form some or all the evidence in an analysis?

Those, in brief, are the topics that we shall be surveying in this overview chapter. The readings that follow will flesh out many of the topics through theoretical and methodological discussions and case studies. In the next section, we set the context a little further, by categorising different contexts in which structured expert judgement (SEJ) may be used. Section 1.3 discusses how judgements should be elicited from experts. As we have suggested, simply asking them risks answers biased by potentially flawed thinking. The judgements of several experts can be aggregated in several ways. They can be elicited individually and then combined by some mathematical process or they can be elicited consensually from the group through qualitative discussion. These are the topics of Sects. 1.4 and 1.5, respectively. In Sect. 1.6, we consider how SEJ studies should be reported. Although we believe that SEJ is now a mature technology that can be—and indeed has been—applied in many complex risk and decision analyses, there are still many areas requiring further research and development. We indicate some of these in Sect. 1.7. Finally, in Sect. 1.8, we give an overview of the following chapters.

1.2 Contexts

Experts may be consulted for their advice on risks and uncertainties in a number of contexts. French (1985) introduced three broad categories: the *expert problem*, the *group decision problem* and *textbook problem*, though often, individual problems reflect aspects of two or all of these.

- In the *expert problem*, a group of experts are asked for advice from a problem-owner or decision-maker, external to the group. Responsibility and accountability for the potential consequences rest with that person and the experts are free from those, and the many pressures that might bias their judgements. In this context, the emphasis is on the decision-maker learning from the experts.
- In the *group decision problem*, the group itself is jointly responsible and accountable for the decision. They are their own experts and are both experts and decision-makers. They have responsibility and accountability for the decision. The group may, and indeed probably will, wish that their actions appear rational, fair and democratic. Thus, they may wish to combine their judgements in some formal structured way; but in voting, each will surely wish to be guided by their own personal views having heard the opinions of the others.
- In the *textbook problem,* a group of experts may simply be consulted for their judgements for others to use in as yet undefined circumstances. Here, there is no predefined risk or decision problem, but many potential ones that are as yet only vaguely perceived. An example here is the reports of the Intergovernmental Panel on Climate Change.

The careful distinctions between the roles of expert and decision-maker implicit in the above descriptions are important. Whatever the context, the experts are asked for their opinion on the *science* of the situation: either how likely something is to happen or their subjective estimate of an unknown quantity. They are not asked for *value* or *preference* judgements. This reflects recognised practice in the relationship between science advisors and decision-makers, the scientific culture of being evidence-led, and also a technical perspective from the theory of rational decision-making in which uncertainties and value judgements are separate components of decision-making under uncertainty.

The expert and group decision problems have at their heart a specific risk or decision problem; the third does not. This specific focus provides a structure against which possible approaches can be judged—which does not imply that their resolution needs to be similar. What may be appropriate to one problem may be less suited to the other. For instance, in the expert problem, it seems entirely appropriate for the decision-maker to process the experts' judgements if she believes that they may be biased in some way, i.e., poorly calibrated. In the case of the group decision problem, equity arguments would suggest that the group members should be able to vote according to their best beliefs even if others believe them to be poorly calibrated. Similarly, arguments drawing on democratic principles may suggest that all experts should be treated equally in the group decision problem, whereas in the expert problem, it may not be reasonable to assume that all experts are equally knowledgeable and, hence, the decision-maker might weigh them differently.

Both the expert and the group decision problems have been well explored with many theoretical and methodological contributions over the years (see, e.g. Clemen and Winkler 1999; Cooke and Goossens 2000; O'Hagan et al. 2006; Hora 2008; Burgman 2015; Dias et al. 2018). However, little has been written on the textbook problem, perhaps because its lack of structure makes it more difficult to address. Developments in web-based public participation, stakeholder engagement and deliberative democracy, however, are giving the topic some importance (French 2012), as different groups seek to draw on previous SEJ studies to provide evidence in a different context.

Cooke (1991) argues from a methodological point of view that in all the above contexts, the goal of an SEJ elicitation is to enhance rational consensus which is attainable if problem owners commit in advance to the way the experts' views are elicited and aggregated. He formulates necessary conditions for rational consensus in the form of four principles: *Scrutability/Accountability*, "all data, including experts' names and assessments, and all processing tools should be open to peer review and results must be reproducible by competent reviewers"; *Empirical Control*, "quantitative expert assessments should be subject to empirical quality controls"; *Neutrality*, "the method for combining and evaluating expert opinion should encourage experts to state their true opinions, and must not bias results"; *Fairness*, "experts should not be prejudged, prior to processing the results of their assessments".

