DEPARTMENT OF MANAGEMENT SCIENCE

University of Strathclyde Business School

Structured Expert Judgement for dependence in probabilistic modelling of uncertainty

Advances along the dependence elicitation process

Christoph Werner

Supervisors:
1st: Professor T. Bedford (University of Strathclyde)
2nd: Professor J. Quigley (University of Strathclyde)
3rd: Professor R. M. Cooke (Resources for the Future/Delft University of Technology)

A thesis presented in fulfilment of the requirements for the degree of:
Doctor of Philosophy (PhD)

Glasgow, Scotland, UK - April 2018

Declarations

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Date   Christoph Werner

Within this thesis several contributions (peer-reviewed journal articles and a book chapter) have been made which are either published or currently under review for publication.
I have conducted the research through literature reviews and developing original research ideas under the guidance of my supervisors (Prof T. Bedford, Prof J. Quigley and Prof R. M. Cooke). In the following, I intend to clearly outline the extent of my own contribution to these published (or to be published) papers for which I have been jointly responsible:

- ***Chapter 3:*** Werner, C., Bedford, T., Cooke, R. M., Hanea, A. M. and 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.

The idea for this paper, which reviews the literature on how expert judgement methods can contribute to dependence models together with the possible choices we can make with regards to the elicited forms, was mine. I started this work with the two latter co-authors during a research visit to Delft University of Technology and continued with all of the co-authors after that.
I have written the initial draft of this paper, as well as all subsequent drafts together with the final version.
Discussions with T. Bedford on the paper led to major changes in the overall structure. Further, he contributed to section 3.3 (that it is in this thesis) by deriving the general framework on the dependence modelling context and how choices here affect choices for the expert judgement process. R. Cooke had a main contribution when reviewing the part on scoring rules and their application in dependence elicitation contexts (briefly discussed in section 3.2 as of this thesis).

- ***Chapter 4:*** Werner, C., Hanea, A. M., Morales-Nápoles, O. (2018). Eliciting multivariate uncertainty from experts: Considerations and approaches along the expert judgement process. In: Dias, L. C., Morton, A., Quigley, J. (eds.) Elicitation: The science and art of structuring judgement, New York: Springer International Series in Operations Research and Management Science, 171-210.

The idea for conducting a literature review, presenting an overview of the different elements of structured expert judgement processes for eliciting dependence, was mine. I then discussed the scope of this publication and its structure with my co-authors.
The initial draft, all subsequent drafts and the final version were written by myself. This included conducting the literature reviews on the different processes' elements. In particular, less explored elements of dependence elicitation processes, such as training as well as knowledge and belief structuring, but also biases and heuristics, required a more throughout discussion for determining which research and literature we consider relevant for dependence elicitation. As such, we decided, for example, to include a brief overview on the literature

about teaching statistical and probabilistic dependence parameters to students. This offers insights and potential learning opportunities for designing experts' training.
The study shown in section 4.5.2 was conducted by my co-author O. Morales-Nápoles, who contributed it to the chapter, together with the results presented in Table 4.3.

- ***Chapter 5:*** Werner, C., Bedford, T. and Quigley, J. (under review). Mapping Conditional Scenarios for Knowledge Structuring in (Tail) Dependence Elicitation, Journal of the Operational Research Society.

  Awarded Winner of the Donald Hicks Scholarship by the UK Operational Research Society

The need for original research in the area of knowledge and belief structuring was first proposed by T. Bedford as part of the preparation of my first annual PhD review. Following that, together with my co-authors, I then developed the method to structure conditional scenarios. During the development of the method, my co-authors mainly contributed by suggesting potentially interesting elements for the method, providing feedback on my ideas and by discussing/refining the scope of the paper. The latter mainly influenced in which areas I did most of my research.
I then proposed, prepared and conducted both elicitations (for eliciting marginals and dependence). The method used for identifying parametric copulas fitting the expert judgements has been proposed by R.M. Cooke.
Lastly, I wrote all drafts of this paper, including the final one.

- ***Chapter 6:*** Werner, C., Bedford, T. and Quigley, J. (2018). Sequential Refined Partitioning for probabilistic dependence assessment, Risk Analysis, doi: 10.1111/risa.13162.

  Awarded Runner-up in the INFORMS Decision Analysis Society Best Student Paper Competition

I identified the need for a sequential elicitation method that allows for eliciting detailed assessments (in particular for tail dependencies) feasibly and intuitively. This identified need has been based on my previous research in dependence modelling and expert judgement, such as chapters 3 and 4 in this thesis.
Following the initial idea for the method, I then continuously further developed the SRP method and provided the first proofs on the feasibility bounds of particular assessment sequences. Based on these first proofs, I then generalised the method, whereas my co-authors made their main contribution when generalising the proofs.
I organised and conducted the case-study elicitations presented in the paper together with two collaborators within the framework of the European Cooperation of Science and Technology (COST).

Finally, I wrote all versions of the paper, including the final one.

- ***Chapter 7:*** Werner, C., Bedford, T., Colson, A. and Morton, A. (to be submitted). Risk assessment of future antibiotic resistance - eliciting and modelling probabilistic dependence between multivariate uncertainties of bug-drug combinations, Working Paper.

Based on the initial proposal of A. Morton and A. Colson to elicit dependence between future antibiotic resistance rates, I wrote all versions of the above chapter. The motivation for the study is to improve the decision-making for research and development of antibiotics through dependence assessments.
Together with A. Colson, I prepared and conducted the dependence elicitation presented therein. The elicitation of marginal future resistance distributions of selected antibiotics was done previously by A. Colson. The other co-authors provided feedback throughout the writing of the chapter and elicitation protocol.

Date ______________________ Christoph Werner ______________________

"The story that I have to tell is marked all the way through by a persistent tension between those who assert that the best decisions are based on quantification and numbers, determined by the patterns of the past, and those who base their decisions on more subjective degrees of belief about the uncertain future. This is a controversy that has never been resolved. [...] Which matters more when facing a risk, the facts as we see them or our subjective belief in what lies hidden in the void of time? Is risk management [own remark: and assessment] a science or an art? Can we even tell for certain precisely where the dividing line between the two approaches lies?"
- Against The Gods (p.6),
Peter Lewyn Bernstein

Acknowledgements

The completion of this thesis would have not been possible without the effort of numerous people. First of all, I am extremely grateful to the dedicated and constant support of my supervisors - not only did I enjoy the regular meetings in which we (most of the times) had interesting and fruitful discussions about the research presented in this thesis, but I am also thankful for teaching me the techniques of probabilistic dependence modelling and assessment together with guiding me during times when I was stuck on how to further develop my ideas.
Further, I would like to thank everyone in the Department of Management Science at the University of Strathclyde as I truly enjoyed the atmosphere there, both from and academic point of view but also on a personal level.
Next, I would like to recognise the support from the European Cooperation in Science and Technology, COST Action IS1304 - Expert Judgement Network, which allowed me to meet numerous fellow academics working in the research area on several occasions, such as conferences, meetings and on a so called Short Term Scientific Mission. The latter has made a decisive impact on this thesis as it allowed me to spend some time with Dr. Anca M. Hanea and Oswaldo Morales-Nápoles with whom the work presented in chapters 3 and 4 has been conducted.
Lastly, I want to thank the constant support of my family, who especially helped to keep my spirits up during difficult times, and all friends, I have made along the way here in Glasgow, who made my time as a PhD student so much more fun outside of working hours.

Abstract

In decision and risk analysis problems, modelling uncertainty probabilistically provides key insights and information for decision makers. A common challenge is that uncertainties are typically not isolated but interlinked which introduces complex (and often unexpected) effects on the model output. Therefore, dependence needs to be taken into account and modelled appropriately if simplifying assumptions, such as independence, are not sensible. Similar to the case of univariate uncertainty, relevant historical data to quantify a (dependence) model are often lacking or too costly to obtain. This may be true even when data on a model's univariate quantities, such as marginal probabilities, are available. Then, specifying dependence between the uncertain variables through expert judgement is the only sensible option. A structured and formal process to the elicitation is essential for ensuring methodological robustness.

This thesis consists of three published works and two papers which are to be published (one under review and one working paper). Two of these works provide comprehensive overviews from different perspectives about the research on dependence elicitation processes. Based on these reviews, novel risk assessment and expert judgement methods are proposed - (1) allowing experts to structure and share their knowledge and beliefs about dependence relationships prior to a quantitative assessment and (2) ensuring experts' (detailed) quantitative assessments are feasible while their elicitation is intuitive. The original research presented in this thesis is applied in case-studies with experts in real risk modelling contexts for the UK Higher Education sector, terrorism risk and future risk of antibacterial multi-drug resistance.

Keywords: DEPENDENCE ELICITATION, DEPENDENCE MODELLING, UNCERTAINTY ANALYSIS, STRUCTURED EXPERT JUDGEMENT, DECISION ANALYSIS, RISK ANALYSIS

Contents

Contents **xiv**

List of Figures **xvii**

List of Tables **xix**

I Introduction and Preliminaries **1**

1 Introduction **2**
1.1 Dependence between multivariate uncertainties 2
1.2 Structured Expert Judgement 3
1.2.1 Who is an expert? 3
1.2.2 What is elicitation? 4
1.2.3 Dependence in the subjective probability context 4
1.3 Research objectives, questions and scope 5
1.4 Epistemological considerations and research methodology 6
1.5 Contribution of thesis 10
1.6 Thesis outline 10

2 Preliminaries **13**
2.1 Chapter introduction 13
2.2 Uncertainty 14
2.2.1 Different forms of uncertainties 14
2.2.2 Measuring uncertainty 17
2.2.3 Interpreting uncertainty 19
2.2.4 Operationalising subjective probabilities 24
2.3 Risk 27
2.3.1 Qualitative aspects and definitions 27
2.3.2 A quantitative definition 28
2.3.3 Some alternative risk definitions and misconceptions . . . 29

II Literature Overviews **32**

3 SEJ for common probabilistic dependence models **33**
3.1 Chapter introduction 33
3.2 Generalisations of univariate elicitation processes for eliciting dependence 34
3.3 Guide to modelling and elicitation context 36
3.4 Dependence models and expert judgement 39
3.4.1 Elicitation for direct modelling 39

3.4.2 Elicitation for indirect modelling with auxiliary variables 47
3.4.3 Elicitation for modelling propagation of output 48
3.5 Forms of elicited dependence parameters 48
3.5.1 Probabilistic methods 49
3.5.2 Statistical methods 55
3.5.3 Other methods 58
3.6 Dependence elicitation in the empirical literature 59
3.7 Chapter conclusions and further research 63

4 Considerations and approaches along the SEJ process when eliciting dependence 66
4.1 Chapter introduction 66
4.2 Structured Expert Judgement processes: an overview 67
4.3 Biases and Heuristics for Dependence Elicitation 70
4.3.1 Causal reasoning and inference 72
4.3.2 Biased Dependence Elicitation: an Overview 73
4.3.3 Implications of biases for the elicitation process 79
4.4 Elicitation process: Preparation/Pre-elicitation 80
4.4.1 Problem identification and modelling context 80
4.4.2 Choice of elicited parameters 81
4.4.3 Specification of marginal distributions 85
4.4.4 Training and motivation 85
4.5 Elicitation process: Elicitation 87
4.5.1 Knowledge and belief structuring 87
4.5.2 Quantitative elicitation 88
4.6 Elicitation process: Post-Elicitation 89
4.6.1 Aggregation of expert judgements 89
4.6.2 Feedback and robustness analysis 92
4.7 Chapter conclusions 94

III Original Research 96

5 Mapping conditional scenarios for knowledge structuring in (tail) dependence elicitation 97
5.1 Chapter introduction 97
5.2 Tail dependence models and the resulting variables of interest for the elicitation 98
5.3 Tail dependence elicitation 99
5.4 Scenario definition and desiderata of structuring methods 101
5.5 Applicability and features of existing structuring methods 102
5.6 Mapping conditional scenarios 105
5.6.1 Overview of the mapping process 105
5.7 An illustrative example: Managing risk in the UK Higher Education sector 108
5.8 Discussion and chapter conclusions 113

6 Sequential Refined Partitioning for probabilistic dependence assessment 115
6.1 Chapter introduction 115
6.2 Eliciting detailed dependence information feasibly and consistently through *sequential refined partitioning* 117

6.2.1 General set-up of sequentially refined partitioning 118
6.2.2 Commonly assessed quantile partition sequences 121
6.3 Modelling the form of directly and indirectly assessed probability masses through minimum information 125
6.4 An illustrative case-study: assessing spatial dependence of political violence/terrorism risk in insurance underwriting 126
6.4.1 Pricing terrorism risk in insurance 127
6.4.2 Expert judgement for adversarial risk 127
6.4.3 Expert judgement for spatial dependence of terrorism attacks 128
6.4.4 Eliciting the marginal probabilities 129
6.4.5 Applying the SRP method for quantifying spatial dependence of terrorism risk 131
6.5 Discussion and chapter conclusions 136

7 Eliciting and modelling probabilistic dependence between multivariate uncertainties of bug-drug combinations 139
7.1 Chapter introduction 139
7.2 Dependence between uncertainties of resistance rates through multidrug resistance 141
7.3 Methods for dependence elicitation and modelling of multidrug resistance 142
7.3.1 Eliciting the marginal distributions of future antibacterial resistance rates 142
7.3.2 Eliciting and modelling the dependence between future antibacterial resistance rates 143
7.4 Results of dependence elicitation for multidrug resistance rates . 148
7.5 Discussion on informing the value of future antibiotics' R&D and chapter conclusions 154

IV General Conclusions and Bibliography 156

8 General Conclusions 157
8.1 Reflections on the research objectives and the connection between the research contributions 157
8.2 Re-addressing the research questions 160
8.3 Limitations of the presented research 163
8.4 Future research 163

Bibliography 168

List of Figures

1.1 Prescriptive decision/risk analysis process and research objectives 12

3.1 Schematic representation of modelling and elicitation context . . 37
3.2 Example Bayesian network with one child and three independent parents. 39
3.3 Conditional exceedance probabilities at u^{th} quantiles (rank correlations: 0.2 to 0.9). 43
3.4 Expert's conditional probability assessment as a function of the product moment correlation coefficient. 50
3.5 Conditional quantiles to rank correlations (perspective). 55
3.6 Conditional quantiles to rank correlations (contour). 55
3.7 Different perspectives on elicited dependence parameters' use in the case study literature. 62

4.1 Overview of the expert judgement process adjusted for eliciting dependence (steps discussed in this chapter are in grey). 69
4.2 Exemplary elicitation question with visualisation 89

5.1 Schematic representation of eliciting the conditional 50^{th} and 5^{th} quantile. 100
5.2 Convergence of exceedance probabilities for (selected) parametric copula families (rank correlation 0.3 (left) and 0.7 (right)). 100
5.3 Overview of Scenario Mapping Process. 106
5.4 Fitting parametric copula forms to combined assessments (rank correlation=0.55). 110
5.5 Unconditional scenarios for 50th quantile. 111
5.6 Conditional scenarios for 50th quantile. 112

6.1 Modelling context of the SRP method. 117
6.2 Partition example $QP(\tilde{u}, \tilde{v}; p, q)$ with solid lines for previously elicited quantiles and dashed lines for new ones. 119
6.3 Maximum case of 16 partitions (right) resulting from partitioning 9 rectangles (left). 120
6.4 Example of a quantile partition for assessing the upper tail. . . . 122
6.5 Quantile partition of the joint distribution from (i) to (iv). . . . 123
6.6 Further refining the assessment on the joint upper distribution tail. 124
6.7 Resulting quantile partitions after further refining the previous assessments. 125
6.8 Regions of interest for dependence assessment. 129
6.9 Outcome of eliciting the marginal distribution for each region. . . 131
6.10 The experts joint distribution: overall (left) and assessed upper quadrant (right). 132

6.11 Unconditional scenarios for assessments. 134
6.12 Conditional scenarios for assessments. 135
6.13 Different set of rectangles in $QP(u, v)$. 137

7.1 Schematic representation of elicitation sequence 147
7.2 Elicited marginal distribution (5^{th}, 25^{th}, 50^{th}, 75^{th} and 95^{th} quantile) for the rate of resistance of Escherichia coli isolates resistant to Carbapenems in the UK in 2021. 149
7.3 Elicited marginal distribution (5^{th}, 25^{th}, 50^{th}, 75^{th} and 95^{th} quantile) for the rate of resistance of Escherichia coli isolates resistant to third generation Cephalosporins in the UK in 2021. 149
7.4 Elicited conditional exceedance probabilities for quantiles $u = 0.5$ and 0.95 with fitted copulas per expert. 150
7.5 Elicited conditional exceedance probabilities for quantiles u=0.5 and 0.95 with fitted copulas per combined weighting. 150
7.6 Unconditional meta scenario of all experts for Escherichia coli resistance to third generation Cephalosporins in the UK in 2021. 152
7.7 Conditional meta scenario of all experts for Escherichia coli resistance to Carbapenems in the UK in 2021. 153

List of Tables

4.1 Main biases and heuristics for dependence elicitation 71
4.2 Overview of elicited forms 83
4.3 Dependence calibration results based on rank correlation, Gaussian (H_G) and Hellinger (H) distance (Morales-Nápoles et al. 2016b) 93

5.1 Features of existing qualitative structuring methods 103

6.1 Overview of dependence elicitation procedure and results. 132

7.1 Individual and combined multidrug dependence assessment results. 151

Part I

Introduction and Preliminaries

Chapter 1

Introduction

In decision making under uncertainty it is vital that dependencies between uncertain variables are appropriately modelled, as otherwise the model may not be fit for purpose. Dependent uncertainties may arise either directly because variables in the model are correlated, or indirectly when an uncertainty analysis of model parameters is carried out to explore model robustness. Both cases exhibit complex interrelations and dependencies which need to be considered if assumptions, such as independence, are not justifiable.
However, it is often not straightforward to either model or quantify dependence. In particular whenever no relevant historical data are available, the only sensible way to achieve uncertainty quantification is through eliciting expert judgements. When performed rigorously, the elicited quantities, often aggregated from multiple experts, offer reliable information for model quantification.
A challenge for anyone wanting to elicit dependence assessments from experts is the lack of research on structured expert judgement processes and elicitation methods for multivariate uncertainties. This thesis consists of three published works and two papers that are to be published (one under review and one working paper) which aim to address some of the main research gaps of dependence elicitation. Therefore, the original research of this thesis proposes procedures for some of the main elements of a structured expert judgement process for dependence - (1) the structuring and evocation of experts' knowledge and beliefs about dependence relationships together with (2) the flexible and feasible elicitation of detailed (quantitative) dependence assessments, in particular for quantifying models that include tail dependence. We apply both new methods in case-studies for validation purposes. The need for the original research in these particular areas has been identified by two extensive literature reviews which for themselves have made an original and publishable contribution to the dependence elicitation literature given that such comprehensive overviews of this area, including the link between dependence modelling and assessment as well as including the detailed elements of dependence elicitation processes, had been non-existent so far.

1.1 Dependence between multivariate uncertainties

In this thesis we consider expert judgement methods for various types of dependence models, justifying particular modelling choices in applications and case-studies based on a proposed taxonomy as well as model conveniences. Therefore, we discuss dependence modelling in this introduction more generally while addressing assessment for specific models later on.

Typically, the variables of interest in dependence elicitation (i.e. the elicited quantities) serve as inputs to some multivariate model. The form in which the information is elicited, e.g. through a conditional probability or a correlation coefficient, is chosen by an analyst who manages the elicitation process. Together with the definition of the underlying scenarios (the rationale) that determine(s) a specific assessment, the value and mathematical interpretation of the elicited quantity constitutes the knowledge (and type of information) we want to capture from experts.
In decision and risk analysis, a mathematical representation of dependence is a collection of random variables describing measurable risk characteristics [269], such as the number of lives lost, monetary losses and so forth. Formally, these random variables are denoted by $X_i \in \mathbb{R}$ for some $i \in \mathbb{N}$. For dependence, we are interested in the distribution of the random vector $\mathbf{X} = X_1, X_2, \ldots, X_n \in \mathbb{R}^n$ described by the joint probability distribution $P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n)$. Assuming probabilistic independence greatly simplifies the modelling process as we simply use the product of the marginal distributions to determine the multivariate distribution. However, whenever this assumption is not sensible, it follows that:

$$P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n) \neq \prod_{i=1}^{n} P(X_i \leq x_i)$$

In this case, we need a dependence model that approximates the unknown distribution by capturing the most important features of the dependence relationship. For instance in the later case-study applications, we are often concerned that the (in-)dependence structure in the central part of a model might not be representative of the structure in the tails. Therefore, we include tail dependence, i.e. the strength of association in the tails of the joint distribution, if applicable. Neglecting this aspect of a random variables can lead to distorted model output and hence poor decision making. For example, [263] discuss common (false) modelling assumptions in financial mathematics and their impact by reflecting on the famous *Wired* article about “the formula that killed Wall Street”.

1.2 Structured Expert Judgement

In research that aims at contributing on a practical as well as technical level to the area of expert judgement for decision and risk analysis, the two terms that are mentioned most frequently should be outlined at the beginning. Therefore, in the following the concepts of an *expert* and *elicitation* are briefly discussed and explained.

1.2.1 Who is an expert?

While a complete discussion about the definition of *expertise* together with *experts* and what distinguishes them from other professionals is given elsewhere (see, e.g. [50, 155, 85]), briefly and for purposes here an expert is (first of all) someone who is likely to make the best estimates of a target quantity (e.g. a model input).
Thus, a desirable feature of an expert might be a vast amount of domain knowledge (substantive expertise). However, we need to more precise when it comes to substantive expertise given that such knowledge is not only about facts (and rules) but also how to adapt this knowledge to new situations [50]. Therefore,

[50] links this ability of adapting experiences also to verification through feedback when solving new problems. This includes knowledge and experience on acquiring relevant data within new problem situations. In this regard, we note that experts typically have access to data as well as access to mathematical modelling tools whereas an expert could also “go back” and use a model and his experience in the interpretation of the data.
Substantive expertise is often mentioned together with normative expertise. The latter refers to formal methods to express domain knowledge [85]. When eliciting a dependence parameter (such as a conditional probability or correlation coefficient) from an expert, an understanding of these concepts is highly desirable. As we will re-address for instance in Chapter 4 of this thesis, the extent to which experts in a particular domain are expected to have normative expertise is likely to influence some main decisions of an expert judgement protocol, such as the form in which dependence is elicited. An example (that we will refer to later) is from weather forecasting where experts obtain very frequent data on specific correlation coefficients.
Nevertheless, being part of a specific domain does not guarantee good normative expertise and while the choice of experts through substantive expertise might be identified by considering (quantity and quality of) relevant publications, identifying normative expertise is more challenging [50]. For marginal probabilities, we do so to some extent through the Classical model for expert judgement [330, 79] (as experts might be excluded when aggregating results), even though this is still different to the earlier (first) identification of suitable experts. For dependence (as we will see), both, identification as well as aggregation, are less well explored.

1.2.2 What is elicitation?

Elicitation refers to a procedure that supports the expert to state his or her belief truthfully. For instance, [165] define it as “the process of formulating one’s knowledge about uncertain quantities in the form of a (joint) probability distribution for those quantities”. They refine this definition by stating that “an elicitation is done well if the distribution that is derived accurately represents the expert’s knowledge, regardless of how good that knowledge is”. Thus, a successful elicitation might be defined as the process that assists an expert in the rational and thoughtful evaluation of knowledge in order to eradicate personal bias, irrationality and superstition [306]. Going even further, it may be accepted that elicitation constitutes a misnomer as this would imply the crude idea that information for a judgement or an estimation is readily available to be “read off” an expert’s mind, indicating that the expert would simply look up the information within a mental “storage bin”, which seems rather controversial as shown by the Information Processing Theory which is further discussed in section 1.4 in the context of descriptive decision analysis research. Common guiding principles may therefore be often based on further properties of an elicitation, such as *reproducibility*, *accountability*, *empirical control*, *neutrality* and *fairness* which are all discussed in more detail in [79].

1.2.3 Dependence in the subjective probability context

Throughout this thesis we use the word *dependence* in a general sense (in contrast to specific association measures) when referring to situations where there are multiple uncertain quantities and gaining information about one would change uncertainty assessments for some others. More formally, two unknown quant-

ities X and Y, are independent (for experts) if their beliefs about X are not changed when given information about Y. For higher dimensions experts regard all quantities independent of one another if knowledge of one group of variables does not change their belief about other variables. Dependence is simply the absence of independence. It is a property of experts' belief about the quantities. This definition relates to [246] who reminds us that in a subjective probability context one expert's (in-)dependence assessment might not be shared with another expert possessing a different state of knowledge.
A particular consideration when discussing dependence between uncertainties in the subjective probability context is that of *observability*. In chapter 3, we list observability as a desirable property of an elicited form. Depending on the nature of joint or conditional events about which we elicit expert judgements in form of a specific dependence parameter, this might not be guaranteed. In particular for statistical dependence parameters (see chapter 3) and concordance probabilities, we note that these are not observable if they cannot be interpreted in terms of frequencies, i.e. if they are not based on populations. This is why we do not recommend their elicitation and rather suggest the elicitation of, for example, conditional probabilities which fulfil the desideratum of observability and are mathematically related. Nevertheless, some researchers suggest that even when an interpretation in terms of a population is not given, we can still elicit statistical forms and concordance probabilities whereas experts then need to think of hypothetical populations in order to make sensible dependence assessments. See [306, 74] for a more comprehensive discussion.

1.3 Research objectives, questions and scope

The research presented in this thesis is motivated by a holistic perspective on the overall expert judgement process and how the elements of such a process might be affected by different dependence model choices. In line with that, the following research objectives and questions have been developed as guidance for the research conducted in this thesis. They determine the scope and foci of the different parts which constitute this thesis. In other words, the following research objectives and questions specify and define the research problem that we hope to solve.

Research Objectives and Questions:

Research Objective 1: Propose a taxonomy for the current research on expert judgement for dependence and identify its future agenda in decision and risk analysis.

- Resesearch Question 1: Which dependence models are most prevalent in the decision and risk analysis research and for which of these has expert judgement been used to address the lack of relevant historical data? What are the foci of these models and how can they be considered in assessment methods?

- Resesearch Question 2: Which *key* elements of processes for eliciting dependence from experts have partly or completely been neglected in past and current research?

- Resesearch Question 3: What is the status-quo of related research topics of interest for cross-fertilisation from the elicitation of experts' (univariate) probabilities (e.g. behavioural research on cognitive fallacies) and to what extent have these taken into account in current dependence elicitation methods and processes?

Summarised Research Approach: Several systematic reviews of the literature on dependence elicitation have been conducted with different foci. Guidance is provided in each literature review on the current research gaps.

Research Objective 2: Propose methods that address the research gaps identified.

- Resesearch Question 4: How can a method support experts in structuring their knowledge and beliefs about dependence relationships in order to mitigate common cognitive fallacies and enhance experts' confidence in their dependence assessments?
- Research Question 5: How can we support experts in making feasible dependence assessments while not restricting the level of detail and flexibility of a dependence model as desired by a decision-maker?
- Research Question 6: How can we ease experts' cognitive burden when making detailed dependence assessments, in particular when assessing tail dependence?

Summarised Research Approach: Novel methods for qualitiative and quantitative elicitation of dependence have been developed and validated in case-study applications.

We will re-address the research objectives and questions when concluding the thesis and reflecting on the research findings.

1.4 Epistemological considerations and research methodology

When conducting research and contributing to a particular scientific community, we should address questions regarding the *scientific rigour*, *validity* and *methodological robustness* of our research. By considering these questions, we can shape our research design to adequately address the above research questions through a fitting research paradigm and methodology. A research paradigm encompasses the underlying set of presuppositions and values of our chosen methods. The presuppositions and values are heavily influenced by our philosophical assumptions whereas the main ones are about *the way we think the world is* (ontology), how this influences *what we think can be known about the world* (epistemology) and *how we think it can be investigated* (methodology and research techniques) [148]. A methodology can be seen as the strategy (or plan of action) behind the selection of particular methods whereas it links the use of methods to a desired outcome and uses them to gather and analyse data related to the research question [100]. In the following, we briefly discuss these concepts, their relationship

and the impact of their considerations on the research presented in this thesis.
Generally, the solution to a scientific problem can be illustrated by a triangle connecting the concepts of *theory*[1], a *(mathematical) model* and the *real world*[2] sitting on its vertices. This triangular embeds the idea that the solution to a scientific problem is obtained through a theory, explaining a phenomenon of the real world that is of interest, while producing a (mathematical) model which simplifies the real world phenomenon with the aim of manipulating it, e.g. for the purpose of prediction. The idea of triangular relationship originated with the Vienna Circle which comprised some of the main foundational contributors to the area of epistemology [52].
While the reader might think of various famous natural science theories and models that are embedded in the idea of the triangle, such as from physics, it similarly applies to the area of the research presented in this thesis. For example, according to [166] "the scientific contribution of Operations Research/Management Science is in the development of decision-aiding models". This view might be extended to similar and overlapping areas of risk and decision analysis. A well-known theory in this regard (and one that underlies the ideas of this thesis and will be addressed in more detail later) is the subjective expected utility theory by L. J. Savage.
For theories and models in decision and risk analysis, a traditional distinction between normative, descriptive and prescriptive research and approaches exists [37].
Normative research concerns ideal reasoning, judgement and decision-making under uncertainty. Theories, such as formal logic, probability theory and decision theory, provide rules for rational inferences and decision processes that are reasonable and consistent [365]. Irrationality is the systematic deviation from these rules. For a more detailed discussion and debate on the concept of rationality, see [379, 378]. More specifically, following these rules allows for choosing and maximising the optimal alternative among those available. Hence, by acting rationally, a decision-maker chooses the optimal combination of probability and utility. In this regard, expected utility theory [404, 405] together with its subjective version [357] provides the most established axioms for rational choice (whereas the axioms concern preference relations). The difference between both theories lies in the operational definition of uncertainty which is introduced in more detail in the next chapter. The evaluation of normative approaches is done through their theoretical adequacy [404, 405].
The descriptive research tradition stems from a psychological motivation and aims at exploring how people actually make decisions rather than ought to make them as emphasised in normative approaches. A central concept is bounded rationality together with the principle of *satisficing* (satisfy and suffice)[367]. Instead of choosing an optimal solution, a decision-maker chooses an adequate solution-based on a satisfactory perception of the decision's important aspects. Hence, it is argued that this is a more realistic approach to decision-making in practice by taking into account the limitation of human information processing. A prominent model for human thinking and learning is Information Processing Theory in which the human mind not solely responds to stimuli, but rather manipulates the incoming information by i.a. selectively perceiving, encoding,

[1] In the general sense, a set of propositions starting from premises and/or axioms and arriving at rigorous solutions.

[2] We accept a common-sense definition of "real world" even though it is much more discussed in the philosophical area of ontology.

storing (in short- and long term memory) and interpreting it with the use of experience [364]. Descriptive models can be evaluated through the extent to which they agree with observed behaviour, in other words, their empirical validity [37]. Lastly, for prescriptive methods and models, the aim is typically to address the (broad) research question of how can we improve the quality of decisions made in practice? Thereby, the discrepancy between normative and descriptive research findings has motivated the development of prescriptive methods. Thus, prescriptive research exploits proposals of normative theories together with empirical findings of descriptive researchers in order to support people in making better decisions, whereas better means rational and coherent as well as feasible [37]. Broadly, prescriptive approaches propose structured procedures for modelling a decision situation for i.a. increasing the understanding of the problem, exploring different courses of action and as well ensuring a sensible choice for a model's input data. A central focus of prescriptive methods is therefore the elicitation of values about consequences and judgements about uncertainties. Evaluation is typically achieved through a method's pragmatic values, i.e. its ability to support people for making better decisions [37]. The original research presented in this thesis proposes such prescriptive decision-aiding tools.

This discrepancy between normative, descriptive and prescriptive research approaches, which is also reflected in the differing backgrounds of the corresponding researchers[3] has led to the development of theories, models and methods under several research paradigms.

A common paradigm that determines the research strategy in the natural sciences is that of positivism. Briefly, in positivism, we (as researchers) aim at confirming or contradicting a hypothesis through evaluation and observation. Knowledge production occurs by means of a search for knowledge that is general and valid to the formulation of hypotheses [345, 394]. This view is more common among normative decision researchers.

It follows that in social sciences the question arises whether we can use a natural science approach to the research process while reality is regarded as a social construct and the researchers' subjective view on it determines the knowledge of the object studied - typically under the paradigm of interpretivism. Hence, social sciences seek to describe, understand, and reflect on human beings and their actions [345, 394]. Descriptive decision-aiding research can often fit in this paradigm.

Positivism and interpretivism are two opposing epistemological positions. However, both have as their mission the search for the truth, and their goal is to describe, explain, and advance knowledge in a given area [113].

Before considering a suitable paradigm for prescriptive research, let us briefly recall that, more generally, knowledge production can be understood as "the construction of universally accepted truths in a given historical time or as a process of learning of the subject" [410]. Thereby, [389] states that knowledge is produced through information from two sources (1) research authors who structure knowledge and (2) users applying knowledge when solving problems (in the real world).

For us as prescriptive researchers, the latter is particularly relevant. While on the one hand our research outcomes should ensure scientific rigour, on the other hand they should also be of relevance for practitioners. Scientific rigour in a

[3]Decision and risk analysis has particularly drawn the interest of disciplines such as mathematics, statistics, economics, computer science but also psychology, behavioural science and artificial intelligence.

research methodology is traditionally the key for its validity, for which research methods comprise a set of steps that are recognised by the corresponding academic community for creating rigorous scientific knowledge [118]. In the natural and social sciences, these steps are to explore, describe and explain. In prescriptive research however, we might not only explore, describe, and explain a given phenomenon, but rather we also study designing and creating artifacts [118, 366]. Generally, such artifacts are artificial objects or solution concepts whereas we can characterise them in terms of goals, functions, and adaptations. They are designed to effect some change in a system, solve problems and allow for a better performance of the system as a whole [366]. Here, these artifacts comprise prescriptive methods, models or algorithms, which serve a decision-aiding purpose. This allows to meet the other main criteria for evaluating prescriptive models, such as usefulness and pragmatic value to provide decision-makers with suitable and relevant assistance to improve their decision-making [37]. This is an important consideration as for instance [259] states that "the big problem with management science models is that managers practically never use them".

Design science is a research methodology which incorporates the idea of such artefacts and hence lends itself strongly to be a methodology for any research that aims to be of prescriptive nature. For a comprehensive overview of design science, see [118]. Traditionally this methodology has been developed and proposed for research in engineering where the principle of "build-test-build-test" is a preferential research paradigm. In management science, this is known as the management science process [419]. This is different to the natural and social sciences for which the common research paradigm is (as aforementioned) based on empirical observation. Nevertheless, this does not imply that we can ignore the importance of empirical observation in prescriptive research. In fact, we might see design science research as complementary to more traditional research methods as it occupies a middle ground between traditional scientific approaches and context-specific, problem-solving approaches that seek for knowledge creation by solving practical problems and testing novel ideas within applied contexts [118]. Validation is thus two-fold: (1) internal, from ensuring logical soundness and coherence together with a robust grounding in relevant theory and (2) external, by evaluating the proposed methods in the real world, e.g. through case-studies. For the former, we therefore base all proposed novel models and methods on such a robust grounding, while for the latter, we have chosen the research method of case-studies. With regards to how our research questions on the original research part in this thesis are formulated, this method complies with the idea that "case studies are the preferred method when 'how' or 'why' questions are being investigated, the investigator has little control over events, and the focus is on a contemporary phenomenon in a real-life context" [435]. Within the methodology of design science research, case-studies are a main way of obtaining data for verifying the robustness of new models and methods due to the breadth of ways to gather data, e.g. through observation, interviews, or questionnaires [118]. For us, in particular observations and interviews/feedback (after elicitations) have been a main source of verification data. The methodological robustness of our case-study research has been ensured by following generally accepted planning steps, such as *definition of conceptual structure*, *defining data gathering and analysis*, *defining ways to control the study*, *piloting the studies*, *recording data*, *analysing the data* and *generating research reports* [435]. Note that another research method in design science (which has not been applied in this thesis however) is action research [118]. As such, for evaluating prescriptive decision-making models and their pragmatic value, i.e. their ability to provide decision-makers with appropriate aid for better decisions, is a main criterion.

1.5 Contribution of thesis

Several contributions have been made within this thesis which are either published or currently under review for publication. The following papers (journal articles and a book chapter) have resulted from research for this thesis:

- ***Chapter 3:*** Werner, C., Bedford, T., Cooke, R. M., Hanea, A. M. and 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.

- ***Chapter 4:*** Werner, C., Hanea, A. M., Morales-Nápoles, O. (2018). Eliciting multivariate uncertainty from experts: Considerations and approaches along the expert judgement process. In: Dias, L. C., Morton, A., Quigley, J. (eds.) Elicitation: The science and art of structuring judgement, New York: Springer International Series in Operations Research and Management Science, 171-210.

- ***Chapter 5:*** Werner, C., Bedford, T. and Quigley, J. (under review). Mapping Conditional Scenarios for Knowledge Structuring in (Tail) Dependence Elicitation, Journal of the Operational Research Society.

  Awarded Winner of the Donald Hicks Scholarship by the UK Operational Research Society

- ***Chapter 6:*** Werner, C., Bedford, T. and Quigley, J. (2018). Sequential Refined Partitioning for probabilistic dependence assessment, Risk Analysis, doi: 10.1111/risa.13162.

  Awarded Runner-up in the INFORMS Decision Analysis Society Best Student Paper Competition

- ***Chapter 7:*** Werner, C., Bedford, T., Colson, A. and Morton, A. (to be submitted). Risk assessment of future antibiotic resistance - eliciting and modelling probabilistic dependence between multivariate uncertainties of bug-drug combinations, Working Paper.

1.6 Thesis outline

The remainder of this thesis is structured as follows. Chapter 2 (in this first part) presents the preliminaries that underlie the research presented thereafter. This clarifies some main definitions and the viewpoints of the author on the topic.
In Part II both, chapters 3 and 4, provide overviews of the literature on dependence elicitation and modelling in decision and risk analysis. Chapter 3 reviews several probabilistic dependence models together with common ways of eliciting dependence information for them from experts. In this regard, a taxonomy on how to classify dependence elicitation methods in different modelling contexts is presented. Chapter 4 provides an overview of the main elements of structured elicitation processes, typically used for eliciting univariate quantities, and discusses adjustments together with additional elements that are necessary in the

case of dependence elicitation.
Part III, consisting of chapter 5, 6 and 7, presents the original research of this thesis together with case-study applications. Chapter 5 introduces a method for mapping conditional scenarios which allows for structuring and recording experts' knowledge and beliefs on dependence relationships prior to a quantitative elicitation. Chapter 6 addresses the potential feasibility issue when eliciting detailed expert judgements for dependence - a topic of relevance when a decision-maker requires a high level of detail for a dependence model. Both previous chapters include a case-study to apply their novel methods with experts on real questions of interest - the former for a risk assessment in the UK Higher Education sector, the other on terrorism risk as of interest for insurance underwriters. Chapter 7 uses various of the earlier introduced methods in an application of future antibacterial resistance risk assessment.
Finally, Part IV concludes the thesis with the overall conclusions (chapter 8) and the bibliography.
Figure 1.1 shows how the chapters, which are published contributions, are connected and located within an overall framework of prescriptive models in risk and decision analysis. The framework is based on the discussion of my research design and methodology in section 1.4. The derivation of the framework and how the chapters fit into it is discussed in detail in the concluding chapter.

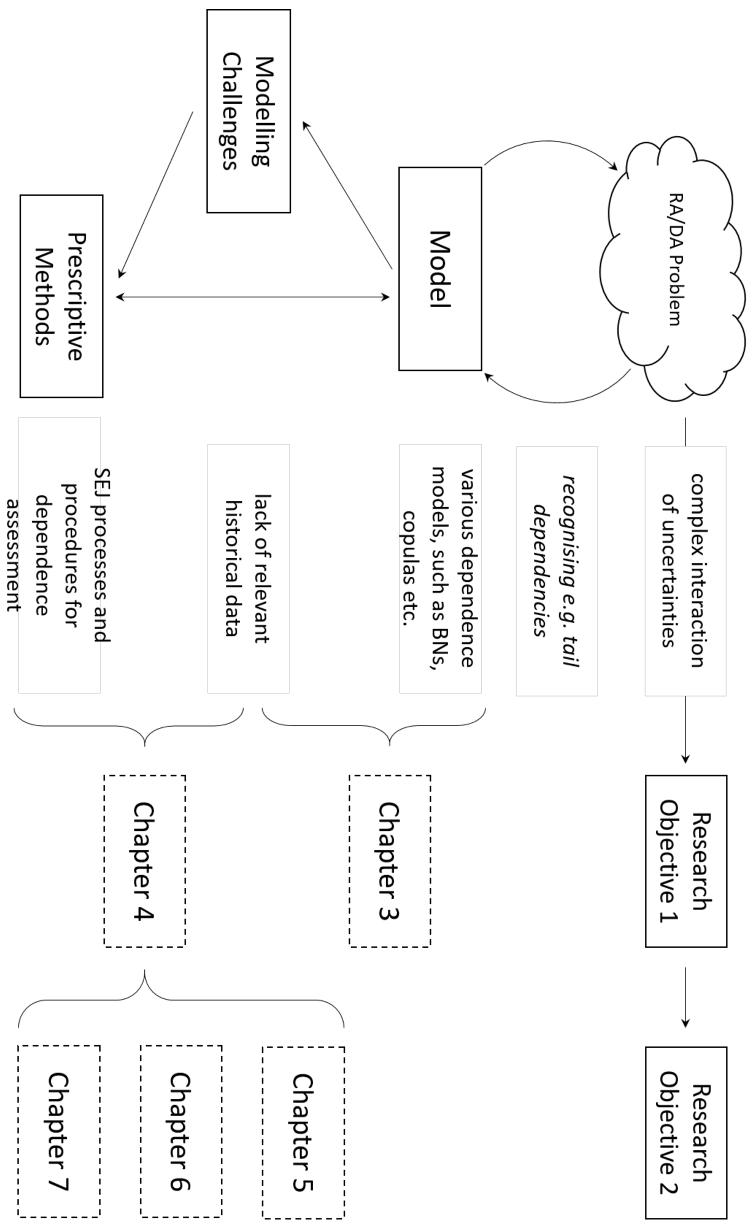


Figure 1.1: Prescriptive decision/risk analysis process and research objectives

Chapter 2

Preliminaries

This chapter introduces the different interpretations of uncertainty and ways to measure it while outlining its relation to the concept of risk and dependence. This allows for a precise and common understanding of subjective uncertainty which is fundamental to expert judgement elicitation as commonly used in probabilistic modelling of risk and dependence.
We clarify some theoretical and philosophical foundations as these fundamental concepts are often applied in an unclear manner despite their central importance in the decision and risk analysis literature. Some developments in this regard are non-probabilistic (alternative) interpretations of uncertainty and the confusion between uncertainty and risk.

2.1 Chapter introduction

The terms *uncertainty* and *risk* appear in various research areas and as such, numerous approaches to define, elicit and model it are proposed. This chapter presents the views of the author on these concepts which outlines the underlying principles of the research presented in the following chapters.
The definition of dependence as we use it here (see Part I, Introduction) relates directly to the scope of this thesis. The word *dependence* is used in many ways within decision and risk analysis, Operational Research (OR), Management Science and related fields, and it is worth clarifying how its use here differs from its meaning in other contexts. The underlying framework adopted is that of subjective probability which plays a key role within expected utility maximisation for decision making. Dependence then, refers to the way we model and assess the probability dependence structure required for such decision support processes. We do not consider non-probabilistic representations of uncertainty, nor do we consider approaches to represent dependence between criteria used to model the preferences of the decision maker as discussed widely in the multi-criteria decision analysis (MCDA) literature.
The foundations of subjective probability are drawn from a wide literature, in which [357] provides one of the most sophisticated accounts. In this account, probabilities can be assessed through preferences over lotteries, and there are implied consistency rules for preferences which can be empirically validated. It is well known that there is a distinction between normative and empirical validation, so the degree to which researchers choose to be led by normative or empirical consistency has led to many different approaches. For instance, [121]

provide a theoretical framework which attempts to tie these strands together in the context of possibility theory, and the implications of this are discussed in detail by [81]. The modelling of dependence between attributes in MCDA is the subject of a wide literature, and as discussed above, is outside the scope of this thesis. In the following, subjective probabilities are set in context to other approaches and their measurement together with some implications are discussed. The structure of this chapter is as follows. The next section 2.2 discussed the topic of uncertainty in more detail together with its types (2.2.1), measurements (2.2.2), interpretations (2.2.3) and operationalisation (2.2.4). Section 2.3 outlines the different views and definitions on risk. In 2.3.1, ways to define risk qualitatively are presented, before its quantitative definition (2.3.2) together with alternative definitions and misconceptions (2.3.3) is discussed.

2.2 Uncertainty

When discussing approaches to measure uncertainty, we should first clarify its different interpretations, which have been put forward throughout history, together with ways for establishing a categorisation according to its various forms. Before we do that however, we briefly note that in this thesis our starting point is to consider *uncertainty as a lack of certainty* rather than ambiguity. While [84] discuss this in more detail, it holds that the latter is removed by linguistic convention. As aforementioned, some non-probabilistic models of uncertainty often include ambiguity as a form of uncertainty which might lead to some fundamental distortions in its measurement (see [81]).
In order to approach a definition of uncertainty and its various interpretations, cognitive sciences offers a common distinction for probability judgement strategies and reasoning under uncertainty - the idea of reasoning from the *outside* versus the *inside* [248]. To illustrate this, [248] use an illustrative, stylised example about four brothers, Harpo, Zeppo, Chico and Groucho[1], setting bets on a horse race outcome.
The four brothers have a different approach to place their bets, Harpo and Zeppo reason in terms of similar events (outside) whereas Chico and Groucho use sensible degrees of belief focusing on the unique case at hand (inside). Despite the great simplification of this example, it indicates that uncertainty is what can be reduced by valid information or evidence that predominantly comes from observations or physical knowledge. An example for the former is the initial phase of some random process (similar to a classical card game). In this regard, the type of information indicates the interpretation of uncertainty. For [423] "uncertainty [...] is a function of the information that is available" and for [84] uncertainty can be reduced by observation.

2.2.1 Different forms of uncertainties

Identifying different forms of uncertainty follows already from thinking about and defining the concept of uncertainty itself as shown in the previous example. As discussed later, in risk analysis generally a distinction between those types is more a practical than theoretical necessity, e.g. the common understanding of

[1] If these names sound familiar to the reader, this might be due to their appearance in the family comedy *The Marx Brothers*, a fact pointed out to me by John Quigley as it was omitted in the original source.

probability concerning long-run repetition of events is insufficient for the kind of uncertain quantity for which typically expert opinion is wanted.
According to [423], the question of whether or not "such distinctions between different types of uncertainty have a solid foundational basis" should be negated therefore as he sees it as "fundamentally flawed", while he confirms that such a distinction has useful consequences for modelling and analysing complex systems. This can be seen as a reasonable approach to this discussion as at a certain level of detail (at least from a philosophical point of view) some of the following distinctions can be quite tantalizing.