These principles are operationalised in the Classical Model for SEJ which will be discussed in more detail in Sect. 1.4.3 and in Chap. 10 of this book. Even though sometimes criticised (e.g. French 2011), no alternative principles have been formulated; hence, they remain important guidelines for SEJ protocols. The *Scrutability/Accountability* principle is essential in how SEJ studies should be reported, and it will be further discussed in Sect. 1.6.

Throughout we assume that the experts are asked to quantify their uncertainty using probabilities, or numerical estimates corresponding to quantiles of probability distributions. We recognise that others have proposed different formalisms for encoding uncertainty numerically, but no other methodology has the power, axiomatic and empirical validity of probability theory. We do not assume, however, that experts are only asked for their numerical estimates. A good elicitation also gathers the experts' reasoning behind their statements and reports this too. Such qualitative material is as important to sound risk and decision analyses as their quantitative judgements, providing, for instance, a qualitative commentary on the validity of the models used.

Finally, while we have spoken of risks and decisions, we emphasise that SEJ methods are also important in Bayesian statistical inference in developing informative prior distributions for analyses (French and Rios Insua 2000).

1.3 Elicitation

To use SEJ, it is inevitable that the analyst asks the experts for their probabilities. That is a far more skilled task than at first might seem. The problem is that when anyone—experts included—is asked such a question, they may respond based on very superficial thinking. Behavioural scientists and psychologists have investigated how people frame and respond to questions relating to uncertainty, since Ward Edwards (1954) asked whether real people were as rational in their behaviour as economic and Bayesian theories of expected utility would suggest. Evidence quickly accumulated that in general they were not. Their behaviour was governed by many heuristic patterns of thought that could lead to judgements and actions that were systematically biased away from the assumptions underlying theories of rationality. Empirical behavioural studies identified many 'heuristics and biases', and this area of research became known under that title (Kahneman and Tversky 1974). Nowadays, one distinguishes System 1 Thinking and System 2 Thinking. The former refers to simple, fast, heuristic patterns of thought on the borders of consciousness; the latter to more conscious, slower, explicit and auditable analysis, one that can be tested against and corrected to be consistent with some norms of rationality (Kahneman 2011). A caveat: our description suggests a dichotomy between these two systems of thinking, but they may represent two ends of a scale with many forms of thinking between the two, moving from the subconscious to the conscious. Indeed, there is much debate within the psychological and behavioural sciences about the precise details of such systems of thinking (Evans and Stanovich 2013). The potential for subconscious patterns of thought to lead to irrationalities is unquestioned, however.

The heuristics of System 1 Thinking may lead to biased responses from the experts during elicitation of their probabilities. For instance, the *availability* heuristic leads individuals to overestimate the probability of events that are easily recalled because of horrific consequences. The *anchoring* heuristic suggests that if a question contains a potentially relevant number, the responder will anchor on that, giving a numerical reply biased towards it. Overconfidence is considered "the most significant of the cognitive biases" (Kahneman 2011) and can play a paramount role in expert's ability of quantifying uncertainty. We do not survey and summarise the many heuristics and biases that need to be taken into account during elicitation: there is a large literature doing precisely that (e.g. Kahneman and Tversky 2000; Gigerenzer 2002; Bazerman 2006; Kahneman 2011). Nor do we provide guidance on the forms of questioning that nudge experts into more System 2 forms of thinking and thus reduce the potential biases in their stated probabilities (see, e.g. for surveys: Wright and Ayton 1987; O'Hagan et al. 2006; Hora 2007). What we would emphasise is that elicitation is a skill that needs to be acquired from training and guided by mentoring; it is not easily developed simply by reading a textbook.

We would also note the importance of developing a detailed elicitation protocol *before* embarking on any elicitations from experts to ensure that all are treated in the same way. Elicitation protocols are as important in expert judgement studies as experimental designs are in empirical research.

1.4 Mathematical Aggregation

1.4.1 Introduction

Taking a rather simplistic view, there are two ways of producing a combined judgement from a group of experts. Firstly, we could ask each to give their assessments and then take their numerical arguments and combine them by some mathematical process. Secondly, we could ask them to discuss the uncertainties and provide a consensual numerical group judgement, thus aggregating their individual opinions behaviourally. Here, we discuss the former approach, leaving the latter to the next section. We confine attention in this section to opinion pools, Cooke's Classical Model and Bayesian approaches. Arguably, these span all the mathematical aggregation approaches though proponents of some approaches may prefer different terminologies. For wider reviews, see, i.e., French (1985), Genest and Zidek (1986) and Jacobs (1995).