Epistemic and aleatory uncertainty

The main distinction of uncertainty is made between *epistemic*, relating to imperfect states of knowledge or belief, and *aleatory* uncertainty, induced by randomness [306], i.e. regarding frequencies or proportions caused by stochastic process in the world [180]. The former originates from the Greek word *episteme*, meaning "pertaining to knowledge". Aleatory on the other hand comes from the Latin word for dice (alea) thus indicating its relation to randomness. Hence considering the idea that through learning we can acquire more knowledge about a state of uncertainty, this distinction is also often called reducible and irreducible respectively, even though in the common literature it is not limited to these classifications [196]. While the earlier example was quite simplified and therefore the distinction between epistemic and aleatory for each approach to reason under uncertainty seems straightforward, this is much more complex in real-world applications. This is also why some authors do not agree with this differentiation at all. For example, [431] stresses the point that "there is only one kind of uncertainty stemming from the lack of our knowledge concerning the truth of a proposition". Similarly, [423] introduces a coin tossing example with a fair coin and concludes that only a lack of knowledge bears uncertainty as in principle all necessary information about relevant physical laws could be known, reducing uncertainty completely. This position might not be without reason as from the 18th century classical determinism's perspective likewise epistemic uncertainty was the only acknowledged classification and only with an advance in physical sciences, such as the emergence of quantum mechanics, a more intrinsic nature of uncertainty was more and more accepted [428]. However, that this discussion cannot simply be reduced to the non-acceptance of the limit of knowledge is for instance shown by Heisenberg's (1901-1976) uncertainty principle, saying that it cannot be known both, not the position nor velocity of a particle, but this is still epistemic [308].
While this theoretical discussion remains open for different viewpoints, the practical usefulness of the distinction is commonly appreciated. Problems in which this becomes clear are when approaching a solution by breaking a complex problem down, whereas not rarely an effective way to do this is by decomposing necessary inputs, such as probabilities or utilities. [196] summarizes the usefulness of decomposing for epistemic and aleatory uncertainty by claiming: "When a distinction between stochastic and subjective uncertainty is not maintained, the deleterious events associated with a system, the likelihood of such events, and the confidence with which both likelihood and consequence can be estimated become commingled in such a way that makes it difficult to draw useful insights". Further, [84] list three explicit aims for making such a distinction, (1) making modelling choices clear, (2) providing the basis on which quantification takes place and (3) demonstrating the decision-maker the degree to which learn-

ing and reducing uncertainty (epistemic) can make a difference in the model output. Stating this usefulness and reaffirming that from a normative viewpoint both interpretations of uncertainty can be regarded as valid, the rest of this section will introduce other forms of uncertainty before showing how the notions of epistemic and aleatory uncertainty for a measurement and hence assessment approach can be used.

Parameter and model uncertainty

Two other types of uncertainty that are usually considered together are *parameter* and *model* uncertainty. A simplified modelling process can be viewed as three step approach, whereas after identifying important variables to be included and classifying how these variables interact in terms of a suitable structure, in the final step, necessary parameters are to be assessed [253]. Then a statistical model can be regarded as representing data in terms of variables' probability distributions with aleatory and parameters with epistemic uncertainty [306].
Parameter uncertainty is inherently epistemic as parameters are concerned with populations rather than samples and given that, whenever expert opinion is used, the whole population is simply not known, otherwise all the necessary data to make inferences would already been given [306]. Parameter uncertainty can be defined as the "uncertainty about the true value of a parameter in a mathematical model" [84]. Usually expert judgement is concerned with eliciting some unknown value in order to make inferences about a parameter, however the direct assessment of model parameters should be avoided as they are often fictional, not directly observable, and therefore not have a simple and intuitive interpretation (e.g. [431]). Further, viewing the elicitation of parameters critically, leads us to the reasoning that parameters can be described as "artefacts of a model" [423], so that [84] question the real-world interpretation of a parameter, hence also questioning how meaningful the notion of a true value from the above definition is? As an expert's subjective uncertainty, expressed in a personal probability, is only meaningful when representing uncertainty about the truth of some meaningful proposition, a meaning can only be given to parameter uncertainty if representing an observer's uncertainty about model predictions' accuracy on observable quantities [84].
Model uncertainty is in principle similar to parameter uncertainty and might be even regarded as a special case of it. A definition such as model uncertainty being "uncertainty about the truth of the model", again serves little given the common understanding that "all models are false" [84]. A meaning for model uncertainty must come hence from taking a different perspective due to the fact that good modelling practice requires a model no more refined than needed for the specific application [253] and usually accurate predictive quality serves this purpose [84]. One such a perspective is, as aforementioned, that model uncertainty is a special case of parameter uncertainty. This is the case whenever there are several models with uncertainty about which one fits best one's needs, e.g. has the most accurate predictive quality. Then, a single "supermodel" could integrate all models and a discrete parameter could indicate, which model to use, i.e. assessments of probability distributions for model parameters can be made upon the condition of the model being suitable [84, 423]. Another perspective comes from [253] replacing the term model uncertainty with structural uncertainty implying also a clear practical distinction of parameter and model uncertainty as herewith in the simplified three step modelling process those two uncertainty types can only occur in different steps, neglecting iterative amending and refining. A final per-

spective states that model uncertainty is "unequivocally epistemic" [308]. They ([308]) derive this position with an example about a modelled real-world process, a risk assessment of atmospheric dispersion for a chemical installation with a residual variability, which is surely stochastic as considered part of the natural process, therefore assumed to be aleatory. However, with further refinement, more conditions and details could be added, reducing or eliminating variation, thus implying that the removed component was epistemic. This discussion leads back to the earlier question whether any uncertainty is aleatory? Apart from the aforementioned theoretical discussion, it can be seen that for grasping modelling uncertainty, the distinction of aleatory and epistemic is fundamental and reflects mainly the first and third reason from the three given by [84] about why this distinction is useful.

2.2.2 Measuring uncertainty

After having clarified the main types of uncertainty, this section considers technical practicalities when eliciting an uncertain quantity from an expert - its (mathematical) language. Herewith, two common and closely related measures are *probability* and *expectation*. The origin of both concepts is the systematic analysis of uncertainty that can be traced back at least to Paccioli (1445-1517) and a riddle, known as game of *balla* or problem of points. Roughly it constitutes the situation when stakes in an unfinished game have to be divided, a similar idea that was behind Cardano's (1501-1576) theory of gambling and the mail correspondence of Pascal (1623-1662), Fermat (1601-1665) and De Méré (1607-1684) (actually Gombaud) [39]. Before talking about probability and expectation as ways to measure uncertainty, it shall be noted that for a full mathematical language or representation, three components must be given, (1) axioms that give a specification for the formal properties of uncertainty, (2) interpretations that establish the link between the axiomatic definitions and observable phenomena and (3) measurement procedures that offer practical ways to interpret the axiomatic structure [84].

Probability

According to [423], "probability is the mathematical language of uncertainty" and it is often emphasized in glowing terms, such as by Butler (1692-1752), English bishop and philosopher, who stated that "probability is the very guide to life" [428]. And indeed, even though in the general literature on expert judgement, various numerical expressions for use in elicitations are mentioned, i.e. to make statements about an uncertainty, e.g. relative and natural frequencies as well as percentages of chance or direct probabilities [306], in fact all those numerical expressions should give the same result for an assessment: a probability as a normalised measure. Though, at this point it should be mentioned that psychological research indicates that the formulation makes a difference, but this is not of concern in this chapter.

Loosely, states of knowledge can be indicated by probability distributions, so e.g. no knowledge can be shown with a uniform distribution whereas a distribution with a certain peak and range gives the most likely value together with its extremes. Formally, a probability is a positive normalised measure of uncertainty following certain mathematical properties. In common notation, $P(A)$ denotes the probability that event A occurs, whereas it can take any value on the interval ranging from 0 to 1. $P(A) = 1$ implies that event A occurs with certainty and thus $P(A) = 0$ signifies the impossibility of event A's realisation. A way to introduce probability is known in the respective literature as set theory, in which a σ-field is the collection of events which a probability can be assigned to. Then, Ω represents the non-empty set, a collection of all the possible outcomes or possible worlds, which need to be distinct and definite. On the other hand, $\varnothing$ denotes a null or empty set. A set $A \in \Omega$ is called an event and if all elements of set A are part of a set B, then A is said to be a subset of B, i.e. $A \subseteq B$ if and only if $\forall \omega \in \Omega : (\omega \in A \implies \omega \in B)$. Following this, a subset is called a proper subset $A \subset B$ if it not constitutes the whole set, i.e. there is an element $\omega \in \Omega$ being an element of B but not A. Further to these definitions, set theory allows operations for understanding the relation of different events. The most common ones herewith are *union*, *intersection*, *complement* and *difference*.

Formal properties and Kolmogorov's axioms While the earlier brief excursus into historical developments showed that the systematic analysis of uncertainty started at least around the 17th century, it was as late as the beginning of the twentieth century that formal properties of probability were outlined. Some prominent ideas brought forward were Keynes' partial ordering on a set of propositions [84], whose explicit aim it was to formalize the theory of probability by introducing it as logical relation [181], and conditional probability (e.g. [329]), i.e. relativised to an indicated set of outcomes as whenever a situation is modelled probabilistically, the set is delimited to outcomes one is ready to tolerate. For example, with a throw of a dice one might exclude the possibility of it landing on an edge by defining the outcome set as $\{1, 2, 3, 4, 5, 6\}$ [183]. Particularly followers of a subjective interpretation make the argument that an individual's probability is always based on prior knowledge and/or belief. From today's perspective, it is agreed that the Soviet mathematician, Kolmogorov (1903-1987) [235] succeeded in laying the formal foundation of probability when introducing the first axiomatic structure based on measure theory, where probability is axiomatised as a normalised measure [18]. It is commonly accepted for being simple, intuitive and suitable in many applications [84]. The axioms are:

- Non-negativity: $P(A) \geq 0$, for all $A \in \mathcal{F}$ where $\mathcal{F}$ is the event space and P the probability function;
- Normalisation: $P(\Omega) = 1$ where Ω is the sample space;
- Finite additivity: $(A \cup B) = P(A) + P(B)$, for A,B, $\in \mathcal{F}$ such that $A \cap B = \varnothing$

The third axiom can be applied similarly to an infinite sequence of mutually exclusive events, thus stating:

$$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

Then, $\{\Omega, \mathcal{F}, P\}$ is called the *probability space*. Consequences of the above axioms are the deduction of some common calculation rules for probabilities, such as the probability of the empty set, monotonicity, probabilities being bounded between 0 and 1, or as well the sum rule of probabilities.

Expectation

Another measurement of uncertainty is *mathematical expectation*, also known as mean, expected value or first moment. Consequentially, the n^{th} moment of a random variable X is the expectation to its n^{th} power, i.e. $E(X^n)$. In addition to the aforementioned contributors of measuring uncertainty, usually Van Huygens (1629-1695) is credited for having developed an axiomatic structure of what then became the theorems of expectation rather than probability [417]. Similar to above, the axioms of mathematical expectation are outlined as a formal foundation, whereas $E(X)$ denotes expected value of the random value X.

Formal properties and axioms As a normalised positive linear operator, expectations need to satisfy the following four axioms:

- Monotonicity: if $X \geq 0$, then $E(X) \geq 0$;
- Constants: if c is a constant, the $E(cX) = cE(X)$;
- Linearity: $E(X + Y) = E(X) + E(Y)$;
- Axiom 4: $E(1) = 1$

Despite the fact that expectation can be derived from coherent previsions as will be seen later when discussing de Finetti's principle of coherence, most arguably the more common approach is to induce it from probability-weighted average of all possible values of a quantity. This can be denoted by the outcome set Ω with a countable set of values $\{\omega_1, \omega_2, \omega_3, \ldots, \omega_k\}$ for the finite case, in which the expected value of random variable X is then denoted by:$E(X) = \Sigma_{t=1}^{k} P(\omega_i)X(\omega_i)$. Accordingly for the continuous case, Ω represents a real line and ω a scalar, while the expected value of random variable X is: $E(X) = \int_{-\infty}^{\infty} X(\omega)f(\omega)\mathsf{d}x$ for all $X(\omega)$ with a defined integral as well as being absolutely convergent, whereas $f(x)$ constitutes the probability density on Ω following: $f(x) \geq 0$, $\int_{-\infty}^{\infty} f(\omega)\mathsf{d}x = 1$.

2.2.3 Interpreting uncertainty

The next component necessary for probability and expectation to offer a full mathematical representation is a suitable interpretation linking formal properties with observable phenomena. The various interpretations are subject of ongoing debate and thus different perspectives are outlined to draw some practical implications from these, to be considered throughout this thesis. In this context, the subjective one is of main interest as it builds the foundation of subjective measurement processes and hence the application of SEJ. Several metaphysical desiderata have been imposed onto the different interpretations with the aim of establishing criteria of adequacy (e.g. [355]). However, as these rather philosophical approaches do not genuinely succeed in establishing an order of more and less adequate interpretations, in the following the main views over the course of history are put forth.

Classical

A first interpretation is often known as classical due to its early pedigree already championed by Laplace (1749-1827). It is the oldest way to interpret probability beginning with the analysis of games of chance, based on equally likely outcomes

[84, 423]. Within this interpretation, the probability of an event equals the number of outcomes that encompass the event divided by the number of all possible outcomes.
This way of quantifying a random situation is simple and intuitive, but also quite limited, not only in terms of the situations that can be analysed, but also restricting those games of chance itself to premises such as a perfect dice or a perfect coin. This means that by considering equally likely outcomes, a theoretical assumption is made which gives this interpretation also the name of theoretical probability. The term equally likely might become circular if seen as equally probable, i.e. making use of the notion of probability while establishing it. This can be overcome by the principle of indifference. While Laplace's version is limited to finite spaces, some approaches have been made to extend the classical interpretation to infinite cases, appealing to the principle of maximum entropy, whereas it can be regarded as a generalization of the principle of indifference as advocated by [205].
A rather curious example which transforms a typical situation of classical probability into an empirical one is mentioned in [428], referring to Weldon (1860-1906), an English biometrician, who conducted the laborious manual experiment of rowing $26,306$ rolls with 12 dice. He counted $106,602$ occurrences of 5 or 6, $1,378$ more than the expected number of $105,224$, i.e. an excess of $0,0044$ in the probability above $1/3$. In order to avoid a throwing bias those dice were rolled down a wavy cardboard.

Frequencies

As seen above in Weldon's simulation of rolling dice, another approach to interpret probability is by doing so empirically, such as in terms of its relationship to relative frequencies. This concept has with the classical interpretation in common that it takes an objective view and thus looks at interpreting probability independently of an individual's belief. Its simplest form relates to finite frequencies, for which simply a finite sequence of events is considered counting all actual outcomes in a long series of identical trials instead of possible ones as the classical interpretation would do. Pioneering work in form of a transition from a logical (see below) to a frequentist interpretation was mainly elaborated by Venn (1834-1923), who concluded, while working on the probability of male and female births, that "probability is nothing but that proportion" [400, 159]. Other prominent contributors are Von Mises (1883-1953) and Reichenbach (1891-1953), who both considered the infinite case of frequentism that extend limitations of the former case by identifying the limited relative frequencies of events within an infinite reference class, i.e. a hypothetical sequence of trials. Then, a probability states the limited relative frequency of *would be* events if the sentence was to be extended under identical conditions, concerning independent repetitions of a random experiment ad infinitum. Therefore it is also termed aleatoric interpretation [18]. Formally this is expressed as:

$$P(A) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} 1_A(\omega_i)$$

with ω_i being the outcome of an experiment i and thus $1_A(\omega_i)$ giving the indicator function.
An apparent problem with the frequentist interpretation is the non-integration of unrepeatable events, also known as the single-case problem. Further issues

arise when trying to claim a definite existence for a unique limit independent of the events' context, i.e. unaffected by whom repeats the experiment at what time [153], thus questioning the meaning of identical. Often the weak law of large numbers is stated to claim this existence, though this might introduce circularity as its assumptions must be defined beforehand, otherwise they themselves invoke probability [84]. To get around this, limiting relative frequencies are relativised to sequences as well as reference classes, a manipulation that brings its own problems. First of all, the question of ordering the outcomes within a sequence and then finding right reference classes, if this ever exists. [403] introduces what he calls collectives for solving this by defining algorithmic randomness and introducing two axioms of convergence and randomness, allowing that in a random sequence, or collective, the probability of a specific outcome is not altered by the selection of the sub-sequence. Given these theoretical issues, a common opinion in the respective literature is that this interpretation serves as evidence for probability assignments rather than making conclusive statements. For instance [335] claims in relation to the subjective interpretation (see below) that "the very idea of partial belief involves reference to a hypothetical or ideal frequency". For maintaining conclusiveness too often however, this view is for instance attacked from representatives of the logical interpretation, such as [229]: "to argue from the mere fact that a given event has occurred invariably in a thousand instances under observation, without any analysis of the circumstances accompanying the individual instances, that it is likely to occur invariably in future instances, is a feeble inductive argument [...]". Indeed objections could go towards the validity of inductive statements based on empirical observations (only) as shown next.

Logical

With the idea that probabilities can be determined a priori by considering all possible outcomes within a space, the logical interpretation shows similarities with classical and by introducing an empirical element with frequentist's reasoning. But it differs given that probability is seen as generalization of an empirical and epistemic relation between propositions. This generalization is a logical entailment in deductive logic, i.e. if valid, the truth of the premises guarantees the truth of the conclusion [159]. Then the central idea is capturing to which degree of confirmation (or: validity, partial entailment) some piece of evidence E supports hypothesis H, denoted as $c(H, E)$. Hereby only conditional probabilities (given evidence) are considered [184].
Logicism joins the subjectivist's view, which is of main interest for expert judgement, in the family of modern epistemic interpretations (distinguished from the classical one, i.e. not retaining the strict doctrine of determinism). Most prominently this interpretation has been elaborated in the 19th century by Boole (1815-1864) and De Morgan (1806-1871) and in the 20th century by Johnson (1858-1931), Keynes (1883-1946), Jeffreys (1891-1989) and Carnap (1891-1970). As main pioneer, Boole strongly advocates a normative position on probability theory which according to him must be derived from logic, thus implying that the logical interpretation does not describe how to actually reason under uncertainty but how to ought to. A strong implication for the kernel of the logical position is therefore the accent of rationality. This emphasis was expressed similarly by De Morgan, who even came quite close towards stating a subjectivist position while developing a normative perspective by rejecting the idea that probability can be objective and claiming: "by degree of probability we really mean, or ought to

mean degree of belief [...] I throw away objective probability altogether, and consider the word as meaning the state of mind with respect to an assertion, [...] on which objective knowledge does not exist" [159]. Rather than "throwing away" objective probability in the sense of a subjectivist however, De Morgan refers not to the actual belief of an individual but to that ought to be adopted, i.e. the universal assessment of everyone examining the subject in question, viewing the human mind as transcendental [159]. Johnson in [211] shares this view which is noteworthy given that he as lecturer at the University of Cambridge at that time was influential on later personalities such as Keynes, Ramsey and Jeffreys. In the 20th century, most famously Keynes argues that probability theory should be part of inconclusive, i.e. inductive logic, stating: "the Theory of Probability is concerned with that part [of knowledge] which we obtain by argument [rather than directly]". Therefore, Keynes rejects the frequentist position by stressing the important connection of induction and probability, similar to Bayes (1701-1761), and therefore introduces the concept of weighted arguments as a core ingredient of his theory, referring to the idea that favourable evidence increases weight, consequently affecting a probability: "the question comes to this - if two probabilities are equal in degree, ought we, in choosing our course of action, to prefer that one which is based on a greater body of knowledge?". According to [91] this introduces the difficulty of probabilities for probabilities, which might be problematic but is not further discussed here. In agreement with his predecessors it is clear for Keynes that given the same amount of knowledge, the logical relation demonstrating probability is the same for everyone [159], being "objectively fixed, [...] independent of our opinion". It might be questionable first of all whether ever several individuals can be said to have the same amount of information due to an individual's mental processing of information and secondly, it induces non-measurable probabilities. Thus, as a main critic of Keynes, Ramsey in [333] rejected the idea that "a probability may [...] be unknown to us through lack of skill in arguing from given evidence". This subjectivist view that an unknown probability does not make sense, a point also stressed by de Finetti in various occasions, is underpinned by Ramsey's belief of a psychological foundation of probability [385]. A position that is further explored in the next section.

Subjective

While sharing the conviction that probability is epistemic, the subjective interpretation disagrees that it is determined by a given body of evidence [159]. In other words, degrees of belief constitute an individual's actual degrees of belief rather than some objective or normative requirement in a transcendental sense as seen in the logical position. This interpretation goes thereby beyond the previous views as, being a property of a decision-maker, it can make use of probabilistic properties of physical situations or as well include evidence from past observations, but still evaluate relevance of information or assess single-case probabilities by resemblance. However, this position is not completely unconstrained as an individual still has to meet a single rationality requirement known as coherence, further outlined below.

Pioneering work in this interpretation was advanced by Donkin (1814-1869), Borel (1871-1956), Ramsey (1903-1930), De Finetti (1906-1985) and later Savage (1917-1971). Among these, Donkin might be debatable as his view on "quantities of belief" is very close to the logicist position of De Morgan outlined earlier. Due to his work on belief conditioning however, stating that proportionality among

probabilities assigned to options must preserve even if changing one probability of a hypothesis given new information, he shows a subjectivist mind-set [159]. More clearly does Borel pave the way for a subjective interpretation, defending the idea that the probability assigned to an event by two individuals with the same information can indeed be different. For him probability judgements are necessarily relative to a "body of knowledge", which is "necessarily included in a determinate human mind, but not such that the same abstract knowledge constitutes the same body of knowledge in two distinct minds". Further, Borel's epistemic account has strong affinity with the later work of Ramsey and de Finetti as he (Borel) appeals to betting as a method that "permits us in the majority of cases a numerical evaluation of probabilities". Similarly, Ramsey holds the position that probability as degree of belief and herewith probability theory as logic of partial belief has "no precise meaning unless we specify more exactly how it is to be measured" [385]. While an operative definition that specifies how degrees of belief can be measured will be discussed in more detail later, it should be noted that betting is one method to do so. In Ramsey's words, it is "the oldest established way of measuring a person's belief to propose a bet, and see what the lowest odds are which he will accept" [385]. Ramsey already knew about several drawbacks of betting methods, such as diminishing marginal utility of money, a certain arbitrariness due to an individual's "eagerness or reluctance to bet" and the issue that "the proposal of a bet may inevitably alter [an individual's] state of opinion", which is why he then refines and formalizes his approach. To do so, in *Truth and Probability*, degrees of belief are defined in terms of more general personal preferences, based on expectation and relative to a set of alternatives, obeying a set of axioms and yielding an expected utility representation. By satisfying the laws of probability, consistency similar to de Finetti's idea of coherence is assured, the only condition for rationality on degrees of beliefs [159].
Independently, de Finetti developed a theory with analogous conclusions, which prepared the ground of a "grown-up" subjectivism. A main point of his contribution was de Finetti's notion of *exchangeability* in his *Representation Theorem*. In this context, for one of his illustrations regarding drawing from an urn with unknown composition, he concludes therefore:

> "what is unknown here is the composition of the urn, not the probability: this latter is always known and depends on the subjective opinion on the composition, an opinion which changes as new draws are made and the observed frequency is taken into account"

In a more philosophical perspective, he sees the Representation Theorem as an answer to Hume's sceptical empiricism by justifying "why we are also intuitively inclined to expect that frequency observed in the future will be close to frequency observed in the past" [385]. It relates to the point that for de Finetti, the notion of an objective or true probability "out there to be found" is thus meaningless as he made clear with his famous sentence: "Probability does not exist". For him, the main problem with this objective probability is that it confuses defining and evaluating probability, so that looking for a correct objective probability, constraints these interpretations unnaturally. For his subjective view, de Finetti knew that in practice probability assessments are to be influenced by several factors, such as an assessor's competence, optimism, pessimism and alike, so that the topic about how to attain good probability assessment emerged in particular in cooperation with Savage [356]. The latter refined the subjectivists' approach by synthesising Ramsey's and de Finetti's work and adding Von

Neumann's (1903-1957) and Morgenstern's (1902-1977) expected utility theory. Before regarding a respective measurement procedure for our purpose, two other ways to conceptualize probability are briefly sketched for completeness.

Propensity and best system

Two further, however less common, ways of interpretation are the propensity and best system interpretations. While their discussion is out of the scope of the research presented later, the interested reader is referred to [329].

2.2.4 Operationalising subjective probabilities

After having looked at ways to interpret probability and contrast subjectivism with other views, a practical focus on the axiomatic structure is adopted by outlining the operationalisation of degrees of beliefs. For this, it is first of all necessary to consider what is measured and how it is done. [258] makes clear that "there are no absolutes in the world of measurement", meaning that "all measurement is based on a comparison with a standard". [81] outlines that this standard ought to have an operational definition, i.e. a rule indicating how mathematical notions are to be interpreted. Besides its general imperative, this might be even more critical for subjective uncertainty given the need to define in empirically observable terms what a respective measurement represents. This concerns the question and respectively the type of quantity to be elicited, how those might be interpreted by an expert and the meaning an analyst gives to them.

In the remainder of this section therefore, the betting approach of de Finetti is described, building onto the earlier discussed foundational ideas of Ramsey and de Finetti, though the focus here is on the technical elaboration. Following this, Savage's rational preference is presented as a way to offer a refined theory and axiomatic structure of subjective probability by synthesizing Ramsey's consistency and de Finetti's coherence with expected utility theory.

It should be mentioned that those are not the only approaches to operationalise degrees of belief. However, according to [81], Savage's way "is the best from a philosophical viewpoint" by fulfilling the criteria of using observable phenomena in terms of betting and more general choice behaviour based on rational preference, doing so by integrating operational definitions for both, probability and expectation.

De Finetti: Betting odds, coherent previsions and the Dutch Book Theorem

As aforementioned, betting odds are a suitable starting point for operationalising degrees of belief by inducing coherence through probability axioms, the only demand for rationality, proven e.g. via the notion of a Dutch Book. Here, de Finetti already outlines potential issues of using monetary amounts in betting (similar to Ramsey's concerns presented earlier) and comments that despite avoiding those by considering sufficiently small stakes, he admits that using expected utility "could be better". For an introduction to the original ideas, see [107].

In a betting situation (with an illustration from [313]), let Φ be the indicator event that a certain event occurs, e.g. in sport betting that a certain team wins, in which case Φ is true and takes value of 1, otherwise 0. The bet itself can

be seen to be worth a stake S if Φ is true (1) and nothing if false (0). The bookmaker then sells the bet at price $\pi_\Phi S$ in units of the stake. When betting in favour of Φ, the odds are denoted as $\pi_\Phi S : (1-\pi_\Phi S)$ and the action of betting itself is represented by $a_{\Phi,S}$. Considering someone willing to buy this bet, a net gain of $(1-\pi_\Phi)S$, i.e. the stake less the respective price, is obtained with $\Phi = 1$. Otherwise, whenever $\Phi = 0$, net gain is $-\pi_\Phi S$, i.e. a loss. For selling this bet, (plus/minus) signs are reversed. Once the odds are finalised, a seller of a bet is assumed to be indifferent on selling or buying, known as having fair odds. More formally, fair odds are denoted through indifference by $a_{\Phi,S} \sim a_{\Phi,-S}$, whereas in contrast, $\succ$ designates a situation with preference, e.g. $a_{\Phi,S} \succ a_{\Phi,-S}$ signifies that odds in favour of Φ are considered by a bookmaker.
Based on such a betting situation with fair odds, requirements for an individual's degrees of belief to be coherent can be derived. The argument is then that a degree of belief has to satisfy the axioms of probability as otherwise it is susceptible to a *Dutch Book*, meaning that it can be exploited. A Dutch Book is possible, if irrespective of Φ, the gain was strictly positive, meaning the seller of a bet would be a sure loser through incoherence [313]. In this case the seller of a bet with respective degrees of belief about some event is worse off, e.g. economically, irrespective of the state of the world. [428] outlines a parallel when thinking about coherence in subjective probability and the principle of no risk-free profit, i.e. no arbitrage, in the activity of trading.
In order to avoid a Dutch Book by following probability axioms, two structural assumptions need to be in place. The first is that the odds are fair, leaving aside concerns about its unique existence. This guarantees that a bookmaker's knowledge about the uncertainties is implies the odds rather than the aim to make a profit. The second assumption is that there is no limit on the number of bets that can be bought or sold as long as this number is finite. In de Finetti's words this is in place to "purify" the notion of probability from factors related to utility and it implies that the next monetary unit is valued as the former [183, 313]. Given those assumptions, a degree of belief can be "plugged" into formal probability calculus, represented e.g. by Kolmogorov's axioms, even though other axioms, such as for expectations, can also be used. For the corresponding proofs, see [313].

Savage: Rational preference behaviour and the axiomatic development

As mentioned previously, Savage developed, besides a more general approach to the axiomatic structure of subjective probabilities, a system that integrates coherence with expected utilities by interpreting degrees of belief as rational preference, operationalised with choice behaviour [81]. For Savage, preference can be, roughly speaking, conceptualized as choosing between acts for the sake of those act's consequences, which themselves depend on the state of the acting subject [84].
Nowadays, the standard literature for rational preference is Savages *The foundations of statistics*, which has been praised the "most brilliant axiomatic theory of utility ever developed" [147]. The fundamental idea shall be illustrated in the following simple example from [84]. Here a subject is offered the choice between "receiving 100 if it rains tomorrow, 0 otherwise" or "100 if the Dow-Jones goes down tomorrow, 0 otherwise". Supposing that 100 are preferred to 0, then, given the subject chooses the first bet, we can conclude that the subject's degree of belief about the event of "rain tomorrow" is *at least as likely* as "the Dow-Jones

going down". While this is a simplified example, it can be concluded that by observing enough choices, subjective probabilities can be derived accurately. While a complete and formal elaboration of the axiomatic structure for rational preferences can be found in [357] as well as in [84], for the remainder of this section, only main points of Savage and later general provisions of subjective probability axioms are provided.

Generally speaking, the purpose of Savage's theory can be summarized as (1) defining more probable than relationships among states of the world, (2) elaborating these relationships to derive probabilities on the states of the world and (3) working with von Neumann and Morgenstern's theory to integrate expected utility. Before this is achieved, Savage uses his first and fifth axiom to assure no indifference among a set of states, also called *transitivity*, shown more formally later. Following, a cornerstone of Savage's system (captured in his second axiom) contemplates the *sure-thing principle*, which resembles the von Neumann and Morgensterns independence axiom, necessary for the fundamental notion of *conditional preference*. It can be regarded as fundamental as it concerns preferences between actions after some conditioning event has been observed, so it relates to the connection of axiomatic theories and statistical practice [147].

After these initial clarifications, unique probability distributions can be derived from preferences, achieved with (1) qualitative probabilities in form of more likely than statements and (2) a transformation into quantitative terms by imposing additional restrictions. For the first, the definition of a null state is needed, simply meaning that one is indifferent if comparing two acts conditional on a null state, i.e. the subjective probability of a null state is zero. Based on this, Savage's third axiom "is so couched as not only to assert that knowledge of an event cannot establish a new preference among consequences or reverse an old one, but also to assert that, if the event is not null, no preference among consequences can be reduced to indifference by knowledge of an event" [357]. Similarly important for the nature of the relation of preferences, the fourth axiom ensures that as long as respective utilities stay in the same relation, e.g. $U_1 > U_2$, then the preference relation should not change. These steps so far make it possible to define a more likely relation and thus the idea of qualitative probabilities has been completed. For transforming those qualitative probabilities into quantitative ones, several possibilities to achieve this exist. One relates to [107], who proposes a finite partition condition, i.e. splitting up the set of all possible worlds into an arbitrarily large number of equivalent subsets, so that the quantitative probability of each of these is the inverse of it. Savage though takes a different approach in his *Archimedean axiom* (his sixth axiom), in which he embeds the finite partition condition directly in terms of preferences. While it leaves aside the direct restriction onto the qualitative probabilities directly, it requires splitting up the set of all possible worlds in small enough pieces so that preferences are unaffected by a change of consequences. Herewith no consequences can be infinitely better or worse, similar to von Neumann and Morgenstern's third axiom, as well as the set comprising the states of the world is rich enough to split it up into smaller pieces [147].

At this point, the axiomatic structure offers all necessary and sufficient conditions for a unique probability representation, so that in the next steps, Savage follows von Neumann and Morgenstern's expected utility theorem to derive utilities and a representation of preferences.

While this representation only gave a brief insight into Savage's theory, it is nowadays a more general idea of axiomatic structures for subjective probabilities to start from qualitative or comparative probability relations, which is usually seen as criteria for coherence [147]. Herewith usually three types of subjective likelihoods are distinguished:

- $A \sim B$, events A and B are thought to occur *equally likely*;
- $A \succcurlyeq B$, event A is thought of to occur *at least as likely as* event B;
- $A \succ B$, event A is thought of to occur *strictly more likely than* event B

From this basic notation some general axioms can be outlined, whereas broadly they can be split up into a group that enjoys no serious criticism and is uncontroversial, while the other group is debated thoroughly [147]. The ones that qualify for the first distinction are:

- Asymmetry: if $A \succ B$, then not $B \succ A$;
- Non-Triviality: $\Omega \succ \varnothing$;
- Non-Negativity: $A \succcurlyeq \varnothing$;
- Monotonicity: if $A \supseteq B$, then $A \succcurlyeq B$

The more challenged axioms in the respective literature are thus:

- Transitivity: if $A \succ B$ and $B \succ C$, then $A \succ C$;
- Additivity: if $A \cap C = \varnothing$, then $A \succ B \iff A \cup C \succ B \cup C$;
- Complementary: if $A \succ B$, then not $A^C \succ B^C$

Despite their categorisation here however, it should be noted that axioms like the ones above are fundamental for any theory of subjective probabilities.

2.3 Risk

[434] state that "if we were to read ten different articles or books about risk, we should not be surprised to see risk described in ten different ways". After having discussed the vast topic of uncertainty in the previous section, it follows almost "naturally" to consider understanding the concept of risk together with outlining respective measurement approaches. This will allow for making a clear distinction between risk and uncertainty as well as introducing at the same time the conceptual relationship between the two, which seems necessary due to some inherent potential for confusion. Having a clear idea about the concept of risk and how to work with it, both, qualitatively and quantitatively, will be particularly important in later sections on the presented original research.

2.3.1 Qualitative aspects and definitions

A common misconception when discussing and defining risk concerns the difference between risk and uncertainty. While probability theory builds the foundation of many quantitative risk assessments and earlier, probability was introduced as a way to measure uncertainty, it seems clear that uncertainty is involved whenever we talk about risk, but as it is not the same concept. Thus, something else needs to be present to give a first definition. A common and straightforward qualitative definition of risk that is used here as a starting point is thus: *Risk =*

Uncertainty + Damage [223]. In this first approach to conceptualise risk, some sort of loss or adverse event happening is considered. One option to do so is taking into account a common interpretation where risk is related to variance in expected returns or standard deviation. In contrast, another interpretation often seen in the psychology literature links risk to the probability or size of potential losses [126].

Besides the confounded use of the terms uncertainty and risk, another expression often found is that of *hazard*. It stems from the Arabic word *al zahr*, which translates into "dice". It refers to some of the earliest (gambling) games known to humankind, which used *astragali* (knuckle-bones) instead of dices [39]. Nowadays, hazard usually refers to a "source of risk" [84], meaning that risk itself incorporates the likelihood that this source actually results in a loss or damage. Additionally, for a second qualitative definition, the concept of a *safeguard* needs to be considered as several factors, such as the awareness or acknowledgement of a risk per se, can either increase or decrease the likelihood of the hazard to be triggered. Together this, a second definition can be expressed as: *Risk = Hazard / Safeguard* [223]. This definition is important as it verifies the earlier given statement that risk can never be reduced entirely, which forces us to compare and choose between risks.

After having outlined some intuitive interpretations of qualitative nature, below a way to quantify risk is given, which was chosen as it is regarded as a complete approach and offers practical validity together as well as a suitable starting point for the discussion of different definitions that follow.

2.3.2 A quantitative definition

Similarly to the qualitative definitions towards the conceptualization of risk, [223] offer a quantitative approach which shall be taken as an appropriate idea behind the concept. Their quantitative approach is built onto three questions to be answered by a risk analysis:

- What can happen?
- How likely is that it will happen?
- If it does happen, what are the consequences?

Based on the answers to these questions, a scenario list is drawn up, consisting of the actual scenarios, s_i , the likelihood of each scenario, p_i , and the consequence of each scenario, i.e. the measure of damage, x_i. These form a triplet and the risk R can then be expressed as the set of triplets, i.e. $R = \{s_i, p_i, x_i\}$, $i = 1, 2, 3, \ldots, n$. Re-arranging the scenarios in increasing order regarding the severity of their consequences, i.e. $x_i \leq x_2 \leq x_3 \leq \cdots \leq x_n$, and adding the cumulated probability, P_i of the scenarios, now allows to plot a staircase function with $\{x_i, P_i\}$. Smoothing it out, so that it displays an actual risk curve, will allow to consider the continuous case instead of the discrete one, which resembles reality better, whereas each scenario has a categorical function, meaning that within it, several different sub-scenarios, resulting each in different consequences, are entailed. A consequence of this representation is that risk can be defined as "probability or likelihood and consequence" instead of the prevalent notion in the respective literature of "probability times consequence". Though the former is not preferred without reason as with a single scenario it avoids the equivalence of "high probability - low damage" and "low probability - high damage" scenarios while with multiple scenarios it avoids the consequence of representing risk as the expected value of all consequences, i.e. the mean of the risk curve only.

Nevertheless, presenting the whole risk curve still seems to be representing risk not precisely enough as one might question the confidence in the curve itself, or if the curve represents the state of confidence, one might question the confidence in the state of confidence, a thinking easily leading to a problem of infinite regress. In other words the probability here is seen as a measure of uncertainty about future events and consequences based on some assessor's background knowledge. A tactic to overcome this is commonly known as the "probability of frequency approach". This additional detail however does not change the overall representation of risk apart from offering the possibility to assign confidence to different risk curves, even though as it introduces a second-order probability. It should be mentioned that those are not completely uncontroversial among philosophers as well as statisticians [282]. Defining risk in this way offers a representation for integrating uncertainty suitably for common risk analysis tools, such as fault and event trees as well as the risk assessments methods presented later in this thesis. For the latter, we might set a particular focus on considering multiple dimension of a risk definition whereas it seems intuitive that the triplet can be extended in different ways that may refine it. As already pointed out earlier, one number as risk measure seems too simplified and even a curve might not be satisfying. In this regard, [223] specifically mention that the consequence or damage, x_i, can be seen as a vector, entailing different damage types rather than a scalar and thus transforming the risk curve in a multidimensional space. A further refinement respects possible interdependences within the vertical line of s_i as well as the consequences x_i. Thus it overcomes the assumption of independent scenarios.

2.3.3 Some alternative risk definitions and misconceptions

As aforementioned, the topic of risk with its different approaches and definitions is a rather complex area and some might even call the present characterisations weakly justified and inconsistent [20]. Therefore a detailed discussion of as many definitions as possible gets confusing rather quickly. However, some grouped classifications that were identified by [21] are outlined to provide some structure.

A first one is *Risk = Expected Value (or Loss)*, a view that is most arguably the oldest definition of risk going back to the French mathematician De Moivre (1667-1754). Similarly, sometimes risk is referred to the expected loss solely [421], to utility, expressed as the product of probability and utility, or expected disutility solely [62]. As mentioned earlier, these definitions and hence measurements are closely related to finance theory as well as economic perspectives on risk. Even though in practice, utility is not often used directly, Cost-Benefit Analyses often determine the acceptability of taking a risk, which in turn is indirectly but still strictly related to utilities [23]. In fact, in these cases utility is expressed often in monetary values in form of Net Present Value (NPV), which is not trivial when thinking about expected loss of lives or irreparable damage to nature. This shows a certain degree of arbitrariness when working with utilities (directly or indirectly) in risk analyses despite its acceptance as a rational way to make decisions.

Another common understanding and measurement of risk is based on *Risk = Probability (of an undesirable event)*. Within this view, the damage or consequence is specified so that solely its likelihood is considered. This can be done

in most arguably its simplest form by assigning a probability to the chance that the loss occurs and respectively obtaining the probability that it does not. It becomes clear that one crucial consideration is specifying the amount of precision that goes into the definition of the damage itself, i.e. particular time frame, circumstances etcetera [230]. On the bottom line it becomes clear that within this view measurement of risk and risk itself are the same and so an assessor's perception and, as some might argue for its existence, an objective risk is not distinguishable. In a related but different perspective on risk, *Risk = Uncertainty*, risk itself and its measurement can be seen as different ideas though, which is welcomed by some researches arguing that "uncertainties beyond the probabilities should be taken into account" [19]. Referring back to the introduction of this chapter, alternative ways to measure uncertainties might be questioned.
Another grouping of definitions is summarized as *Risk = Objective Uncertainty* whereas the distinction is made between risk as objective and uncertainty as subjective probability. This idea came to "fame" especially in economical contexts, most often attributed to Knight (1885-1972). There are various objections to this definition, one of the most obvious ones being that risk analysis concerns a situation in which there is seldom knowledge about the objective distribution if it exists at all from a philosophical point of view. Further, as seen before, uncertainty is an inherent part of the concept of risk, so that taking those two apart only adds confusion if following the approach to risk that was outlined above.
While the so far presented groupings have concerned uncertainty itself or a certain measure of it, some other definitions, traditionally coming from an engineering context , take the starting point from the earlier described triplet or a derivation of it, so that scenarios and consequences are included. For instance, the definition of *Risk = Event or Consequence* as expressed e.g. by [346] in terms of "risk is a situation or event where something of human value (including humans themselves) is at stake and where the outcome is uncertain" belongs into this range as all components of the earlier idea of a triplet are included. However, it constitutes a far less practical concept in the daily usage of risk applications, such as the ALARP principle, as it does not allow for discriminating between high or low risk, comparing different risky options and so forth [25].

Part II

Literature Overviews

Chapter 3

SEJ for common probabilistic dependence models

This chapter[1] addresses the point that in expert judgement studies, a structured approach to eliciting variables of interest is desirable so that their assessment is methodologically robust. One of the key decisions during the elicitation process is the form in which the uncertainties are elicited. This choice is subject to various, potentially conflicting, desiderata related to e.g. modelling convenience, coherence between elicitation parameters and the model, combining judgements, and the assessment burden for the experts. While extensive and systematic guidance to address these considerations exists for single variable uncertainty elicitation, for higher dimensions very little such guidance is available. Therefore, this chapter offers a systematic review of the current literature on eliciting dependence. The literature on the elicitation of dependence parameters such as correlations is presented alongside commonly used dependence models and experience from case studies. From this, guidance about the strategy for dependence assessment is given and gaps in the existing research are identified to determine future directions for structured methods to elicit dependence.

3.1 Chapter introduction

When performed rigorously, elicited quantities, often aggregated from multiple experts, offer reliable information for model quantification. Nevertheless, there are several different broad approaches and many choices to be made by the analyst, all of which can affect the elicitation burden for experts and ultimately also the reliability of the outcome.
While research and reviews that offer guidance exist for methods addressing the elicitation of univariate quantities [79, 150, 155, 207, 306, 310], and while dependence modelling is an active research area [244], little guidance exists about the elicitation of dependencies. The exceptions are Bayesian (Belief) nets (BNs), though also for these modelling and elicitation challenges remain, as shown later. In fact, developing defensible elicitation processes for multivariate quantities is still much under development despite its fundamental importance for decision

[1] Based on: Werner, C., Bedford, T., Cooke, R. M., Hanea, A. M. and 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

as well as risk analysis [298, 372]. Some of the first studies that elicit dependence are [89, 227, 243, 172] and [215]. Since then more ways for quantifying multivariate distributions and models through experts have been investigated, yet on the actual elicitation only little discussion and guidance is available. References that introduce some aspects are [104, 244, 306] and [165]. However, a complete and systematic way of comparing different dependence parameters as elicited quantities, and reflecting their use in dependence models in the form of a literature review has been non-existent so far. Therefore, research and applications of several dependence measures in models and their elicitation methods are presented. With a practical focus, case studies are discussed whenever available. This chapter addresses elicitation processes for dependence information and aims at providing understanding of their use in applications. It offers guidance on making robust choices about which summary of expert knowledge on multivariate distributions should be elicited, and how they might be used within a dependence modelling context, as these are key decisions within the overall elicitation process. This is achieved by outlining how much is understood about the complexity of approaches to dependence modelling and the cognitive assessment burden for experts.

A comment on the scope of this review is that while we discuss the cognitive complexity of assessing dependence in various ways, such as already considered by [241], and while insights from psychological studies are mentioned, corresponding research streams for causal and association judgements are not reviewed exhaustively. Normative and descriptive models for causal reasoning or mental conceptualisation of correlations, which origin is often attributed to [371], are found for instance in [279, 177, 41] and [13]. An overview and introduction to these areas is given in [192] and [359].

The chapter is organised as follows. Section 3.2 discusses the extent to which findings from eliciting univariate quantities apply to the elicitation of multivariate ones in order to provide the reader with an indication for the scope of the overall topic. Section 3.3 introduces the modelling context which shows how modelling and eliciting dependence are related. This offers an overall structure to the research problem. Then, Section 3.4 discusses how elicitation is approached for quantifying various dependence models. Section 3.5 presents dependence parameters that are commonly elicited together with its implications for experts' assessment burden. Section 3.6 provides an overview of the empirical contributions in the literature based on which Section 3.7 formulates directions for future research and concludes the chapter. We refer to Appendix B whenever a technical term needs a more detailed explanation, however the original references should be considered for an extended introduction.

3.2 Generalisations of univariate elicitation processes for eliciting dependence

Structured processes for the elicitation of dependence follow historically from findings made when eliciting univariate quantities. In the early days of uncertainty modelling, formal processes for eliciting univariate uncertainties, such as marginal probabilities, were developed to ensure a methodologically robust approach to parameter quantification in the face of lacking relevant historical data. From these, methods to elicit dependence followed given the need of accounting for relationships between uncertainties. [82] discusses the historical development of expert judgement in uncertainty analysis and its achievements in more detail.

This development is not surprising as univariate quantities are (typically) more intuitive to experts and their specification is required (at least implicitly) prior to eliciting dependent distributions for two or more uncertain quantities.
In this section we discuss some main foci of structured expert judgement studies and evaluate the extent to which findings for univariate quantities are generalisable in the multivariate case. This discussion outlines where in a process adjustments are necessary when eliciting multivariate uncertainty and therefore provides an indication for the scope of dependence elicitation. Given the overall focus of the chapter, we outline only the relevant considerations for the elicited dependence parameters and the aggregation of judgements. However, it should be noted that an elicitation process is much more complex and other decisions in it, such as how to design the statistical training for experts prior to an elicitation, might vary as well considerably when eliciting multivariate uncertainty.
Already the earliest expert judgement studies for univariate quantities have shown that assessment outcomes can differ greatly depending on the use of directly or indirectly elicited query formats [375]. As a result, an extensive literature on heuristics and biases is available on the matter of framing elicitation questions and choosing a form for the query variable. Further, recommendations are made on the theoretical suitability of the elicited format, e.g. objections are made to non-observable quantities [214]. For eliciting multivariate quantities on the other hand, the same conclusions are not readily applicable. As will be seen, the effect of direct and indirect elicitation approaches is less well-understood and findings are often conflicting. The objection to non-observable quantities is less clear and indeed we show later that eliciting non-observable quantities performs well in terms of empirical accuracy and mathematical coherence. Similarly, for heuristics and biases only some extensions for the multivariate case exist, such as *illusory correlation* [66], stemming from the availability bias, and *confusion of the inverse*, originating with the representativeness bias [306] (for both see Appendix B). While these findings indicate an overlap for the existence of common biases, a lack of empirical research on the effect of framing for multivariate elicitation does not allow for generalisable conclusions.
Once the dependence information has been elicited in the form of some dependence parameter (which is thoroughly addressed in the following sections), a well-researched topic for univariate uncertainty, which generalisation would be desirable for multivariate elicitation, is the use of scoring rules. Roughly, a scoring rule is a numerical evaluation of a probability assessment based on observations. In expert judgement studies, they are typically used for two reasons, first to present an incentive for truthful assessment and second to measure the quality of an assessment after the elicitation, usually to inform a weighted combination of the judgements. In other words, they are used to define desirable properties of the assessment itself and they serve as a reward structure when evaluating an assessment. While an incentive is given by using (strictly) proper scoring rules which ensure that experts achieve their maximal expected score if and only if stating their true belief, a main property of measuring the quality of an assessment is its calibration, i.e. the statistical accuracy after observing an event of interest. Suppose an expert provides a probability distribution P over a set of n mutually exclusive events i. Then, after observing the events of interest, we can construct the sample distribution S with $S(i)$ equal to the number of times that i is observed divided by n. While it appears reasonable to state at first thought that an expert is not well calibrated if $S = P$, this might be false if we suppose that true values represent independent samples from a

random variable with distribution P. In this case, P relates to "reality" but we will never have $S = P$ due to statistical fluctuations. Loosely, an expert is therefore said to be well-calibrated if the true values of the uncertain quantities can be regarded as independent samples of a random variable with distribution P [79].