Note that we do not discuss mathematical approaches to the group decision-making problem here, since that would lead us to review game theory, adversarial risk analysis and social choice literatures all of which have large and continually growing literatures (Osborne 2003; French et al. 2009; Banks et al. 2015; Sen 2017).

1.4.2 Opinion Pools

Opinion pools take a pragmatic approach. Assuming that a group of experts have each provided assessments for a number of different possible outcomes (typically we take outcomes that are exclusive and exhaustive, that is, one, and only one, of them has to occur), they simply average the individual expert's assessments. Intuitively if not conceptually, these approaches take the experts' judgements as probabilities in their own right. The process may use a weighted arithmetic or weighted geometric mean or perhaps something rather more general:

$$P_{DM} = \sum_{e=1}^{\mathrm{E}} w_e P_e \quad \text{or} \quad P_{DM} = \prod_{e=1}^{\mathrm{E}} P_e^{w_e} \quad \text{or} \quad P_{DM} = \phi(P_1, P_2, \ldots, P_E)$$

The P_e are the experts' probabilities indexed over the E experts, the numbers w_e are weights adding to 1 and the ϕ function denotes a general mathematical formula. The combined probability is subscripted DM for decision-maker. In their 'vanilla' form, the weights in an opinion pool are often simply given by the decision-maker or analyst based on their judgements of the experts' relative expertise, seldom with any operational meaning being offered for the concept of 'relative expertise'. Alternatively, the weights may be taken as equal, perhaps on some Laplacian Principle of Indifference, or of equity, or, even, on the basis that all the experts are paid the same.

A suggestion that the weights might be defined as some measure of the reputation and influence of the experts on social networks has been investigated but not found useful (Cooke et al. 2008).

There have been many attempts to investigate axiomatic justifications of opinion pools, requiring such properties as follows:

- *Marginalisation.* Suppose that the experts are asked for a joint distribution over (X, Y) say. Then the same result should be obtained for the marginal distribution over X whether the experts' distributions are marginalised before forming the combination or the combination formed and the result marginalised (McConway 1981).
- *Independence Preservation.* If all experts agree that (X, Y) are independent variables, the variables should remain independent in the combined distribution (French 1987; Genest and Wagner 1987).
- *External Bayesianity.* If relevant data become available, then the same result should be obtained whether the experts update their individual distribution through Bayes Theorem before they are combined or their distributions combined and the result then updated (Madansky 1964; Faria and Smith 1997).

If, for instance, one insists that any opinion pool should satisfy marginalisation, then one is limited essentially to weighted arithmetical, i.e., linear, pools (McConway 1981). But if one adds in other requirements such as the other two above or further ones, then impossibility results quickly accumulate showing that no pool can simultaneously satisfy all the requirements (French 1985).

It can be argued (see, e.g. Bedford and Cooke 2001) that independence preservation is less important than marginalisation, and hence that it makes most sense to adopt the linear opinion pool.

The equally weighted linear opinion pool is extensively used in applications though the justification for doing so is seldom discussed in any detail. Possible arguments are appeals to fairness between experts and the lack of any reason to deviate from equal weights. However, we have argued above that fairness/equity arguments should not be applied to expert problems, and Cooke's approach described below provides reasons to deviate from equal weights.

1.4.3 Cooke's Classical Model

Cooke developed the *Classical* Model to combining expert judgement (Cooke 1991; see also Part II of this book). In this, the weights are defined empirically on the basis of the experts' relative performance on a calibration set of variables. This is in accordance with the *Empirical Control* principle for rational consensus. This principle is the one that justifies the collection of calibration data so that the quality of each expert's input can be assessed and their judgements weighted accordingly. Here, the experts do not know the true values of the calibration quantities, but the analyst does. Comparing the experts' answers with the true values over the calibration set allows Cooke to form measures of the calibration and informativeness of each expert, and from these he constructs *performance-based* weights. Cooke justifies his weights on the basis of a particular group scoring rule, which follows the *Neutrality* principle.

Cooke's Classical Model has been used in over 150 published studies. Several case studies are reported later in this book: see also Dias et al. (2018) and the special issue of *Reliability Engineering and System Safety* (2008, **93**(5)). There is also a growing database with the data from these SEJ studies available online at http://rogermcooke.net/.