When evaluating experts' performance, we have to distinguish between scoring rules for individual variables and scoring rules based on sets of assessments together with sets of realisations. The first, assigning scores to each individual assessment and summing these scores over a set of variables, is often suggested in the literature for the purpose of rewarding, yet it is not a sensible approach. A main issue is that the resulting scores cannot be interpreted in a meaningful way without knowing the number of quantities assessed and their overall sample distribution. This is due to the possible additive decomposition of these types of scores into a "calibration" and "resolution" term [111]. Resolution measures how well experts partition the variables into statistically distinct categories while not considering whether the distributions assigned to these categories correspond to the experts' assessment. This becomes problematic when high resolution overpowers low statistical accuracy. A more detailed presentation of this drawback and some intuitive examples are given in [79, 83]). Therefore, scoring rules for average probabilities are highly encouraged for evaluating and combining experts. While some main properties of scoring rules are applicable in the multivariate case, others cannot be readily used.

[213] discuss (for the univariate case) the inclusion of order information (requiring an ordered state space). Ordered events allow for rewarding that takes account of nearness to an event's realisation. In the multivariate case the lack of natural ordering means that this approach is not possible. Further, [212] discuss a wide class of scoring rules, called generalised divergence scores, that allow for any baseline distribution (rather than a uniform by default), and which reward according to a measure of distance between the assessed distribution and the baseline distribution. Of interest for multivariate elicitation is the derivation of a weighted scoring rule that is closely related to the Hellinger distance which is a measure of divergence that has been used in the calibration of experts' multivariate assessments.

3.3 Guide to modelling and elicitation context

The main purpose of eliciting dependence is to quantify a multivariate stochastic model when this cannot be done wholly by conventional statistical estimation (which, in our view is a common situation). This section discusses broad approaches to dependence modelling in order to provide a clear structure for the next sections by highlighting the link between dependence modelling and expert judgement. Figure 3.1 shows this general view on the modelling context with three different broad approaches to assessing dependence and illustrates the relationships between model input and output variables.

In this general context, S represents the vector of stochastic variables in the model, and T the vector of output variables which depends deterministically on S. R represents another set of auxiliary variables used to evaluate the uncertainty on S. The solid arrows show deterministic relationships between the variables, and hence the direction in which uncertainty can be propagated.

It is not uncommon for there to be dependence between the output variables T. This can arise simply as a result of the functional dependence represented in

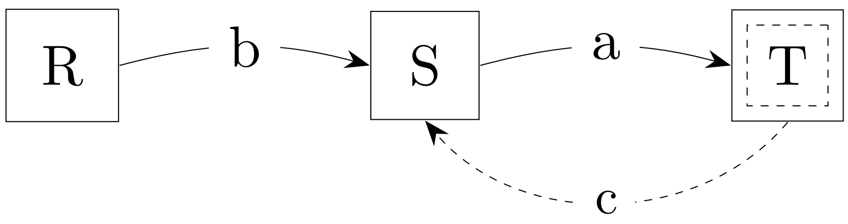


Figure 3.1: Schematic representation of modelling and elicitation context

arrow a, even if the stochastic variables in S are modelled as being stochastically independent. In many applications, however, it is not appropriate to model the variables in S as independent, and so we should find a way to model and assess dependence in S.

Approach a In Approach (a) we model the dependence relations between the variables in S directly. The main techniques are BNs, copulas, parametric families of multivariate distributions (e.g. the multivariate Gaussian distribution), and Bayes Linear methods. We provide examples for each method in the next section. Having assessed the dependence and hence having specified the distribution of the variables in S, uncertainty is then propagated through the model (arrow a) to the output variable (or variables) T. As we shall see later, direct assessment of dependence on the variables S is most predominant in the literature. However, two other approaches are also important and worth discussing.

Approach b In Approach (b) we introduce a new set of auxiliary variables R, which are not directly part of the model variables (though may in practice have some overlap with the variables S). The variables R are chosen so that their uncertainty is easier to quantify - in particular one might choose these variables so that they can be considered stochastically independent, with the dependence in the variables S arising as a result of the complex relationship between the "explanatory" variables R and those in S. This is shown in Figure 3.1 as arrow b. This approach is of interest particularly when change of variables methods (frequently used in multivariate statistics) can be used to simplify the variable set from S. A common model type used in this context is a regression model and an example of introducing and assessing auxiliary variables is given in Section 3.4.2.

Approach c In Approach (c) we "calibrate" the uncertainties on S through considering some set of output variables T on which the uncertainties can be assessed. Obviously, to be useful, this would have to be a different situation than the one in which the overall model is to be used (see dashed node inside T), as we would otherwise be simply directly assessing the uncertainty in the variables of interest. This calibration of uncertainties relies on the backwards propagation of uncertainty from T back to S, shown by arrow c. The dotted arrow is used to indicate a key difference with the solid arrows a and b. In general, more than one distribution on S will forward-propagate to the given distribution on T, that is, the inverse problem has no unique solution (or even worse, it has no solution). Other criteria (such as max entropy) are then used to select a unique inverse. That solution then defines a dependence structure on S, which can be propagated back through arrow a to look at other output contexts. This is called

Probabilistic Inversion (PI) [80, 239, 244] and we show an example in Section 3.4.3.
This approach is of interest when the dependence structure in S is difficult to determine directly, but must satisfy reasonable conditions on output variables that are easier to understand and hence easier to quantify.

A common theme in the latter two approaches is the model boundary. In both cases we choose to extend the model to include other input or output variables in addition to those which are strictly necessary for direct modelling. Indeed it may happen that the auxiliary variables represent simplifications of more complex issues which are insufficiently understood to be included explicitly in the model but which are known to collectively impact the behaviour of the system significantly. An example of this is the modelling of common cause events in risk analysis [84] where the range of underlying causes is too wide to be modelled individually, but which together have a substantial effect in inducing dependencies in the overall system behaviour.
We illustrate the dependence structures shown in Figure 3.1 with the following simplified project risk management example which shows how choices can be made in the various modelling contexts. We are managing a project which has an overall cost (model output variable T). The cost is determined by individual activities with associated costs (variables in S) that are of importance for the project completion. If we want to model the stochastic dependence between activities in order to obtain information about the overall cost, a first option is to do so directly by specifying the dependencies directly between the cost elements. The dependence models used here are part of modelling context a. If modelling the dependence between the individual activities directly does not produce a satisfactory model output, we have the choice to include explanatory variables (R) that help us to understand the relationship better. For instance, we can include factors like environmental uncertainties if we belief that our project's activity costs are (partly) influenced by them. The techniques used here are part of modelling context b. Recall that we are choosing to extend the model which relates to the earlier discussion on the model boundary. The reason for modelling dependency in this way is because it may be easier to consider the impact of certain factors explicitly rather than implicitly when only using approach a. If the model output resulting from the inclusion of additional factors is still not satisfactory, we might choose to model some systemic impacts of the project. For instance, factors like the availability of qualified staff might be present and result in a subtle dependence relationship, leading to the distribution for the overall cost (the model output variables T) being incorrectly assessed. With methods used in c, we would have a separate assessment of the distribution (or at least for features of this distribution) for the overall cost which would lead to a changed model for the joint distribution of the activity costs (modelling context a or b). We could also consider modelling a more complex situation in which we manage several projects. In this case, the overall cost becomes multivariate instead of univariate (i.e. T becomes a vector of variables). Then, we can use methods (from c) that allow propagating our uncertainty from one project about which we have information backwards in order to make inference about the distribution of the activities (S) and hence the distribution for overall costs (T). The common objective is to find a good model for the uncertainties relating S and T. Conceptually, we can only ever specify part of the required information for this model, so that in practice our model is always underspecified

(though this point is often not appreciated because modellers often adopt low-dimensional parametric families of models early on). Approaches b and c provide complementary approaches to specify further information about the model.

3.4 Dependence models and expert judgement

Before presenting and reviewing dependence parameters as elicited quantities explicitly, in this section we first discuss expert judgement for common dependence models. This includes main challenges when using experts to quantify models as well as the applicability of elicited forms for a satisfactory representation of the experts' information in the model. We present the modelling aspects first given that decisions here precede and strongly affect the choice of which dependence parameter to elicit. In accordance with the earlier framework (see Figure 3.1), BNs and copulas together with probabilistic and non-probabilistic parametric models are introduced for context (a), regression models for (b) and Probabilistic Inversion for (c).

3.4.1 Elicitation for direct modelling

Bayesian (belief) networks

In (a), a common way of integrating high dimensional uncertainty in a probabilistic model is by specifying a multivariate distribution for the random variables through the product of marginal and conditional probabilities. A common modelling framework is a BN [105, 318]. A random variable is described by a node in the graph while arcs represent the qualitative dependence relationships amongst variables. The direct predecessors/successors of a node are called parents/children, and the BN is specified (for example) by determining for every child node its conditional probability distribution given the states of its parent nodes. Hence, it is composed of a directed acyclic graph with marginal distributions for source nodes and conditional distributions for child nodes given the parents. A simple example BN to be used throughout this review is shown in Figure 3.2.

When using expert judgement, [155] views eliciting BNs as an obvious approach for obtaining dependence information. However, while more has been written about eliciting the qualitative dependence structure (the arrows in the BN) [197, 301], eliciting dependence quantitatively has been recognised as a main issue when constructing BNs [119, 339]. Identified difficulties are the elicitation for high dimensional models and the assessment burden due to an exponentially growing number of probabilities to assess (in discrete BNs). Therefore, some alternative modelling approaches have been developed to be used in conjunction

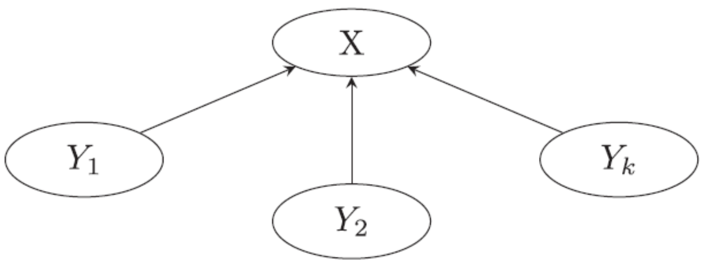


Figure 3.2: Example Bayesian network with one child and three independent parents.

with expert judgement methods.
While in the low dimensional, discrete case, experts provide information in form of conditional probabilities to populate conditional probability tables, in higher dimensions this is intractable and too time-consuming. An alternative approach is to model continuous distributions and to elicit dependence information through (un-)conditional rank correlations. These models are known as non-parametric BNs for which a review of applications can be found in [188]. For these, [290] developed a way of eliciting conditional exceedance probabilities for higher dimensions to derive the required rank correlations. This method is detailed in the next section when discussing elicited forms of dependence parameters explicitly.
In order to address the reduction of the assessment burden (in the discrete case), one way is to reduce the number of necessary assessments. For instance, [426] propose piecewise linear interpolation (see Appendix B) in order to reduce the overall number of required assessments for a full conditional probability table. Their method elicits conditional probabilities which are discussed in the next section as an elicited form. Another method that reduces the number of required assessments is through assumptions on the causal interpretation of a BN. The assumptions on the causal interpretation originate with noisy-OR gates [317] which use an underlying parametric distribution that reduces necessary assessments logarithmically (see Appendix B). The functional OR relationship denotes how individual parent nodes are combined for a common effect and assumes that they are independent of each other with respect to their causal effect on the child nodes. Thus, the presence of one parent node suffices to produce an effect on the child independently of other parents (with a certain probability - hence noisy rather than deterministic). A leaky noisy-OR gate includes a background probability that represents the influence of non-modelled causes. From this, [436] , building onto [120], introduce the elicitation of leaky and non-leaky noisy-OR parameters as alternatives to conditional probabilities. They use parameters introduced by [197] and [117] and a potential framing (for the BN in Figure 3.2) is:

> "What is the probability that X is present when Y_1 is present and all other causes of X (addition for leaky case: including those not modelled explicitly) are absent?"

In an experimental setting, [436] elicit leaky and non-leaky noisy-OR parameters together with conditional probabilities. An artificial dependence relation between three parents and one child node was determined (causes for anti- gravity of an unknown type of rock) and in a small simulation, participants could choose the influence (strength level of presence or absence) of each cause and observe what happens as an effect (anti-gravity or not). Then they assessed the conditional probability distribution with each assessment method, i.e. non-leaky and leaky noisy-OR parameters and a direct conditional probability assessment. The leaky noisy-OR parameter was assessed as most accurate (in terms of Euclidean distance to empirical distribution) while conditional probability was found least accurate. The authors claim that with an increasing number of nodes their method offers a clear advantage over conditional probability elicitation as the latter will become unmanageable. More generally, noisy-OR methods belong to the group of canonical models [317]. For these, assumptions on the underlying probabilistic relationship are made so that a conditional probability table can be generated algorithmically given parameters that are assessed by experts

and which only grow linearly with the number of parent nodes. Usually the parameters refer to conditional assessments which are made about a number of combinations of the states of the parent nodes. An alternative to the aforementioned noisy-OR method is the noisy-MAX method [117]. Within the same group of methods is also the ranked nodes approach [145]. Briefly, ranked nodes are random variables with discretised ordinal scales which are typically assessed by experts through verbal descriptors of the scale.
The usage of verbal classifiers to assess BNs has also been proposed more generally to counteract a high assessment burden. Here, the influence of a node is simply determined verbally rather than numerically. For instance, [395] use a scale containing both, numerical and verbal anchors, and [280] conclude (in a review on the use of expert judgement for BNs in human reliability assessment) that the use of verbal labelling for inferences in BNs is common. We discuss verbal elicitation of dependence explicitly in the next section.
A last way to facilitate judgement is by providing graphical support. [191] provide experts with the pie chart probability tool available in GeNIe Bayesian Network Software to adjust assessments. Probability masses are determined and the resulting distribution is graphically visible immediately. This procedure is repeated until the experts feel comfortable with their assessments.
As shown in Section 3.6, the use of expert judgement for BNs is considered in a variety of empirical areas given the popularity of this dependence model itself.

Copulas

In certain situations of context (a), a multivariate distribution can also be modelled by a copula rather than by the "marginal-and-conditional approach" [74], presented for BNs before. While an extensive introduction to copulas can be found in [125] and [208], recall first that for a continuous random variable X with distribution function F_X, the random variable $U = F_X(X)$ is uniformly distributed. If we have two continuous random variables X and Y, then the distribution of the vector $(F_X(X), F_Y(Y))$ is supported on the unit square and has uniform marginals. Any such distribution is called a (bivariate) copula. This construction can be reversed: Any set of univariate distribution functions combined with a copula represents a multivariate distribution as a result of [368]. The notion of a copula is easily extended to greater than two dimensions.
Often a one-parameter copula family is used, $C_\theta(u, v)$, that can be indexed by a parameter θ related to a rank correlation such as those of Spearman or Kendall (see Appendix B). In fact, both can be expressed in terms of the copula: Spearman's correlation is

$$\rho_C = 12 \iint_{[0,1]^2} C(u, v)\mathsf{d}u\mathsf{d}v - 3 \tag{3.4.1}$$

and Kendall's τ is

$$\tau_C = 4 \iint_{[0,1]^2} C(u, v)\mathsf{d}u\mathsf{d}v - 1 \tag{3.4.2}$$

Within a chosen family of copulas (see Appendix B), expert elicitation can be used to determine the correlation and hence specify the dependence. Whenever the family is uncertain, information on how copulas differ for upper or lower tail concentration, i.e. tail (in-)dependence (see Appendix B), needs be elicited additionally. For this, upper (or lower) asymptotic tail dependence is of interest. The asymptotic upper tail dependence parameter is defined as:

$$\lambda_U(X,Y) = \lim_{u \to 1^-} P(Y > F_Y^{-1}(u) | X > F_X^{-1}(u)) \tag{3.4.3}$$

when a limit $\lambda_U \in [0,1]$ exists. In this case, X and Y are defined as dependent in the upper tail when $\lambda_U > 0$, whereas whenever $\lambda_U = 0$, they are tail independent [208]. In other words, for the former case, it is more likely to observe high values for Y given high values for X. Following naturally from the concept of tail dependence, the tail concentration function distinguishes various copula formats and is defined for any u in $(0,1)$ as $\lambda_U = P(U > u, V > v)/(1u)$. For the (upper) tail, it leads to the tail dependence coefficient in form of $\lambda_U = (12u + C(u,u))/(1u)$.
The review results presented in Section 3.6 show limited experience for expert judgement within a copula modelling framework. One reason might be that copulas are distinguished on the one hand by measures of association such as rank correlations, but on the other hand also by its behaviour along the dependence function as indicated by its family. This constitutes a great deal of complexity to be integrated into an elicitation method. However, both types of information are highly important given that two different copula families exhibit a very different behaviour even for the same rank correlation (as shown in Appendix B). This is particularly crucial for copula families that model extreme joint dependence through asymptotic upper/lower tail dependence (as considered in the first elicitation approach presented below) in contrast to tail independent ones. At this point, it is important to note that the use and elicitation of measures of association related to tail dependence depends (obviously) on whether one is interested in capturing tail dependence explicitly or whether another measure might serve the modelling purpose better, given the increased cognitive complexity for experts to assess tail dependence.
Some proposed methods that aim at a sensible representation of an expert's understanding of dependence in form of a copula are outlined in the following. [17] decompose the asymptotic upper tail dependence coefficient (presented above) and query its components from experts before combining it again. They consider this as a non-asymptotic approximation of $\lambda_U(X,Y)$. The elicitation is as follows: in a first step, all non-negligible causes for X to be "extremely large" denoted as events j , $j = 1, 2, \ldots, J$, are listed. Then, experts assess $P(\textsf{event j}|X = \text{“}\textit{extremely large}\text{”})$, so the likelihood that the chosen event is present if X is in the tail of its distribution. Lastly, experts are queried $P(Y = \text{“}\textit{extremely large}\text{”}|\textsf{event j})$, i.e. the probability that the corresponding event affects Y with the implied magnitude. All assessments are then combined by $\lambda_U(X,Y) \approx \sum_{j=1}^{J} P(Y = \text{“}\textit{extremely large}\text{”}|\textsf{event j})P(\textsf{event j}|X = \text{“}\textit{extremely large}\text{”})$. The proposed framing is:

> "Given that an extremely bad outcome is observed in X, what is your estimate of the probability that Y will experience an extremely bad outcome?"

According to the authors (whose experts were actuaries) this method was perceived as cognitively easy.
Another option that is being researched further by several co- authors of this review but has not been published so far is querying conditional exceedance probabilities for chosen quantiles from experts to fit a parametric copula. This is done by plotting elicited values for each considered quantile together with candidate

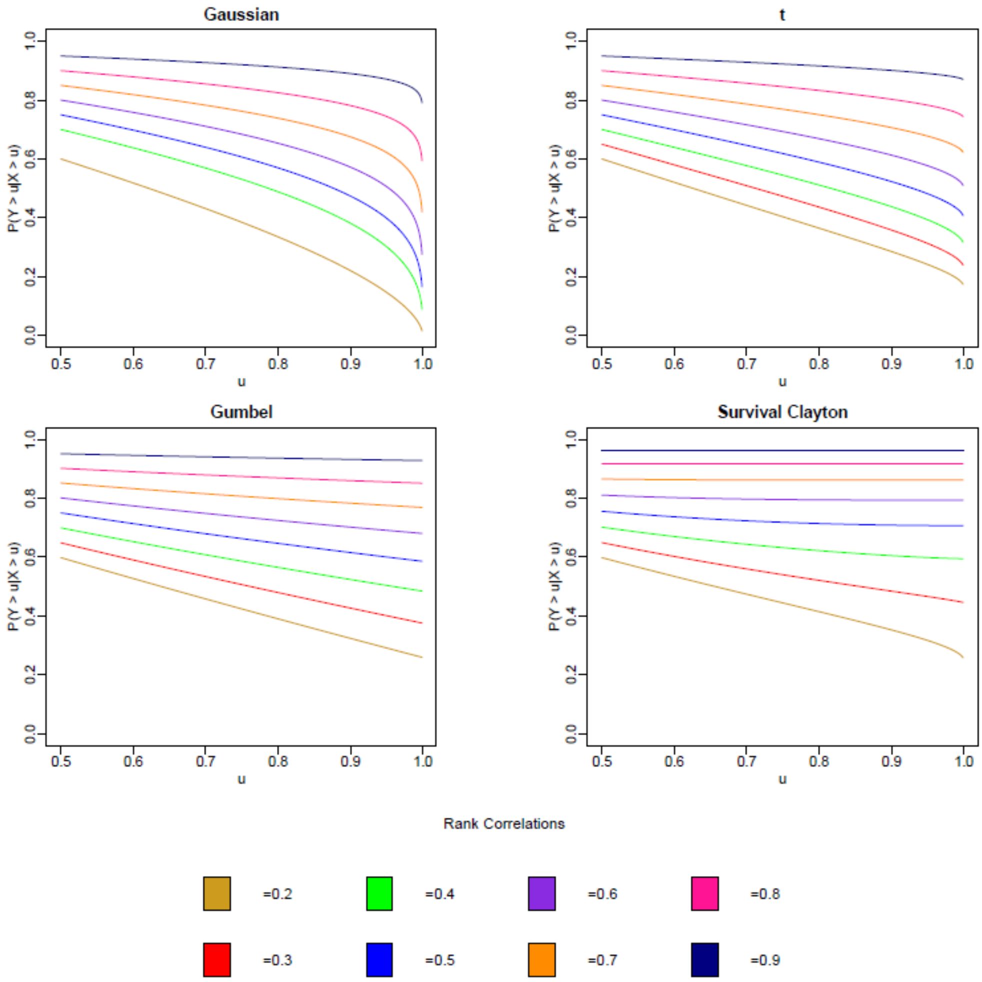


Figure 3.3: Conditional exceedance probabilities at u^{th} quantiles (rank correlations: 0.2 to 0.9).

copula choices and after a first "eyeballing" use conventional goodness-of-fit tests for the distance to parametric families. Figure 3.3 shows simulated conditional exceedance probabilities for several parametric copulas with given rank correlations. With the assessment of the probability that Y exceeds its u^{th} quantile given that X exceeds its u^{th} quantile for a certain number of thresholds u, a sensible copula choice that represents the experts' beliefs can be estimated. We address the details of eliciting conditional exceedance probabilities in the next section.

As a non-standard parametric alternative, [273] discuss using a minimally informative copula with given rank correlation. A copula is modelled by asking experts to provide a dependence constraint between two random variables, and taking the copula which is minimally informative with respect to the uniform (independent) copula. This is further developed in [32] and [30]. Here, experts assess the expectation of functions for the two underlying variables. From that a (min inf) joint probability is constructed which satisfies the expected value constraint. An advantage is that in this approach it is easier to relate a copula parameter to an observable quantity than it is for common parametric families. An example is given for the dependence of failure times between machine components. Minimal informativeness also served as motivation for [236] who

consider a sub-family of generalised diagonal band (DB) copulas which require a dependence parameter. It is specified by experts through conditional exceedance probabilities (given the median value). [397] regards DB copulas as advantageous when using expert judgement as a dependence parameter that relates to its one copula parameter can be defined. We will introduce this dependence parameter in the next section when we address forms of elicited dependence parameters explicitly.
Besides some empirical work in maintenance optimisation [60], the majority of experiences for eliciting copulas, such as the first approach presented above, comes from banking and insurance [17, 47, 336, 361], an area in which the popularity of copulas has increased lately [170]. Here, expert judgement is typically used to assess conditional and joint probabilities of (extreme) loss events. These studies might be helpful for other areas where copulas are gaining increased interest, such as hydrology [169].

(Probabilistic) parametric models: multivariate distributions

Another way to model dependence in (a) is by specifying a multivariate distribution. For an introduction and overview of the distributions discussed here, see [26].
As a main challenge when eliciting a multivariate distribution is that its full specification would be cognitively too complex for experts, we should impose a structure on the distributional choice. While for univariate distributions it might be sufficient to assume a minimal structure such as a continuous and smooth cumulative distribution function which can be specified satisfactorily by a few quantile assessments [306], in higher dimensions this is still unreasonable for practical use. Rather, a parametric multivariate distribution that represents an expert's belief sufficiently is a necessary assumption. Then, an expert's opinion is fully specified by determining a few parameters. While any distributional assumptions have to be in agreement with the experts, they should be as well in accordance with the modelling purpose. For instance, it should be suitable for its use in a specific decision problem for which a distributional form is predetermined or its use as a prior in a Bayesian modelling framework. The latter offers a probabilistic framework to complement the lack of data for some common statistical dependence models. Prior beliefs of experts (see Appendix B) for given parameters are updated once observations are available. A prior is chosen so that it can be most easily updated [306]. Generally, this is a different elicitation situation/purpose than using expert judgements to obtain beliefs about uncertainties without the inclusion of future observations (what is done in most of the literature reviewed here), but this is not of importance for us as with regards to dependence elicitation both methodologies have similar challenges. Hence, both methodologies contribute to the findings presented here.
In the literature on eliciting parameter information for quantifying a multivariate distribution, mainly multivariate normal [9, 10, 115, 162], or t [10, 215] and Dirichlet distributions [65, 135, 437] are considered. A method that specifies a multivariate distribution in a more flexible way (as shown below) is given in [281].
For the common parametric assumption of a multivariate normal or t distribution, the elicitation aims at quantifying the mean vector, μ, and the covariance matrix, Σ. Instead of determining the variables of interest directly, even though this has been attempted through interactive graphical methods [64], typically *hyperparameters* that follow from distributional assumptions on the form of μ

and Σ therefore specify (or index) the multivariate distribution of interest are determined. In other words, the values of the hyperparameters reflect the available subjective prior knowledge about the unknown model parameters. This is typically based on specifying hierarchical priors assuming exchangeability (see Appendix B) for the joint distribution in question. The variables of interest are then conditionally independent given the hyperparameters. This is known as Bayesian hierarchical modelling (see Appendix B) which is a common way to restructure dependence in order to elicit parameters as univariate quantities. Typically, the hyperparameters consist of means, scale parameters, degrees of freedom and the spread matrix which (whenever possible) are elicited through univariate quantities and conditional medians of observable variables. [321, 322] presents how the specification of suitable prior distributions can be simplified and how values of hyperparameters can be elicited from experts through quantiles of predictive prior distributions for a variety of common distributions in the reliability context of mathematical modelling of maintenance. While we explain this approach below (for Dirichlet distributions), it is noteworthy here that a main advantage is that observable quantities can be used. Further, he proposes to elicit fewer quantiles than unknown hyperparameters and use interaction of experts for further adjustments.

A different problem for which a multivariate distribution needs to be specified is whenever an event can take one of k possible outcomes ($k > 2$) and the probability of the i^{th} outcome, p_i, is elicited from experts. This might be denoted as eliciting the opinion about a "set of proportions" [437]. As the sum of all p_i must equal 1, p_i cannot be assessed in isolation. Further, with $k > 2$, a multinomial distribution models the overall outcome given that we have independent trials and the probability of each outcome is the same in each trial. The commonly chosen parametric distribution is then a Dirichlet distribution, the conjugate prior distribution of a multinomial one [306]. One of the earliest approaches in [65] uses an elicitation strategy based on predictive distributions. When considering a specified number of draws from the population of interest, the expectation of the number that belongs to a category is in fact p_i. Given that, they ask their experts for the joint modes of the predictive distribution. Other methods assess the Dirichlet distribution by imaginary observations, i.e. by determining the extent to which experts change their belief given an observation from a draw [306]. More recently, [437] proposed a refinement to acknowledge the strong assumptions of a Dirichlet distribution (due to the small number of parameters that determine its form) and therefore make use of over-fitting. Loosely, they ask experts for more assessments than (strictly) necessary to fit a distribution in order to reject the choice of a Dirichlet distribution if it is inappropriate.

A more flexible method that avoids experts' belief to fit a single pre-specified parametric family is presented in [281]. While the focus of the elicitation is laid on the analyst who seeks to identify the probability density function for a multivariate vector, the posterior distribution is based on the prior distribution as specified by an expert. In order to ensure flexibility on the parametric assumptions, the analyst's prior belief is a Gaussian process which allows the multivariate distribution to take a variety of forms given the experts' assessments. The elicited parameters are univariate quantities and a small number of joint probabilities, unless the elicitation of the latter can be reduced to querying univariate information as well, depending on assumptions for the multivariate vector's probability space.

Given that dependence information for quantifying parametric multivariate distributions is (mainly) elicited through univariate quantities, experimental studies show a similar performance to expert judgement studies with univariate variables of interest. For instance, (conditional) medians are regarded as cognitively easy and reliable to assess [11]. Empirical findings on the elicitation of multivariate distributions are scarce however which is why no indication for a particular application area can be given (Section 3.6).

(Non-probabilistic) parametric models: Bayes linear methods

An alternative to eliciting distributional (prior) beliefs for Bayesian models in (a) is the Bayes linear method (BLM) [173]. It differs by using expectation as basis and is able to represent more complex problems through adjusting beliefs by linear fitting. Without distributional assumptions all required parameters are first and second moments [142]. Hence, eliciting dependence information concerns beliefs about the covariance of parameters (rather than joint probabilities). While not much experience on the actual elicitation is found in the literature, [341, 342] and [340] address expert judgement for BLM specifically. The dependence model considered is $Y = \alpha X + R$ where X is the explanatory variable of Y. R represents the unexplained uncertainty between X and Y (with no correlation between X and R) and α is used to measure the strength of the relationship between X and Y. As a pragmatic way to elicit covariance information, the elicitation of quantiles is proposed whereas the relation between these and the moments needs to be derived. A possibility is through [320], further developed in [225], who propose eliciting from three to five percentiles to obtain means and variances. Hence, with the 5^{th}, 50^{th} and 95^{th} quantiles specified as $x_{0.05}$, $x_{0.5}$ and $x_{0.95}$ for an uncertain variable X, the mean is derived by $\mu_X = 0.63x_{0.5}+0.185[x_{0.05}+x_{0.95}]$ and the variance by $\sigma_X^2 = ((x_{0.95}-x_{0.05})/(3.29-0.1(\Delta/\sigma_0))^2$ with $\Delta = x_{0.95}+x_{0.05}-2x_{0.5}$ and $\sigma_0 = ((x_{0.95} - x_{0.05})/3.25)^2$.
In [341] five elicitation techniques are compared. A first one is the direct elicitation of cross-moments which is omitted here given that it is discussed in the next section as a commonly elicited form. For the remaining methods we assume that the mean and variance of X and Y have been elicited beforehand. In the direct calculation approach, experts assess their updated belief of $E(Y)$ after the observation that $E(X)$ increased hypothetically. While α can be computed, for the uncertain variable R the experts' 5^{th}, 50^{th} and 95^{th} quantiles are elicited through:

> "Given that X is known to be $\bar{x}$ with complete certainty, what are the 5^{th}, 50^{th} and 95^{th} quantiles of Y?"

It follows that $E(R)$ and $var(R)$ can be calculated as shown before and then $E(Y) = \alpha E(X)+E(R)$, $var(Y) = \alpha^2 var(X)+var(R)$ and $cov(X,Y) = \alpha var(X)$. For the adjusted expectation method, experts are asked to re-assess their belief about X based on the true value of Y. When defining the true value as $\bar{y}$ the new belief for $E(X)$ is $E_Y(X) = X_Y$ with observed $\bar{y}$. The covariance can then be calculated as $cov(X,Y) = ((E_Y(X)-E(X))/(Y-E(Y)))var(Y)$. The value of α is again computed and defines the values an expert can assess for coherence reasons. The adjusted uncertainty approach works in the same way as adjusted expectation, with the only difference that the variance of X is updated based on an observation of the true Y. With the adjusted variance denoted as $var_Y(X)$, the adjusted covariance is then derived using $cov(X,Y) = \sqrt{(var(X) - var_Y(X))var(Y)}$. In an experimental setting of the same study, experts were presented with the pairs of variables for life expectancy between males and females (in the same

country), height and weight of male students, as well as mean time to failure between vehicles. All experts were familiar with basic statistical summaries, but not with BLM. The different techniques were compared for accuracy, incoherence and intuitiveness. Thereby, adjusted uncertainty was the only method that exhibited incoherent assessments and also had more inaccurate results with far more assessments of negative or no correlation when all empirical data was positively correlated. Direct calculation on the other hand had the best performance in terms of accuracy and no incoherent assessments. Direct correlation and adjusted expectation barely showed any differences for experts' performance. However, over 15 % of the responses were deemed inconsistent.
While this is the first and only such complete attempt to explicitly focus on the actual elicitation of covariance in BLM, some main references for empirical studies with documented expert judgement approaches are [174], [342], [33], [143] and [307].

3.4.2 Elicitation for indirect modelling with auxiliary variables

Regression models

A common dependence model in context (*b*) is a regression model. For recent overviews, see [353] and [409].
Recall that here information on the dependence is modelled indirectly by restructuring the natural input. Technically restructuring is done using variable transformation techniques. Beliefs about parameters are then elicited while being formulated as univariate query variables. Similar to quantifying parametric multivariate distributions, elicitation here is typically done for prior beliefs in a Bayesian methodology.
The parameter of interest is a regression coefficient, β. The likelihood function $p(Y|X, \beta)$ relates observed data Y to regression coefficients β and covariates X. Experts then specify the prior distribution for $p(\beta)$ typically through hyperparameters which are the mean and the variance of the regression coefficient [204]. Eliciting moments of regression coefficients directly however might be cognitively too complex given that experts would need to understand the effect that a change of covariate X has on Y. Therefore, the literature on eliciting priors for regression models proposes indirect approaches. For these, experts provide a probability of the response value based on specified values of the explanatory variables or vice versa. From this, prior elicitation methods for linear models, normal [215] and multiple [164], piecewise-linear [163] as well as logistic regression models [309] have been developed. For the latter, experts typically assess conditional means, $E(Y|X, \beta)$ [36, 204] for a probability of presence, p_i, with binary responses for observation i modelled as $\mathsf{logit}(p_i) = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \cdots + \beta_j x_{i,j} + \eta_i$ [309]. For instance, [70] elicit the probability of presence for a certain wallaby type at a specified location with fixed habitat characteristics in habitat modelling. Depending on distributional assumptions for the probability of presence (such as a Beta distribution) the mode rather than an arithmetic average or median might be elicited due to the potential skewness of the distribution.
In a similar manner, parameters can be elicited for (multiple) linear regression models. [164] propose a model of the form:

$$E(Y|x_1, x_2, \ldots, x_i) = (\beta_1 x_1, \beta_2 x_2, \ldots, \beta_i x_i) \tag{3.4.4}$$

where again β denotes the regression coefficient and $E(Y|x_1, x_2, \ldots, x_i)$ is the expected (average) value of Y when $X_1 = x_1, X_2 = x_2, \ldots, X_i = x_i$. Experts

then specify the prior distribution of β by assessing hyperparameters. To do so, the authors introduce design points, values at which a prediction is made after hypothetical data are given. Likewise, [215] elicit fractiles for a predictive distribution with specified values at design points, using a bisection method (see Appendix B).
Regression elicitation is further explored in [70, 309] and [11]. [309] present three different elicitation methods with graphical support, similarly to [11] who use an interactive graphics method as well. Empirical studies for expert judgement in regression modelling are mainly found in the area of ecology for which e.g. [70] summarise various approaches.

3.4.3 Elicitation for modelling propagation of output

Probabilistic inversion

In modelling context (c), a common situation is that input parameters of a dependence model are not observable. Therefore, a direct quantification of these variables is not sensible and methods such as PI [80, 244] are used. Its aim is to take the distribution representing the uncertainty on certain observables and translate it on the uncertainty of target variables. While the distribution can come from historical data, PI can be used as well as a method for transforming expert assessments of some observable model outputs into uncertainties on parameter values. A motivation for PI (that was never published as such) originated in the development of expert judgement methods and uncertainty analysis in the nuclear sector (for a historical overview, see [82, 240]) where experts refused to assess transfer coefficients directly. Similarly, [239] elicit outputs of a power law that models spread of lateral plume in atmospheric dispersion in form of $\sigma_y(x) = A_y x^{B_y}$. The output $\sigma_y(x)$ denotes the lateral (indicated as y) spread at wind-speeds x and is determined by the dispersion coefficients A and B. Instead of querying the joint distribution on (A, B), which would require experts to consider all possible effects of this relationship through the model, they are asked to quantify uncertainty on the output at various downwind distances through a univariate elicitation method. In addition to modelling plume spread, the same paper discusses a case study in banking. Empirical findings of the method are however lacking which is why no indication of specific application areas can be given.

3.5 Forms of elicited dependence parameters

This section reviews the proposed forms of dependence parameters for elicitation, i.e. association measures or summary types of an expert's joint distribution that are used in an elicitation question. As well, the corresponding framing of elicitation questions is presented. In addition to outlining the main elicited forms, an evaluation regarding desirable properties is given whenever possible. Chosen desiderata allow for guidance on the suitability of elicited dependence parameters from different perspectives.

Desiderata for elicited dependence parameters

A first perspective concerns theoretical feasibility whereas a common desideratum for expert judgement is that the elicited forms are observable and physically measurable. This allows assessments to be credible and defensible [79]. With

a similar objective, a rigorous foundation in probability theory is desirable.
A further perspective considers the assessment burden for experts. In this regard [214] emphasise practicality, i.e. that experts feel comfortable at assessing uncertainty while their opinion is captured to a satisfactory degree. For the former, query variables should be kept intuitively understandable. For the latter, queried information should be linked as directly as possible to the specific dependence model of interest, ensuring that an expert's assessment is satisfactorily reflected in the final output of the model. As variables are often transformed into some other parameter than the one that populates a dependence model (e.g. due to a potential reduction in the assessment burden), it is important to measure and control the degree of resemblance between the resulting assessments (through the model) and the dependence information as specified by the expert [238]. Note that the transformation of dependence parameters is typically based on assumptions about the underlying bivariate distribution. For instance, when transforming a product moment correlation coefficient into a rank correlation, this is straightforward under the assumption of bivariate normality. However, positive definiteness is not guaranteed which relates to the next desideratum, that of mathematical coherence. Coherence means that the outcome should be within mathematically feasible bounds. For dependence measures, ensuring positive definiteness of a resulting correlation matrix might be a potential issue and methods that adjust experts' judgements might be necessary [262]. Yet, whether an expert agrees with this adjustment or not determines their confidence in the final assessment. Another solution to incoherence is to fix possible bounds for the assessment a priori, even though this can severely decrease the intuitiveness of the assessment. A last desideratum is to calibrate assessments on statistical accuracy. This means, we would like to test experts' performance (in terms of statistical accuracy) against empirical data (if available), often to inform the weighting for mathematically combining judgements.
While no elicited dependence parameter meets all desiderata, their consideration supports comparison and allows a better guidance in terms of suitability within certain modelling situations.
At a broad level, a distinction for elicited quantities can be made between probabilistic and statistical approaches [74, 238, 290]. Whenever possible the presented findings are categorised into one of the groups. Approaches that do not fit in any of these classifications can be found in Section 3.5.3.

3.5.1 Probabilistic methods

In the selected literature popular variables to elicit are of probabilistic nature. This popularity can be attributed to the firm foundation (in probability theory) and the (potential) observability of the elicited variables which accompany this choice.

Forms of probabilistic dependence parameters

Conditional (exceedance) probabilities In the context of probabilistic measures of dependence, *conditional probability* might be the best known one. A common way to elicit conditional probabilities is to provide an expert with the information that the conditioning variable is observed above (or below) its median value (marginal probabilities are elicited first or are known from data) before the probability that the target variable lies above (or below) its median value is enquired. A possible framing of the question is:

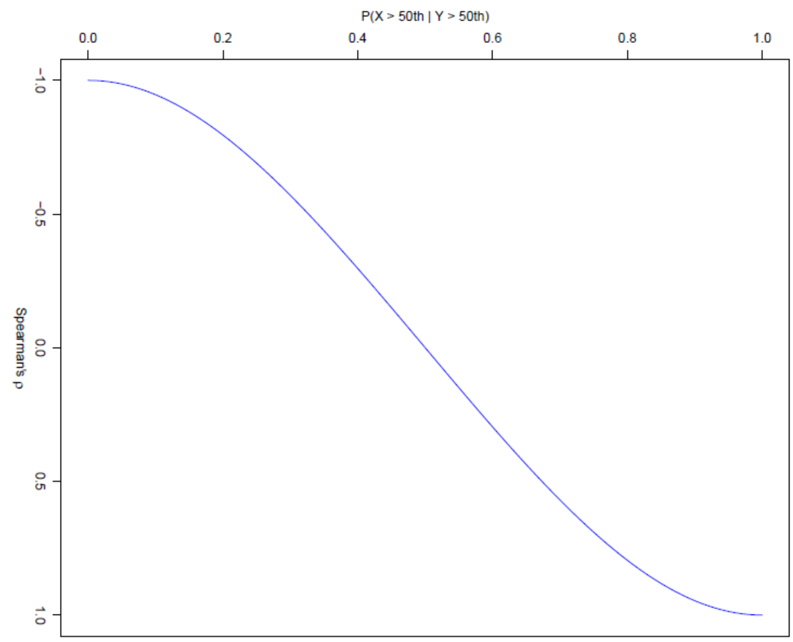


Figure 3.4: Expert's conditional probability assessment as a function of the product moment correlation coefficient.

> "Consider the pair of variables, X and Y. Suppose now that Y has been observed to be above its/your median value for it. What is the probability that X lies also above its/your median value for it?"

This might be extended to any quantile defining for the pair of random variables X and Y the elicited form for a conditional probability as $P_{CP}(x_i, y_i) \coloneqq P(X \geq x_i | Y \geq y_i)$ where $i = 0.5$ refers to the median value, but i might take any other quantile. Experts assess independence between X and Y as $P_{CP} = (x_i, y_i) = P(X \geq x_i)$ implying that learning about $P(Y \geq y_i)$ does not add any information. For a (strong) negative relationship experts state their belief as $P_{CP} \in [0, P(X \geq x_i))$ while for a (strong) positive it is $P_{CP} \in (P(X \geq x_i), 1]$. Given the above form, a conditional probability is sometimes also called conditional exceedance probability. In contrast, another way to elicit a conditional probability is by $P_{CP}(x_i, y_i) \coloneqq P(X \geq x_i | Y = y_i)$. This way can be applied similarly and its use depends strongly on context. However, [306] regard it as less cognitively complex. In order to transform a conditional probability into a product moment correlation coefficient (e.g. for modelling purposes) the relation between the two can be derived as shown in Figure 3.4.

The above derivation is possible only when an assumption about the underlying copula is made [244]. Figure 3.4 was obtained under the assumption of normal copula density for X and Y. The analyst finds the product moment correlation that ensures a positive definite correlation matrix and satisfies the expert's assessments [290].

Experts' performance when eliciting conditional probabilities (in comparison to six other methods) has been investigated in [73]. The assessed pairs of variables are relationships such as height-weight, as well as dependence between individual stocks, their indices and the relation between stocks and their indices. Participating experts were MBA students with some basic statistical training. In this experimental setting, conditional probability is among the worst performing methods for coherence and fourth out of six in terms of accuracy against empir-

ical data. Similar coherence issues when assessing conditional probabilities were observed by [299] who therefore provided their experts with a Joint Probability Table which led to large improvements in performance. Generally, for this method the elicitation of several values to condition on is recommended [89].
In the case-study literature (Section 3.6), the elicitation of conditional probabilities is nevertheless favoured as it often serves as direct model input. Main references where this approach has been formally used stem from the Joint CEC/US-NRC Uncertainty Analysis framework [88]. The experts participating in these studies became familiar with this format which underlines the importance of training experts to ensure familiarity.
An alteration to the elicitation of conditional probabilities which is also closely related to concordance probabilities (see below) is presented in [140]. Experts are asked to assess the median deviation concordance probability which is also known as quadrant probability [241]. It is defined as the probability of the two variables, X and Y, falling both either below or above their medians, i.e. $P_{QP}(x, y) := P((X - x_{0.5})(Y - y_{0.5}) > 0)$ with $x_{0.5}$ and $y_{0.5}$ being the respective medians. This could be asked for as follows:

> "Consider the pair of variables X and Y. You have indicated that there is a 50/50 chance of X being above or below $x_{0.5}$ and Y being above or below $y_{0.5}$. What is the probability that X and Y both will either be above or below their medians?"

The above formulation is a slightly altered version of the original reference to offer a general framing. While the conditional probability cannot be fully represented with a quadrant probability, the author claims that the dependence elicitation concentrates on events that experts "should be capable of making most informed judgements about" [140]. According to [241], this is "perhaps the simplest measure of association between two random variables" and an advantage is that it can be assessed and interpreted on the customary range. This measure is non-parametric, meaning that is has a well-defined interpretation (even) when structural assumptions, such as bivariate normality, do not hold. Further, it is ordinally invariant, i.e. it remains unchanged by monotone functional transformations of its coordinates. This has advantages with regards to modelling convenience as well as in terms of cognitive complexity to assess it. The measure is closely related to Blomqvist β [45] which is defined as $\beta = P((X - x_{0.5})(Y - y_{0.5}) > 0) - P((X - x_{0.5})(Y - y_{0.5} < 0)$.
Similar to [241] when discussing the conveniences of using the quadrant probability, [45] describes his measure of association as being "valid under rather weak assumptions regarding the distribution of the population" and "easy to deal with in practice". Under the assumption of bivariate normality, a relation to the correlation coefficient, ρ, is given by $(2/\pi arcsin\rho)$. Given the advantages from a modelling together with elicitation perspective and as pointed out by a reviewer of an earlier version of this chapter's published version, the quadrant probability and Blomqvist β deserve more attention when eliciting dependence.

Conditional (exceedance) probabilities (for higher dimensions) Eliciting higher dimensions of dependence such as in [289] and [290] requires the assessment of conditional rank correlations in addition to unconditional ones. To do so, the variables of interest that are conditioned onto are ordered according to some order of preference. This corresponds for instance to the relation of parent to child nodes in a directed acyclic graph. Once experts have assessed the unconditional rank correlation ρ_{X,Y_1} (in Figure 3.2) with any of the other techniques presented here, the conditional rank correlations need to be determined ($\rho_{X,Y_2|Y_1}$ and $\rho X, Y_k|Y_2, Y_1$ in Figure 3.2). A probabilistic way to do so is through conditional (exceedance) probabilities for higher dimensions which directly follow from the low dimensional case discussed above. A question (according to Figure 3.2) might be framed as follows:

> "Suppose that not only Y_1 but also Y_2 has been observed above its/your median value. What is now your probability that also X will be observed above its/your median value?"