1.4.4 Bayesian Approaches

Bayesian approaches differ from opinion pools in that they treat the numerical judgements as data and seek to update a prior distribution supplied by or constructed for the decision-maker using Bayesian methods. This requires that the analyst develops appropriate likelihood functions to represent the information implicit in the experts' statements. Specifically, the likelihood function needs to model:

- the experts' ability to encode their uncertainty probabilistically (Clemen and Lichtendahl 2002; O'Hagan et al. 2006; Hora 2007; Lin and Bier 2008);
- correlations that arise because of experts' shared knowledge and common professional backgrounds (Shanteau 1995; Mumpower and Stewart 1996; Wilson 2016);
- correlations between the decision-maker's own judgements and the experts' (French 1980);
- the effects of other biases arising from conflicts of interests and the general context of the elicitation (Hockey et al. 2000; Skjong and Wentworth 2001; Lichtendahl and Winkler 2007; French et al. 2009; Kahneman 2011).

Constructing such a likelihood is a far from easy task. Indeed, developing suitable likelihood functions proved an insurmountable hurdle for many years and only recently have tractable methods with reasonable likelihood functions become available (Albert et al. 2012; French and Hartley 2018; see also Chap. 5).

This does not mean that Bayesian ideas have been without influence. They have provided an analytical tool for investigating the principles of combining and using expert judgement. For instance, French (1987) used a simple Bayesian model to argue against the principle of Independence Preservation and French and Hartley (2018) used a Bayesian approach to critique the European Food Safety Authority's SEJ methodology (EFSA 2014).

1.5 Behavioural Aggregation

Behavioural approaches work with the group of experts to agree on probabilities or quantiles that they can all accept as reasonable input to the risk or decision analysis, even if they themselves would still hold to different values. Simplistically, the analyst might gather the experts together and let them come to some agreement about the numbers to put into the models (DeWispelare et al. 1995). However, it is better to use a more structured approach. The Sheffield method is a version of facilitated workshop or decision conferencing (Reagan-Cirincione 1994; Phillips 2007) in which the group is helped to agree on probabilities that an 'impartial observer' of their discussions might give (O'Hagan et al. 2006; EFSA 2014). The long established Delphi method, or rather family of methods, does not allow the experts to meet but structures a discussion in which their share and revise their opinions through several iterations arriving at an agreed set of values (Dalkey and Helmer 1963; Rowe and Wright 1999; EFSA 2014).

Many factors need to be balanced in choosing between mathematical and behavioural aggregation (Clemen and Winkler 1999). Behavioural aggregation may be affected by many group dysfunctional behaviours, though these may be countered by good facilitation. It can help share knowledge and ensure that all experts share the same precise understanding of the quantities to be elicited. However, there is a risk that behavioural aggregation can win people over, losing concerns about some issues held by a small minority in building a group consensual judgement. Thus in risk management contexts, behavioural aggregation may lose sight of some potential hazards. In planning or regulatory decisions, the explicit, auditable nature of mathematical aggregation can be an advantage in recording all the reasoning implicit in the analysis.

However, there is no reason for the two approaches to be kept separate. Hanea et al. (2018) have developed the IDEA protocol which can combine group discussion with the use of the Classical Model to form the final judgements from the individual assessments.

1.6 Reporting

SEJ studies are necessarily important inputs to a risk or decision analysis. Since such studies are expensive, they are only undertaken when the events or quantities of concern are significant in driving the output uncertainties of the analysis. So it behooves those conducting the studies to report their conduct and conclusions fully. It is somewhat surprising therefore that there is remarkably little guidance on how this should be undertaken. In contrast, the research community and scientific journals have developed and enforced a wide range of principles to govern the peer review, publication and use of empirical studies, alongside which has grown a recognition of the importance of evidence-based decision-making (Pfeffer and Sutton 2006;

Shemilt et al. 2010). The latter developments began within medicine, particularly in the Cochrane Collaboration; but the imperatives of basing decisions on evidence are now changing thinking in many domains.

As mentioned previously, Cooke suggested the *Scrutability/Accountability* principle according to which all data, including experts' names and assessments, and all processing tools should be open to peer review and results must be reproducible. This, unfortunately, is sometimes unachievable as many experts are uncomfortable about having their assessments published under their names. They prefer and expect publication under Chatham House Rules, namely, their participation in the study will be noted, but their judgements and other input will be reported anonymously.

French (2012) considered issues relating to the reporting of SEJ studies from the perspective of potential future meta-analyses that might seek to draw information from two or more. One point here is that in scientific reporting of empirical studies, one should always report the experimental design process underpinning the data collection. In SEJ the elicitation protocol serves the same purpose, though it is not reported in detail in many studies.

EFSA (2014) guidance on running SEJ studies is perhaps the most thorough to date and that does provide a lot of advice on the content of report and the responsibilities of different teams for writing sections of reports. However, the community of analysts involved in SEJ still mainly relies on their experience and personal perspectives in deciding what to include and with what details in their reporting.

1.7 Directions for Future Developments and Research

SEJ methodologies are maturing. The past decade or so has seen a steady growth in applications across many domains. Several are reported in this and an earlier sister book (Dias et al. 2018); and any literature review will easily find many more. We have a toolbox of methods and tools to call on when an SEJ study is needed. But that is not to deny the need for further research. There are several areas in which more developments, both theoretical and methodological, are needed.

Firstly, our current SEJ methods focus mainly on eliciting and aggregating expert assessments of the probability of unknown events or univariate probability distributions of unknown quantities. However, once multivariate probability distributions are needed, things become more difficult and there is still a need for research. Multivariate distributions, whether described by a full multivariate distribution function or represented by a belief network, hierarchical model or some such, require dependences, independences, correlations, etc., if they are to be defined fully. How these should be elicited and aggregated remains a research topic. Quigley et al. (2013) and Werner et al. (2017) are very relevant references; as are Chaps. 2, 4 and 7 in this volume. However, much remains to be done.

We mentioned above that Bayesian methods had not been much used in practice, though they had provided many theoretical insights into SEJ. Things are changing. Chapter 5 in this volume suggests that a new Bayesian method of grouping experts automatically using a calibration set may be achieving a similar performance to the Classical Model. While there may be no great improvement over the simpler Classical approach, the Bayesian methodology will combine seamlessly with other Bayesian models for data analysis, machine learning, risk and decision analysis in an overall analysis. This suggests that further work to improve the Bayesian modelling of the correlations between experts and between them and the decision-maker may bring overall benefits.

Experts do not have unlimited time and effort to give to the elicitation of uncertainties. Apart from the value of their time in other aspects of the analysis and elsewhere, reflected in high consultancy rates, they tire because reflective elicitation requires considerable thought and effort. Currently, much simple sensitivity analysis is used in the early stages of the overall risk and decision analysis to prioritise the uncertainties which should be elicited from experts (French 2003). However, those processes are rudimentary and conducted before the SEJ elicitations. One aspect of elicitation that is not emphasised as much as it should be is that much *qualitative* knowledge is elicited from the experts alongside their quantitative judgements. This qualitative information can be used to shape the overall modelling further, and hence may change the priorities that a sensitivity analysis would determine. Indeed, their quantitative judgements on some uncertainties may constrain possible values of others in complex risk and decision models. Thus there may be benefit in developing procedures and tools to integrate the processes of SEJ with the overall risk and decision analysis process yet more effectively.

Taking this last point further, the overall risk and decision models need building in the first place. There are a host of—arguably under-researched—methods and tools to catalyse this process, which go under various names in different disciplines, e.g. problem structuring methods, soft-OR and knowledge engineering. We might step back and look at the whole process of formulating models, identifying parameters for which there is insufficient data to quantify their values and the uncertainties, and then focusing on those as targets for SEJ. Qualitative information should flow back and forth along this as the model is shaped. We might term this process as iterating between soft and hard elicitation, i.e., between identifying model structure and parameter values. Note that in simpler models a parameter may approximate the average effect of a sub-model in a more complex model. Thus, the distinction between parameters and model structure is to some extent arbitrary and depends on the modeller's perspective. This suggests looking at the modelling process as a whole, from the point of view of eliciting knowledge from experts. If we do that we should pause and think: why do we use a variety of techniques in hard quantitative elicitation to counter possible biases arising from System 1 Thinking, yet use virtually none in eliciting knowledge to structure models? There is a need for research to look at the whole modelling and elicitation process in relation to the potential effects of System 1 Thinking and to develop mechanisms and interventions that encourage comprehensive System 2 Thinking.

Finally, we now have around 30 years of experience in SEJ applications. That is plenty of time for many of the uncertainties relating to the risks and decisions to be resolved. It would be interesting in historical research to relate the actual outcomes to the probabilities derived from SEJ used in past analyses. In short, we could step back and seek to calibrate and validate SEJ overall. There would be many difficulties, of course. The world is far more complex than the microcosmos explored in risk and decision analyses, and the ultimate outcomes may be derived more from unanticipated events and behaviours outside the models. Nonetheless, there are potentially data that might be used to provide validation for SEJ as a whole, and that would do much to reassure the decision-makers and risk owners of today of the value of our methods.