For this the conditioning set of the unconditional case will be extended to $P_{CP}(x_i, y_{1,i}, y_{2,i}) := P(X \geq x_i|Y_1 \geq y_{1,i}, Y_2 \geq y_{2,i})$ for the i^{th} quantile, e.g. $i = 0.5$ for the median. If experts assess (conditional) independence, the estimate will be the same as for $P_{CP}(x, y_1) = P(X \geq x_i|Y_1 \geq y_{1,i})$. Otherwise the positive/negative relationship is assessed as before. Whenever $P_{CP}(x, y_1, y_2) \neq 1$ or 0 it follows that X is not completely explained by Y_1 so that Y_2 adds to the explanation of the former. In psychological research of causal learning theory, Y_1 , Y_2 and Y_k would be referred to as cues that compete for associative strength [279]. The idea of associative strength shows a key difference to the elicitation of noisy-OR parameters presented earlier in the context of BNs.
The intuitiveness of this method might be inhibited given that the choice of the first (unconditional) correlation imposes restrictions of the possible values for the second (conditional) correlation (similar to those of positive definiteness of a correlation matrix). This introduces the necessity to compute (in real time) updated intervals (different than the unrestricted $[-1, 1]$) into which the new assessment can fall, to preserve coherence. Technical details can be found in [287].
In order to test experts' performance when assessing a multidimensional dependence structure, [289] compared conditional probabilities of exceedance with the direct elicitation of pairwise correlation. In their study, a group of 14 experts (with previous training on statistics) was presented with two versions of a graphical model for the relationship between sulphur dioxide emissions and fine particular matter in Alabama, USA. The experts were split into two groups so that different dependence measures could be elicited. For the first model, querying the rank correlation directly exhibited the best performance when averaging out the absolute difference of empirical data and all individual answers. Based on a performance-based measure of accuracy (see next chapter), the top three most accurate experts assessed correlation directly. However, when averaging performances per elicitation technique and model, the conditional exceedance probabilities outperformed direct assessments. Nevertheless, the authors could not formulate definitive conclusions since the different model versions might have had an influence on the differences in experts' performances.

Joint probabilities From conditional probabilities it follows naturally to consider the elicitation of joint probabilities. A joint probability, $P_{JP}(x, y) := P(X \geq x, Y \geq y)$, can be queried for two random variables, X and Y, by asking:

> "Consider the pair of variables X and Y. What is the probability that both are within the lower (upper) k^{th} percentage of their respective distributions?"

If an expert assesses independence between X and Y, the joint probability corresponds to $P_{JP}(x, y) = F_X(x)F_Y(y)$, where F_X and F_Y represent the marginal

cumulative distributions of the corresponding elicitation variables. A positive relationship is assessed by either $P_{JP}(x, y) = F_X(x)$ or $P_J P(x, y) = F_Y(y)$. For a negative relationship $P_{JP}(x, y)$ approximates 0.
A relation to the (product moment) correlation coefficient is derived similarly as in the case of conditional probability. For medians, conditional probabilities are derived by using the relation $2P(X \geq x_{0.5}, Y \geq y_{0.5}) = P(X \geq x_{0.5}|Y \geq y_{0.5})$ [306]. [104] mention a modification to elicit joint probabilities. It is presented in [281], where the elicited probability takes the form $P_{JP}(x, y) \coloneqq P(x_i \leq X \leq x_j, y_i \leq Y \leq y_j)$. It is concluded that this alternative is able to capture the most important features of an expert's distribution with a good accuracy and by just making use of a small amount of data.
Eliciting joint probability directly however is seen as rather cognitively complex and (even) assessing independence in such a way is regarded as non-intuitive [165]. A systematic bias for this kind of assessment is that experts tend to overestimate the probability of conjunctive events and underestimate that of disjunctive ones [306]. This might be due to the requirement that certain knowledge of probability theory is necessary for this method. [73] found that when elicited joint probabilities are transformed to correlations, the obtained values tend to be out their feasible bounds rather frequently. Further, it was the least accurate method when compared to empirical data.

Concordance probabilities A further way to think probabilistically about dependence is by considering concordance (and discordance) of random variables. The concept of concordance probabilities is closely related to the earlier introduced quadrant probability and it is limited to a frequency or cross-sectional interpretation for the pair of variables in question, i.e. it requires a population to draw from [74]. The question can be framed as:

> "Consider two independent draws (x_a,y_a) from their common underlying population a and (x_b,y_b) from population b. Given that $x_a > y_a$ holds for population a, what is your probability that the relation $x_b > y_b$ holds for population b?"

Exemplary populations for a and b might be height and weight of some specified group of people. Formally, the probability of concordance between two random variables, X and Y, considering n independent draws (x_a,y_a) to (x_b,y_b) is given by:

$$P_C(x, y) = \frac{\sum_{\mathsf{a}=1}^{\mathsf{n}-1} \sum_{\mathsf{b}=\mathsf{a}+1}^{\mathsf{n}} 1_{\mathsf{C}*}((\mathsf{x_a}, \mathsf{y_a})(\mathsf{x_b}, \mathsf{y_b}))}{\binom{\mathsf{n}}{2}} \tag{3.5.1}$$

with $\mathsf{C}* = (\mathsf{x_a} - \mathsf{x_b})(\mathsf{y_a} - \mathsf{y_b}) > \mathsf{0}$. It can be assessed by an expert on $[0, 1]$. A value of (or close to) 0 indicates a strong negative relationship, 0.5 represents independence, and 1 refers to a strong positive relationship. The transformation to a rank correlation such as Kendall's tau, τ, is defined as $\tau = 2P_C 1$. With the assumption that X and Y can be approximated by a bivariate normal distribution, the relation from τ to other correlation measures, such as Pearson's product moment correlation, ρ^*, or Spearman's rank correlation, ρ, can be inferred through $\rho^* = sin(\pi\tau/2)$ and $\rho^* = 2sin(\pi\rho/6)$ [241]. Nevertheless, a (transformed) product moment correlation matrix that is positive definite is not guaranteed [238].
Within the psychological literature of causal learning, the concordance probability relates to the term *degree of relatedness*. In the classical experimental design, participants are presented with information about the presence or absence of

an input variable, representing a candidate cause, as well as the presence or absence of an effect/outcome. For instance, medical experts assess the likelihood of a disease from the (non-)occurrence of a symptom. Based on their assessments of discordant and concordant observations the aim is to formulate descriptive rules for inferring causal strength [359].
In [73], this technique performed reasonably accurate in comparison to other methods and only rarely incoherent assessments were made. Similarly, [165], [243] and [172] come to the conclusion that this method is reasonably accurate and might be preferred if a population is given. Yet the importance of an expert's familiarity with the population is emphasised.

Expected conditional quantiles (fractiles/percentiles) The quantile (fractile/percentile) method requires conditional estimates and therefore shares certain characteristics with eliciting conditional probabilities. Experts are presented with information that the conditional value corresponds to a certain quantile (or fractile/percentile) and given that information, the experts assess which expected quantile the other variable takes. A possible framing might be:

> "Consider variables X and Y. Given the value Y has been observed at its i^{th} quantile, q_i. What is your expectation of X's value in terms of its quantile?"

For the pair of random variables, X and Y, this is defined as $E(F_X(x)|Y = y(q_i))$ where $F_X(x)$ is the corresponding distribution function of X and $y(q_i)$ is the value that Y takes at its i^{th} quantile. The relation to rank correlation is given through the standard non-parametric regression function of $E(F_X(x)|Y = y(q_i)) = \rho_{X,Y}(F_Y(y)0.5) + 0.5$ (Figure 3.5 and 3.6). The conditional quantile is bounded by $\mu_{min} \leq E(F_X(x)|Y = y(q_i)) \leq \mu_{max}$ where $\mu_{min} = min[F_Y(y), 1F_Y(y)]$ and $\mu_{max} = max[F_Y(y), 1F_Y(y)]$. If $F_Y(y)$ is above its median, the values close to the minimum refer to a (strong) negative relationship, and the values close to the maximum indicate a (strong) positive one. For independence, experts assess $E(F_X(x)|Y = y(q_i)) = 0.5$. A closely related method is predictive assessment which was mentioned in the context of hyperparameters.
It should be noted that this dependence parameter has certain characteristics which would have similarly justified listing it among the statistical approaches which are presented in Section 3.5.2, after the general discussion on the assessment burden of probabilistic methods.

Assessment burden of probabilistic methods

Despite the limited empirical evidence available for experts' intuitive understanding of different assessment methods, [290] and [73] conclude that probabilistic statements are not perceived as cognitively easy. Conditional as well as joint probability assessments were rated by experts as most difficult among all other methods presented to them. In particular, when moving towards higher dimensions, the growing conditioning sets for conditional exceedance probabilities were met with accordingly growing concern. Additionally, for conditional quantiles (fractiles/percentiles) the expert must understand these location properties of distributions quite well together with the notion of regression towards the mean which might induce cognitive difficulties [74]. A possible advantage of these techniques is that the assessment burden can be decreased for most probabilistic methods by re-framing the questions. For instance, it is often possible to

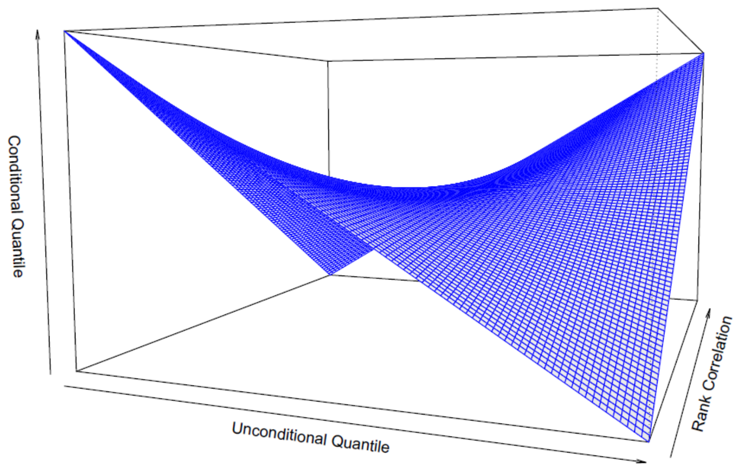


Figure 3.5: Conditional quantiles to rank correlations (perspective).

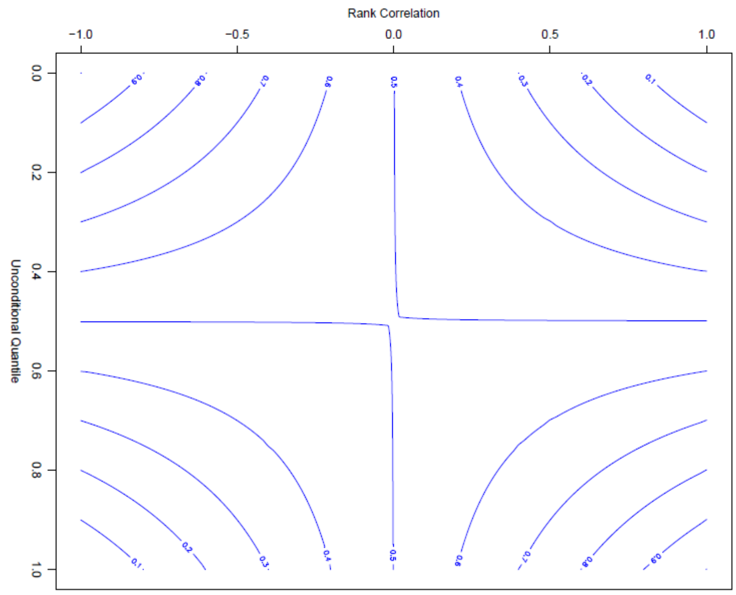


Figure 3.6: Conditional quantiles to rank correlations (contour).

express their forms as relative frequencies which are a more natural way of thinking about probabilities. Such framings were found to have a positive effect both on assessment burden and accuracy in the univariate case [198]. Recognition of the cognitive burden of assessing dependence has existed at least since [241], who supports probabilistic methods, in particular the quadrant probability, due to its intuitive decision analytic interpretation in comparison to statistical methods.

3.5.2 Statistical methods

Despite some objections to the direct elicitation of moments of distributions or even cross moments, such as non-observability [214], the literature offers some interesting findings and conclusions about the direct assessment of statistical measures of association (and alternative formulations).

Forms of statistical dependence parameters

Direct (rank) correlation Directly asking experts for the natural input of a dependence model is seen by some as a natural way of eliciting dependence. Often, this is a correlation coefficient. One option is to ask experts for an estimate of the (rank) correlation between pairs of variables X and Y. A framing might be simply:

> "Consider variables X and Y. What is the (rank) correlation between them?"

This usually refers to the Spearman's rank correlation coefficient (see Appendix B) which is defined on the interval of $[1, 1]$. A value of $\rho = 1$ denotes the strongest possible negative correlation, $\rho = 0$ expresses that X and Y are uncorrelated while $\rho = 1$ refers to the strongest possible positive relation. An advantage of eliciting rank correlations over product moment ones is that the interpretation of the former is independent of its marginal distributions implying that its values are always in the aforementioned interval. Nevertheless, for choosing the appropriate correlation co-efficient, an analyst has to take into account what kind of relationship is assessed. Rank correlations, such as Spearman's version, assume monotonicity while Pearson's product moment coefficient (see Appendix B) can only be meaningful for linear relationships [337].
An obvious precondition for this type of dependence parameter to be intuitive is a certain level of familiarity with statistical measures. Therefore, several (conflicting) conclusions have been made from research on this query variable. Some studies, such as [214], [297], as well as [172], view a direct method as unreliable. The latter for instance conclude that even trained statisticians will have difficulties with this method even when being presented with graphical output in form of scatterplots. This is in agreement with [275] who conclude that experts judge the degree to which variables deviate from perfect correlation rather than directly assessing dependence of variables when shown a scatterplot. Yet according to other research, a direct elicitation has performed better in comparison with other assessment methods. [341], [73] and [74] concluded that eliciting a correlation coefficient is more accurate than other dependence variables (in relation to empirical data) as well as more coherent. The better performance in comparison to other methods is primarily attributed to sufficient normative expertise of the experts.

Ratios of (rank) correlation When considering higher orders of dependence, a direct way to elicit this information from experts is through ratios of (unconditional) rank correlations. In this method, experts assess the "relative strength" of each rank correlation [287]. [288] and [112] present it as an alternative to conditional exceedance probabilities for higher dimensions which have the requirement to assess large conditioning sets that make the elicitation exercise rather unintuitive.
When defining unconditional rank correlations in the exemplary BN of Figure 3.2 as r_{X,Y_1} and r_{X,Y_2}, then for the first conditional rank correlation, $\rho_{X,Y_2|Y_1}$, the ratio $R = r_{X,Y_2}/r_{X,Y_1}$ would be elicited. The corresponding question might be framed as:

> "Given your previous estimate, what is the ratio of r_{X,Y_2} to r_{X,Y_1}?"

Similar to the conditional probabilistic techniques, the values that an expert can assess are restricted for each subsequent ratio. Imposing bounds ensures coherence but makes the elicitation less intuitive. Empirical comparisons to probability of exceedance have neither shown a superior nor an inferior performance. Nevertheless, the proponents of this method found that experts often think in terms of unconditional correlations rather than ratios. The intention of the ratio framing is to prompt experts to think in terms of relative influence between variables. However, there is no way of ensuring the experts will follow the proposed path.

Verbal An indirect statistical approach to elicit experts' beliefs about dependence is through the use of a pre-defined scale. The most common way to do so is by using verbal descriptions that correspond to certain correlation coefficient values. For instance, [73] use a scale of seven points on which the relationship between X and Y is measured as $S_{X,Y}$. The points range from 1 describing a very strong negative relationship up to 7 which denotes a very strong positive relationship. Accordingly, 4 refers to no relationship. The transformation to Spearman's rank correlation is done through $\rho = (S_{X,Y}4)/3$. Despite its obvious subjectivity in determining the scale due to the rather informal translation of verbal qualifiers, a good performance in terms of coherence and accuracy can be observed in empirical studies using this method. Moreover, the method is intuitive which makes it popular. In the area of human reliability analysis, [386] introduce the Technique for Human Error Rate Prediction (THERP) which uses a verbal scale for assigning the dependence level between human errors. The conditional probability for failure between tasks A and B is computed as $P(B|A) = (1+KP(B)/(K+1)$ where K is assessed via verbal qualifiers of complete dependence ($K = 0$) to high ($K = 1$), medium ($K = 6$), low ($K = 19$) and zero dependence ($K = \inf$). The dependence assessment method in THERP is the foundation of various further developments of dependence modelling efforts in this area.

Coefficient of determination A method that has been used rather rarely but that is still possible is to elicit the coefficient of determination. For this, [74] propose to ask for the percentage of variance explained as it would result from regressing one variable on another (R^2). [397] uses this idea to construct a dependence measure which can be used in the elicitation of copula parameters. It is proposed for a common risk factor model within the context of the Program Evaluation and Review Technique (PERT) for which dependence is modelled with a DB copula (see previous section). PERT is an operational research technique for analysing and scheduling projects whereas the uncertainty in completion time is typically of interest. For modelling the dependence between the (aggregated) common risk factor Y (factors influencing project completion time) and random variable X (completion time), first $R(X) = ba$, i.e. the range where realisations of X can be observed, is defined. Next, the range of the conditional distribution, $R(X|Y = y, \phi)$, is specified where the state of different common risk factors that result in the aggregate risk of Y as well as the dependence parameter of the DB copula, ϕ, are known. From this, the dependence measure $\xi(X|Y, \phi) = (1R(X|Y, \phi)/R(X))100\%$ is derived (see reference for full elaboration). This measure can be thought of as the average percent reduction in the range of X when the state of common risk factor, Y , is given. Suppose Y defines the set of possible risk factors, $Y = \{\mathsf{rain}, \mathsf{no\ rain}\}$, and the range of X is the length of an activity, e.g. a project's duration in days. Then the query question is asked as follows:

> "Not knowing the state of the common risk factor, Y, a value of x has been assessed for X. Suppose you knew the state of the common risk factor, Y, on average within a spread of how many days could you now assess the completion of this activity,X?"

An expert's assessment of 5 days would then correspond to 50%, i.e. this is the percentage of uncertainty that is explained by knowing the state of the risk factor. The author highlights that the elicitation question is framed in terms of X which is an observable quantity. While an intuitive appeal for the method is mentioned, no empirical results in terms of performance or cognitive burden for experts have been reported. Extensions for use with different copula families are achieved by slightly altering the formulation of $R(X)$.

Assessment burden for statistical methods

Overall, the statistical methods are seen as intuitively accessible for experts and enjoy favourable feedback in terms of assessment burden [73, 341]. Especially verbal scales are seen as directly applicable and have therefore enjoyed further consideration. [73] report that for statistical methods training and feedback for follow-up studies improved accuracy. This is confirmed by expert studies with frequent feedback on correlation assessments, such as weather forecasters [51]. Similarly, neurological experiments in which experts get frequent feedback on correlation coefficients find evidence for a human ability to "learn" the effect of varying correlation coefficients [432]. Even though not conclusive, there are reasons to believe that statistical methods for dependence elicitation are more intuitively understandable, or at least "learnable", when compared to other approaches. This is nevertheless a signal rather than a strong conclusion also due to the fact that statistical methods have often been tested (only) for simple examples (e.g. heightweight relationships) rather than complex elicitation problems.

With regards to the complexity of problems for which experts might assess a correlation directly, [241] offers perhaps one of the most detailed discussions. He addresses the cognitive complexity required for assessing correlation coefficients directly in terms of their operational, decision-analytic and intuitive interpretation. From this perspective, according to him the necessary level of cognitive processing for assessing a correlation coefficient can be rather high. For instance, when interpreting a (rank) correlation in terms of concordance and discordance of hypothetical observations of a population (which has a clear and intuitive meaning) experts might have to assume (the rather unintuitive idea of) an infinite population (see Appendix B for the definition of rank correlations). The product moment coefficient is seen as (even) more difficult to assess as it is not ordinally invariant which (as aforementioned) inhibits a simple, intuitive understanding given that any assessment is interpreted with regards to the transformations made to the marginal distributions.

3.5.3 Other methods

In the following, methods that do not fit the categories above (for reasons which will be explained) are considered.

One such method is proposed by [2] who elicit joint probabilities through univariate distributions and isoprobability contours. In other words, dependence is elicited indirectly. We present this approach separately because experts express preferences over binary gambles with identical payoffs rather than providing probabilistic (or numerical) responses directly.

Loosely, an isoprobability contour is a collection or set of points which have the same cumulative probability. In order to elicit the 50^{th} percentile of a contour for two variables of interest, X and Y, experts assess first the common quantiles for X, e.g. the median, $x_{0.5}$, the 75^{th} quantile, $x_{0.75}$, and so forth. Then, the experts are offered two gambles, for which the authors propose the framing of:

> A: "You receive a fixed amount, z , if the outcome of variable X is less than $x_{0.5}$ and variable Y takes any value." (short:($x_{0.5}$,y_{max})) B: "You receive the same fixed amount, z , if the outcome of variable X is less than $x_{0.75}$ and the outcome of variable Y is less than y_1 (with $y_1 < y_{\mathsf{max}}$)." (short:($x_{0.75}$,y_1))

The formulation has been altered to fit the wording of the earlier framings for elicitation questions in this review. The value for y_1 is specified and depending on the response of an expert, y_1 is adjusted until the expert is indifferent between the two gambles. If no indifference is achieved, the process ends after a pre-determined number of iterations and upper and lower bounds for y_1 are specified to choose the midpoint. With the same framing, the experts continue choosing between binary deals while varying the quantiles for X and values of y_n, such as $A : (x_{0.75}, y_1)$ and $B : (x_{0.9}, y_2)$ and so forth. Through enough iterations, i.e. a sufficient number of indifferent choices that determine the points on the contour, its 50^{th} percentile is assessed. Once this is achieved, the joint cumulative distribution of any point, $(x, y) \in [x_{\mathsf{min}}, x_{\mathsf{max}}] \times [y_{\mathsf{min}}, y_{\mathsf{max}}]$, can be derived with one additional assessment of a univariate quantity such as a marginal probability for any of the variables of interest, $F_X(x)$, by finding the point (x_1, y_{max}) lying on its isoprobability contour. The joint probability assessment reduces then to a univariate problem through $F(x, y) = F(x_1, y_{\mathsf{ma}}) \triangleq F_X(x_1)$.

This approach was tested with graduate students who assessed the joint probability of weight and height relationships within their university cohort. A monetary incentive was offered for obtaining honest and accurate answers. The authors conclude that this method is sensible with respect to difficulty, monotonicity and accuracy, but still discuss some possible assumptions that might ease the assessment burden. As a main advantage over conventional methods they mention the flexibility in analysing the results by deriving various dependence measures from the elicited outcomes.

Another method that has been proposed for specifying dependence through expert judgements and which fits into this sub-section is [312]. They consider a Bayesian updating procedure for dependent binary random variables. Again, dependence assessments are not made directly, but a threshold copula approach is used to fully determine the dependence structure.

3.6 Dependence elicitation in the empirical literature

Following the previous discussions about elicitation in various modelling contexts and about forms of elicited dependence parameters, this section provides an overview of the common approaches in practice that are prevalent in the case study literature.

While a complete outline of our review methodology can be found in Appendix A, we briefly present how the literature on eliciting dependence has been reviewed. The objective for this literature review is two-fold:

- Assess the application areas and approaches to dependence modelling that are used in case studies published in the literature, in order to evaluate the reach of the different elicitation methods.

- Ensure that the theoretical review is complete and includes a broad variety of perspectives.