1.8 Outline of the Book

This book grew from a conference on *The State of the Art in the Use of Expert Judgement in Risk and Decision Analyses* held in Delft in July 2017. The conference marked the end of a European Co-operation in Science and Technology (COST)[1] Action, which ran from 2013 to 2017. Its main aim had been to create a multidisciplinary network of scientists and policymakers using SEJ to quantify uncertainty for evidence-based decisions, and hence improve effectiveness in the use of science knowledge by policymakers. An earlier sister volume to this had been written at the outset of the Action and had recently been published (Dias et al. 2018).

The conference also had a second purpose: to honour and celebrate the work of Roger Cooke in establishing sound SEJ processes over some four decades. Many of us had worked with him over that period and, indeed, still work with him. Thus, this volume is also a festschrift for him and a recognition of his leadership.

We have divided the chapters into four parts:

Part I. Current Research.
Part II. Cooke and the Classical Model.
Part III. Process, Procedures and Education.
Part IV. Applications.

1.8.1 Part I: Current Research

Part I gathers recent theoretical developments in the field of SEJ. Chapter 2 focuses on expert elicitation of parameters of multinomial models. It presents an extensive overview of the recent research on the topic, along with guidelines for carrying out an elicitation and supporting examples. Chapter 3 discusses whether using performance weights is beneficial and advances the random expert hypothesis. Chapter 4 considers expert elicitation for specific graphical models and discusses the importance of choosing an appropriate graphical structure. Chapter 5 considers how to model dependencies between experts' assessments within a Bayesian framework

[1] See https://www.cost.eu/actions/IS1304/ and https://www.expertsinuncertainty.net/.

and provides a performance comparison with the Classical Model. Still within a Bayesian context, yet in a preventive maintenance setting, Chap. 6 focuses on eliciting experts' lifetime distributions in order to obtain prior parameters of a Dirichlet process. Finally, Chap. 7 proposes an adversarial risk analysis approach for SEJ studies in which the main uncertainties relate to the actions of other actors, usually though not necessarily adversaries.

1.8.2 Part II: Cooke and the Classical Model

Roger Cooke's oration in 1995 on taking up his chair at the Technical University of Delft has never been formally published. It was given barely 10 years after he developed the Classical Model and showed its potential value in early applications; we are proud to publish the oration here as Chap. 8. In Chap. 9, one of us reflects back to that early decade, showing how all the basic principles of the Classical Model and the processes surrounding were laid down then. Chapter 10 provides a current overview of the theory of the Classical Model, providing a deep and comprehensive perspective on its foundations and its application. In Chap. 11, we present an interview with Roger Cooke in which he reflects on the Classical Model and the processes of SEJ.

1.8.3 Part III: Process, Procedures and Education

Part III focuses on processes and procedures for SEJ and on how experiences should be turned into lessons and guidelines that continue to shape what structured elicitation protocols are in the digital age. In Chap. 12, we report an interview with Professor Dame Anne Glover, who served as the Chief Scientific Advisor to the President of the European Commission. She reflects on the role of expert scientific advice to governments. Chapter 13 synthesises the characteristics of good elicitations by reviewing those advocated and applied. It examines the need of standardisation in mature protocols. Chapter 14 discusses the design and development of a training course for SEJ, based on two major experiences in training postgraduates, early career researchers and consultants. Chapters 15 and 16 detail specific experiences with SEJ protocols with the intention of presenting the challenges and insights collected during this journey, and the way those re-shaped what an optimal protocol may look like.

1.8.4 Part IV: Applications

There have been many applications of SEJ over the years. The database maintained by Roger Cooke reports over 100 using the Classical Model alone. The Sheffield Method which is the leading behavioural aggregation approach has had many applications though there is not a specific database of these. In the COST Action, we discussed and promoted applications in many areas including natural seismic and volcanic risks, geographical and spatio-temporal uncertainties relating to pollution and disease, uncertainties in managing public health systems, food safety and food security, project and asset management risks, and the uncertainties that arise in innovation and development. Part IV begins with some reflections from Willy Aspinall on his many experiences in applying the Classical Model in several application domains (Chap. 17). Chapter 18 also provides some related reflections on imperfect elicitation. We present several discussions and applications relating to medicines policy and management (Chap. 19), supply chain cyber risk management (Chap. 20), geopolitical risks (Chap. 21), terrorism (Chap. 22) and the risks facing businesses looking to internationalise (Chap. 23).

References

Albert, I., Donnet, S., Guihenneuc-Jouyaux, C., Low-Choy, S., Mengersen, K., & Rousseau, J., et al. (2012). Combining expert opinions in prior elicitation. 503–532.

Banks, D. L., Aliaga, J. M. R., & Insua D. R. (2015). *Adversarial risk analysis*. CRC Press.

Bazerman, M. H. (2006). *Managerial decision making*. New York: John Wiley and Sons.

Bedford, T., & Cooke, R. M. (2001). *Probabilistic risk analysis: Foundations and methods*. Cambridge: Cambridge University Press.

Burgman, M. A. (2015). *Trusting judgements: How to get the best out of experts*. Cambridge University Press.

Clemen, R. T., & Lichtendahl, K. C. (2002). *Debiasing expert overconfidence: A Bayesian calibration model*. PSAM6, San Juan, Puerto Rico.

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

Cooke, R. M. (1991). *Experts in uncertainty*. Oxford: Oxford University Press.

Cooke, R. M., ElSaadany, S., & Xinzheng, H. (2008). On the performance of social network and likelihood based expert weighting schemes. *Reliability Engineering and System Safety, 93*(5), 745–756.

Cooke, R. M., & Goossens, L. H. J. (2000). Procedures guide for structured expert judgement in accident consequence modelling. *Radiation Protection Dosimetry, 90*(3), 303–309.

Dalkey, N., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management Science, 9*(3), 458–467.

DeWispelare, A. R., Herren, L. T., & Clemen, R. T. (1995). The use of probability elicitation in the high-level nuclear waste regulation program. *International Journal of Forecasting, 11,* 5–24.

Dias, L., Morton, A., & Quigley, J., (Eds.) (2018). *Elicitation of preferences and uncertainty: Processes and procedures*. Springer.

Edwards, W. (1954). The theory of decision making. *Psychological Bulletin, 51,* 380–417.

EFSA. (2014). Guidance on expert knowledge elicitation in food and feed safety risk assessment. *EFSA Journal, 12*(6), 3734–4012.

Evans, J. S. B., Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Pyschological Sciences*, *8*(3), 223–241.

Faria, A., & Smith, J. Q. (1997). Conditionally externally Bayesian pooling operators in chain graphs. *Annals of Statistics, 25*(4), 1740–1761.

French, S. (1980). Updating of belief in the light of someone else's opinion. *Journal of the Royal Statistical Society, A143,* 43–48.

French, S. (1985). Group consensus probability distributions: A critical survey (with discussion). In J. M. Bernardo, M. H. DeGroot, D. V. Lindley & A. F. M. Smith (Eds.), *Bayesian Statistics 2* (pp. 183–201). North-Holland.

French, S. (1987). Conflict of belief: when advisers disagree. In P. G. Bennett (Ed.), *Analysing conflict and its resolution: Some Mathematical Contributions* (pp. 93–111). Oxford: Oxford University Press.

French, S. (2003). Modelling, making inferences and making decisions: The roles of sensitivity analysis. *TOP, 11*(2), 229–252.

French, S. (2011). Aggregating expert judgement. *Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales, 105*(1), 181–206.

French, S. (2012). Expert judgment, meta-analysis, and participatory risk analysis. *Decision Analysis, 9*(2), 119–127.

French, S., & Hartley, D. (2018). Elicitation and calibration: A Bayesian perspective. In L. Dias, A. Morton & J. Quigley (Eds.), *Elicitation: The science and art of structuring judgement.* New York: Springer (in press).

French, S., Maule, A. J., & Papamichail, K. N. (2009). *Decision behaviour, analysis and support*. Cambridge: Cambridge University Press.

French, S., & Rios Insua, D. (2000). *Statistical decision theory*. London: Arnold.

Genest, C., & Wagner, C. G. (1987). Further evidence against independence preservation in expert judgement synthesis. *Aequationes mathematicae, 32,* 74–86.

Genest, C., & Zidek, J. V. (1986). Combining probability distributions: A critique and annotated bibliography. *Statistical Science, 1,* 114–148.

Gigerenzer, G. (2002). *Reckoning with risk: Learning to live with uncertainty*. Harmondsworth: Penguin Books.

Hanea, A., McBride, M., Burgman, M., & Wintle, B. (2018). Classical meets modern in the IDEA protocol for structured expert judgement. *Journal of Risk Research, 21*(4), 417–433.

Hockey, G. R. J., Maule, A. J., Clough, P. J., & Bdzola, L. (2000). Effects of negative mood on risk in everyday decision making. *Cognition and Emotion, 14,* 823–856.

Hora, S. (2007). Eliciting probabilities from experts. In W. Edwards, R. F. Miles, & D. Von Winterfeldt (Eds.), *Advances in decision analysis: From foundations to applications* (pp. 129–153). Cambridge: Cambridge University Press.

Hora, S. C. (2008). Expert judgement. In E. L. Melnick & B. S. Everitt (Eds.), *Encylcopedia of quantitative risk analysis and assessment* (pp. 667–676). Chichester: John Wiley and Sons.

Jacobs, R. A. (1995). Methods for combining experts' probability assessments. *Neural Computing, 7,* 867–888.

Kahneman, D. (2011). *Thinking, Fast and Slow*. London, Penguin: Allen Lane.

Kahneman, D., & Tversky, A. (1974). Judgement under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1131.

Kahneman, D., & Tversky, A. (Eds.). (2000). *Choices, values and frames*. Cambridge: Cambridge University Press.

Lichtendahl, K. C., & Winkler, R. L. (2007). Probability elicitation, scoring rules, and competition among forecasters. *Management Science, 53*(11), 1745–1755.

Lin, S.-W., & Bier, V. M. (2008). A study of expert overconfidence. *Reliability Engineering and System Safety, 93,* 711–721.

Madansky, A. (1964). Externally Bayesian groups. RM-4141-PR, RAND.

McConway, K. (1981). Marginalisation and linear opinion pools. *Journal of the American Statistical Association, 76,* 410–414.

Mumpower, J. L., & Stewart, T. R. (1996). Expert judgement and expert disagreement. *Thinking & Reasoning, 2*(2–3), 191–211.

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, R., Garthwaite, P. H., Jenkinson, D., et al. (2006). *Uncertain judgements: Eliciting experts' probabilities*. Chichester: John Wiley and Sons.

Osborne, M. J. (2003). *An introduction to game theory*. Oxford: Oxford University Press.

Pfeffer, J., & Sutton, P. J. (2006). Evidence-based management. *Harvard Business Review, 84*(1), 62–76.

Phillips, L. D. (2007). Decision conferencing. In W. Edwards, R. F. Miles, & D. von Winterfeldt (Eds.), *Advances in decision analysis: From foundations to applications* (pp. 375–399). Cambridge: Cambridge University Press.

Quigley, J., Wilson, K. J., Walls, L., & Bedford, T. J. R. A. (2013). A Bayes linear Bayes method for estimation of correlated event rates. *33*(12), 2209–2224.

Reagan-Cirincione, P. (1994). Combining group facilitation, decision modelling and information technology to improve the accuracy of group judgement Improving the accuracy of group judgment: a process intervention combining group facilitation, social judgment analysis, and information technology. *Organisational Behaviour and Human Decision Process, 58,* 246–270.

Reliability Engineering and System Safety special Issue: Expert Judgment (2008). Edited by Roger M. Cooke, *93*(5), 655–778.

Rowe, G., & Wright, G. (1999). The Delphi technique as a forecasting tool: Issues and analysis. *International Journal of Forecasting, 15,* 353–375.

Sen, A. (2017). *Collective choice and social welfare, expanded edition.* Harmondsworth: Penguin.

Shanteau, J. (1995). Expert judgment and financial decision making. In B. Green (Ed.), *Risky business: Risk behavior and risk management*. Stockholm: Stockholm University

Shemilt, I., Mugford, M., Vale, L., Marsh, K., & Donaldson, C. (Eds.). (2010). *Evidence-Based decisions and economics: Helath care, social welfare, education and criminal justice*. Oxford University Press: Oxford.

Skjong, R. & Wentworth, B. H. (2001*)*. Expert judgement and risk perception. In *Proceedings of the Eleventh (2001) International Offshore and Polar Engineering Conference.* Stavanger, Norway: The International Society of Offshore and Polar Engineers.

Werner, C., Bedford, T., Cooke, R. M., Hanea, A. M., & Morales-Nápoles, O. (2017). Expert judgement for dependence in probabilistic modelling: A systematic literature review and future research directions. *European Journal of Operational Research, 258*(3), 801–819.

Wilson, K. J. (2016). An investigation of dependence in expert judgement studies with multiple experts. *International Journal of Forecasting*.

Wright, G., & Ayton, P. (1987). Eliciting and modelling expert knowledge. *Decision Support Systems, 3*(1), 13–26.