As a first step, a search strategy was formulated that defined the key words used in order to ensure a thorough search of potential references of interest. For this, we started combining common key words of expert judgement studies such as “expert judgement” (British English)/“judgment” (American English) or “elicitation” itself, with general key words of dependence elicitation and modelling. This was refined by including key words for specific dependence modelling techniques and dependence parameters. Next, appropriate databases were identified, again starting generally before searching explicitly in archives of the topic’s research areas, such as Operational Research and Decision as well as Risk Analysis. For evaluating the relevance of references under equal principles, criteria that specify the fit to this review (and which are outlined completely in Appendix A) had to be defined. The candidate references were then filtered and lastly, the selected findings were distinguished between theoretical and practical contributions as the latter were categorised for the overview in this section.
In total 53 references have been identified in which dependence has been elicited within decision analysis/risk analysis case studies (in some, more than one dependence parameter was elicited). The elicited dependence parameters are categorised as conditional (exceedance) probabilities (CP/CEP), point estimates as well as quantiles, joint probabilities, statistical parameters such as correlation coefficients, verbal and other methods (whereas other methods here differ from the ones presented in Section 3.5). A detailed list of the identified case studies can be found in the additional supplementary material. The empirical references were investigated from different perspectives and Figure 3.7 summarises how the empirical literature is clustered.
In the upper-left corner it can be seen that the predominant dependence model for which dependence is elicited is a BN (61.02%). For that, the main dependence parameters elicited are conditional (exceedance) probabilities (point estimate) and verbal scales. Dependence is elicited much less frequently for copulas, BLM approaches or parametric multivariate distributions.
For dependence parameters per aggregation method an apparent finding is that performance-based methods are used mainly together with conditional (exceedance) probabilities (through quantile assessments). This might not be surprising given that the authors for these studies come from the same expert judgement school that emphasises the use of performance-based combination and quantile (rather than point) assessment. In total performance-based weighting is used in 22.03% of all case studies, just more than equal weighting which is used in 18.64% of all references. Most significant however is that for 37.28% of all case studies the aggregation method is not described or mentioned at all.
When clustering the experts’ domains and substantive expertise (upper-right corner), it is shown that in particular for environmental and ecological studies as well as in risk analyses for infrastructure problems, dependence is elicited through probabilistic variables (CP/CEP), point and quantile assessments, together with verbal methods. Overall, the main domains that experts have substantive expertise in are environmental/ecological (38.98%), infrastructure (23.72%) and energy decision analysis/risk analysis (11.86%). In this context, it is an interest-

ing observation that the relevant case studies (see supplementary material) are mostly published in domain-specific journals rather than journals with a focus on the modelling and hence elicitation methodology. This gives a few indications about the status-quo of the empirical side of the research problem addressed in this review. It shows that modelling dependence together with expert judgement for quantification is a research problem that is (actually) recognised in the identified domains. Interestingly, the domains have an established tradition of applying rigorous risk analysis methods, often stemming from the area of probabilistic risk analysis [84]. Further, this finding indicates that due to a focus on the application in the fields, there is less focus on developing new theory for dependence modelling and elicitation which would be found in journals with a methodological focus. This allows for cross-fertilisation of various findings discussed in the previous sections and our review aims to establish a contribution for this.

While a recommended number of experts from marginal elicitation protocols is between 5 and 10 experts (see aforementioned references on guidance for univariate elicitation), for dependence elicitation this is taken into consideration only in 15.25% of the cases. Slightly more often (22.03%), less than five experts are used. Again, the predominant percentage (33.89%) for “multiple” implies a less clear documentation.

While these findings are not conclusive they offer an indication on the predominant approaches in the case study literature.

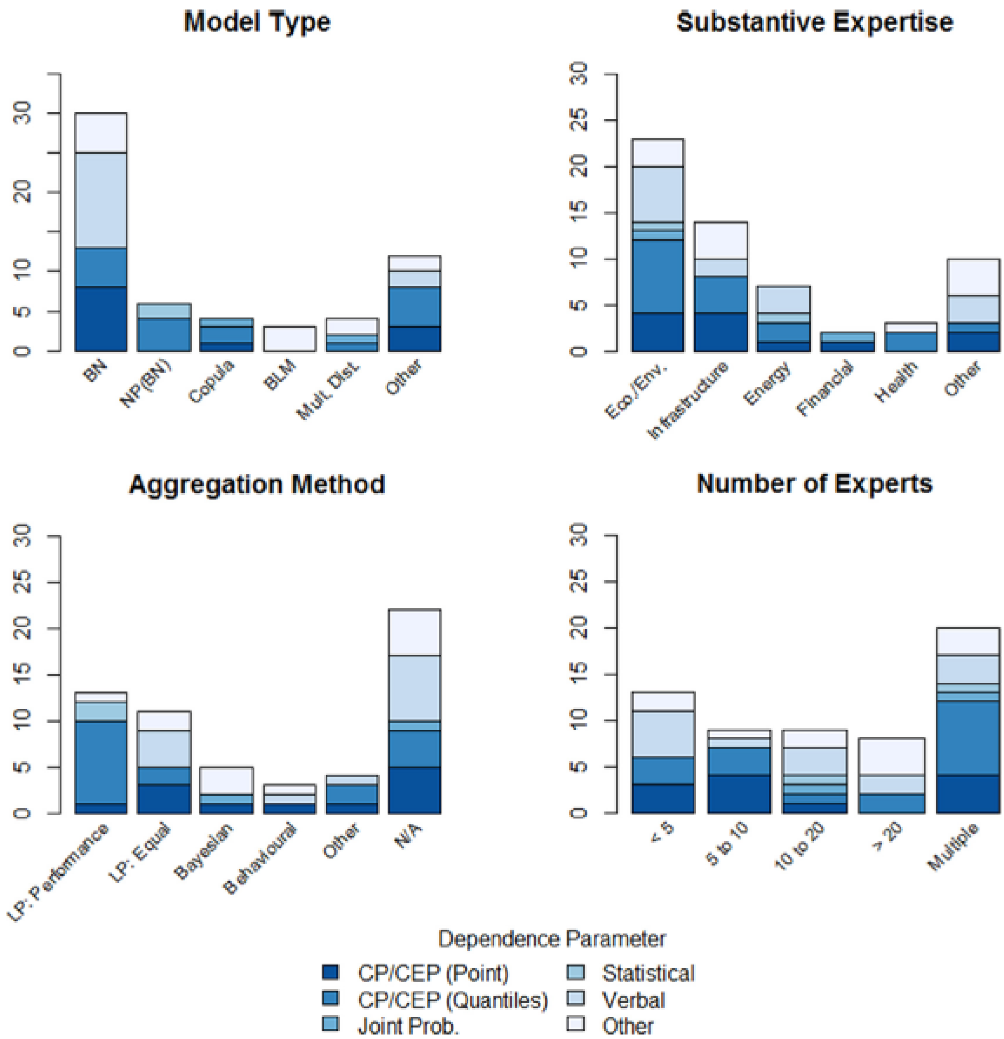


Figure 3.7: Different perspectives on elicited dependence parameters' use in the case study literature.

3.7 Chapter conclusions and further research

We have argued that multivariate decision models under uncertainty are becoming more and more prevalent whether as BNs (continuous or discrete), as parametric multivariate models, or as separate specifications of univariate distributions together with copulas to model the dependencies. We also argued that this immediately leads to the need for elicitation techniques to quantify these models.
The biggest challenge in the use of expert judgement to quantify dependence is in the way we manage the elicitation burden for experts. Implicit in our discussion above is that the elicitation burden has two key dimensions:

- The required quantity of information - there is a danger that large amounts of information required from experts will burden them too much in terms of time and the prolonged intensity of the task.
- The complexity of the required information - there is a danger that the experts might not be able to hold all the required information in the forefront of their minds while considering complex scenarios in which (conditional) probabilities are required.

Both considerations should guide the analyst to choose between ways to reduce the elicitation burden, by: simplifying the parameterisations of models, by considering the qualitative and quantitative steps of elicitation separately, or by finding ways of explaining in practical terms the quantities that are being elicited. However, there is a clear trade-off between easing the elicitation burden and building models that replicate the important behaviour of real world systems. Satisfying both the above requirements is challenging and under research. The qualitative structure provided by a Bayesian network is one example in this direction. However, often it is difficult to decide on a particular form of network. We may have situations, for example, where a multivariate distribution can be estimated from data for moderate values of the variables, but where qualitatively different behaviour can occur in the tails. Expert judgement may be more appropriate in this context, as stochastic behaviour is then driven by different relationships between variables.
The literature review illustrates clearly the challenge faced in finding better ways to elicit multivariate uncertainties: In many cases the reported studies use students instead of (costly) experts. Often, when experts are used, they are asked to only provide guidance on parameters, but the justification for the chosen parametric family is not given. Clearly, for purposes of validity and verification we need to evolve better practices in selecting such families. Otherwise we are not in a strong position to challenge poor operational practice, such as the prevalence of the Gaussian copula used widely in financial modelling prior to the recent crash, and almost certainly still in equally wide use [354].
Finally, in the chapter we have focused on the use of expert assessment in quantifying multivariate distributions. However, the revolution in data analytics is using machine-learning and expert systems rather than human experts. It is therefore worth reflecting on the relative benefits, similarities and complementarities of these approaches. An individual human expert may be considered analogous to a particular machine-learning model, and the empirical result that machine-learning model averaging typically gives better results than any one of the models on their own, reflects older observations in the use of expert judgement that weighted averages of expert assessments are better calibrated than

individual experts. However, the human expert may be able to provide simplifications through parametric model choices, and insights into model "phase changes" that the machine-learning models struggle with, because the data does not go far enough into the tail. The research challenges we have set out above will help us find a more satisfactory approach to combining human and machine expert judgements for uncertainty modelling.

Appendix

Review Methodology

In this Appendix we set out how publications and studies have been searched and selected. The selection protocol is the systematic process shown in Figure 1 below. The subsequent steps allowed for a thorough identification of references of interest by evaluation under common principles. Next, each step of the selection process is detailed.

Search Strategy Formulation

n order to find as many potentially relevant references as possible, appropriate keywords had to be defined. This included expressions such as "elicitation", "expert judgement" (British English) and "expert judgment" (American English) together with keywords for multivariate distributions and dependence measures. These were linked by AND/OR operators as standard in scientific databases. In a first round, a general approach was taken. In addition to the terms "dependence" and "association", common probabilistic expressions, e.g. "multivariate distribution", "joint distribution" and "bivariate distribution", were used. This reflects the varying ambitions in dependence modelling and elicitation, where focus might be on bivariate relationships or higher orders of dependence. Search terms for expert judgement studies were then combined with specific measures of dependence that are often used as elicited variables. These correspond to the ones presented later, e.g. "conditional probability", "joint probability", "correlation" and so forth. A third round considered search terms from dependence modelling techniques. General expressions such as "dependence model", "dependence modelling" (British English) and "dependence modeling" (American English) etc. were used before including specific techniques such as "Bayesian Belief Networks", "Belief Networks" (or "Nets"), "Copulas" and so forth.

Identification

In the identification stage, the search strategy was applied for various scientific databases and other sources of references. Candidate references were identified by scanning titles together with abstracts. General scientific databases such as *Web of Science*, *Ebsco Host* and *Scopus* served as starting points. Then, the focus was narrowed down to more specific archives from areas in which expert judgement and uncertainty analysis are of interest. These relate mainly to Operational Research and Management Science as well as statistics and risk analysis. Lastly, university databases were searched. For all, identification was restricted to the first 300 search results by the ordering criterion relevance. For completeness, bibliographical cross-references and personal archives of contributing scholars in the field of expert judgement were queried.

Evaluation

Next, identified candidate references were evaluated according to their relevance and fit for purpose for this review. References were rejected if they were either *out-of-scope* or *outdated*.

Out-of-Scope. Being out-of-scope might be subject to several criteria. A first criterion concerns a psychological research focus of candidate references. Similar to the heuristics and biases research stream for expert judgement studies in general, dependence elicitation has been considered at least indirectly by psychological researchers which results in an overlap of search terms. In particular causal learning theory makes use of the concept of association and occasionally statistical measures of dependence such as correlation. Main references for normative and descriptive theories on association learning in human cognition as well as causal reasoning are mentioned in the Introduction. Another out-of-scope criterion concerns references studying dependence between experts. Given the high similarity in keywords and abstracts, their evaluation has proven to be tedious. While combining experts is considered in Section 6 of this review, the case of aggregating dependent assessments is not addressed. Some main sources for this topic are Hora (2010), Kallen and Cooke (2002) and Jouini and Clemen (1996). Further, references were rejected if they focus on multivariate modelling but experts are solely used for eliciting marginal distributions. As this process is not always clearly presented within abstracts (nor captured in the titles), filtering these studies out was again time-consuming. Another related field that shares common search terms is investigating structural uncertainty of complex phenomena yet without quantification. In this context, some references had to be sorted out if considering dependence only qualitatively. Lastly, alternative expressions of uncertainty nor MCDA-based approaches were not considered as explained in the Introduction of this review.

Outdated. Apart from being out-of-scope a source might be outdated. While no specific threshold date had been defined, less recent studies were evaluated according to their current importance in the field.

Selection

Lastly, an important distinction was made between empirical and theoretical foci among the selected references. While information from both was used, the former was recorded separately for the overview of the practical experiences in the literature.

Chapter 4

Considerations and approaches along the SEJ process when eliciting dependence

This chapter[1] addresses the main elements of structured expert judgement processes for dependence elicitation. We introduce the processes' common elements, typically used for eliciting univariate quantities, and present the differences that need to be considered at each of the process' steps for multivariate uncertainty. Further, we review findings from the behavioural judgement and decision making literature on potential cognitive fallacies that can occur when assessing dependence as mitigating biases is a main objective of formal expert judgement processes. Given a practical focus, we reflect on case studies in addition to theoretical findings. Thus, this chapter serves as guidance for facilitators and analysts using expert judgement.

4.1 Chapter introduction

A structured approach to eliciting multivariate uncertainty is encouraged as it supports experts to express their knowledge and uncertainty accurately, hence producing well-informed judgements. For instance, cognitive fallacies might be present when experts assess dependence which can inhibit the judgements' accuracy. Therefore, mitigation of these fallacies is a main objective of an elicitation process. Further, a structured process addresses other questions which affect the reliability of the elicited result and hence model outcome, such as aggregating various judgements. Lastly, a formal process makes the elicited results transparent and auditable for anyone not directly involved in the elicitation.

Complementary to the case of eliciting univariate uncertainty, this chapter's objective is to outline the main elements of formal expert judgement processes for multivariate uncertainty elicitation. This is done by discussing theoretical and empirical findings on the topic, though the reader should note that fewer findings are available for eliciting joint distributions than for the elicitation of univariate

[1] Based on: Werner, C., Hanea, A. M., Morales-Nápoles, O. (2018). Eliciting multivariate uncertainty from experts: Considerations and approaches along the expert judgement process. In: Dias, L. C., Morton, A., Quigley, J. (eds.) Elicitation: The science and art of structuring judgement, New York: Springer International Series in Operations Research and Management Science, 171-210

quantities.
The structure of this chapter is as follows. In section 4.2, the importance of formal expert judgement processes is discussed and an overview of the necessary adjustments for dependence elicitation is given. This provides the reader with the scope of the topic. Section 4.3 outlines the heuristics and biases that might occur when eliciting dependence. Then, section 4.4 discusses the preparation of an elicitation (or the pre-elicitation stage) which for instance entails the choice of the elicited forms and the training of experts. In section 4.5, we present considerations for the actual elicitation phase, including structuring and decomposition methods as well as the quantitative assessment. In section 4.6, we review required alterations of the process for the post-elicitation stage, such as when combining the expert judgements. Finally, section 4.7 concludes the chapter by summarising the main points addressed and discussing the status-quo of this research problem.

4.2 Structured Expert Judgement processes: an overview

The necessity for a structured and formal process when eliciting uncertainty from experts, such as in form of probabilities, has been recognised since its earliest approaches. For instance, it has been acknowledged in the area of Probabilistic Risk Analysis (PRA) which comprises a variety of systematic methodologies for risk estimation with uncertainty quantification at its core [84]. From a historical perspective, main contributions in PRA have been made in the aerospace, nuclear and chemical process sector. Hence, after expert judgement was used only in a semi-formal way in one of the first full-scale PRAs, the original Reactor Safety Study[2] by the US Nuclear Regulatory Commission [77], major changes towards a more scientific and transparent elicitation process were made in the subsequent studies, known as NUREG-1150 [78, 227]. When reflecting on the historical development of PRA, [82] highlights the improvements made through a traceable elicitation protocol as a newly set standard and main achievement for expert judgement studies.
Another pioneering contributor to formal approaches for expert judgement is the Stanford Research Institute (SRI). The Decision Analysis Group of SRI similarly acknowledged the importance of a formal elicitation process when eliciting uncertainty from experts. Therefore, they developed a structured elicitation protocol that supports a trained interviewer through a number of techniques to reduce biases and aid the quantification of uncertainty [375, 377].

Following from these early contributions, various proposals for formal expert judgement processes have been made and its various components were further developed. While not one particular step-by-step process to follow exists given the varying and particular objectives of each elicitation, there is agreement regarding which high level steps are essential. Fairly complete elicitation protocols are for instance presented in [274], [176], [87], [407], [75] and [150]. Even though

[2]The study is also known as WASH-1400 and as the *Rasmussen Report* due to Norman Carl Rasmussen. At that time, the use of expert opinion for assessing uncertainties was often viewed highly sceptical, however a main challenge was that until then no nuclear plant accident had been observed. Therefore, the report, together with its use of expert opinion, was only revived due to the Three Mile Island accident (1979). After the incident, the report's results were prescient. In particular, the inclusion of human error as a source of risk made the case for expert judgement.

these references explicitly address the case of eliciting a univariate quantity, they serve as guidance for our purpose of presenting and discussing the considerations for eliciting dependence.

The elicitation of dependence follows historically from advances made for eliciting univariate uncertainty and an overview of the historical development of expert judgement in risk analysis is presented in [82]. This development is not surprising given that marginal distributions need to be specified (at least implicitly) before any dependence assessment can be made. Furthermore, univariate quantities are (typically) more intuitive to assess. Whereas some findings for eliciting univariate uncertainty are still applicable in the multivariate case, for other parts of the process adjustments need to be made. Fig. 4.1 shows the various elements of elicitation processes with the modifications that are necessary when eliciting dependence.

Regarding the different roles during an elicitation, in this chapter we consider the situation of a specific decision or risk analysis problem that is of importance for a *decision maker*. *Experts* assess the uncertainty on the variables without any responsibility for the model outcome or consequences of the later decision. The experts are chosen based on their substantive (also subject-matter) expertise, meaning they are experts on the particular topic of the decision problem. This implies that the experts might not have normative expertise, thus they are not statistical or probabilistic experts. The *facilitator*, who manages the actual elicitation part of the overall process, might be either the same person as the decision maker or an independent third type of attendee at the elicitation workshop. The facilitator clarifies any questions from the experts. An *analyst* on the other hand is usually in charge of the whole process. This includes the preparation of the elicitation and the finalisation of results afterwards. Such a situation with a given, formulated problem and clearly defined roles is often the case, however other ones are possible. [155] discusses various elicitation contexts and their potential implications.

We regard an elicitation as successful if we can be confident that the experts' knowledge is captured accurately and faithfully, thus their uncertainty is quantified through a well-informed judgement. However, the assessments' reliability might be still poor if little knowledge about the problem of interest prevails. This often implies that there is high uncertainty in the area of the decision problem overall.

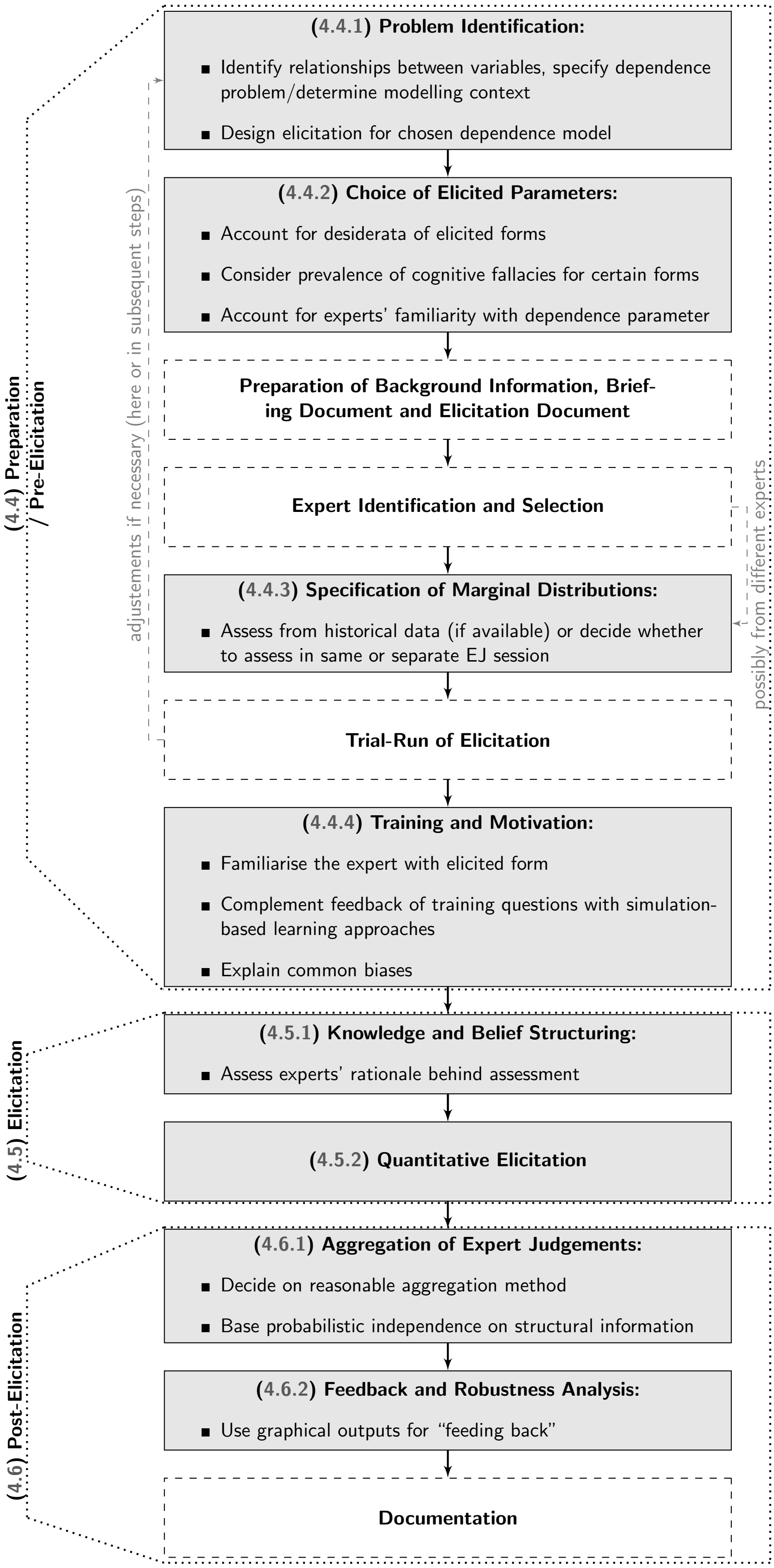


Figure 4.1: Overview of the expert judgement process adjusted for eliciting dependence (steps discussed in this chapter are in grey).

4.3 Biases and Heuristics for Dependence Elicitation

In this section, we review main findings from the behavioural judgement and decision making literature on assessing dependence as psychological research shows that experts are not guaranteed to act rationally when making such assessments. Hence, the goal of this section is to raise awareness of departures from rationality in the hope to minimise them in the elicitation. Briefly, rationality implies that experts make assessments in accordance with normative theories for cognition, such as formal logic, probability and decision theory. Irrationality, on the other hand, is the systemic deviation from these norms. While this definition suffices here, the topic is much more complex and a critical debate on the concept of rationality can be found in [379] and [311]. In contrast to normative theories that describe how assessments ought to be made, descriptive research investigates how assessments are actually made. This relates directly to our earlier definition of a successful elicitation (section 4.2) that states our aim of eliciting accurate and faithful assessments from experts. In other words, a successful elicitation aims at mitigating a range of potential biases.

For expert judgement, in particular two types of biases, cognitive and motivational, are of importance as they can distort the elicitation outcome severely.
Cognitive biases refer to the situation in which experts' judgements deviate from a normative reference point in a subconscious manner, i.e. influenced by the way information is mentally processed [171]. This bias type occurs mainly due to *heuristics*, in other words because people make judgements intuitively by using mental short-cuts and experience-based techniques to derive the required assessments. The idea of a heuristic proof was used in mathematics to describe a provisional proof already by [327], before the term was adopted in psychology, following [367] with the concepts of *bounded rationality* and *satisficing*.
Motivational biases may deviate experts' judgements away from their true beliefs. In other words, experts ought to make the most accurate judgements regardless of the implied conclusion or outcome, yet they do not. Motivational biases happen consciously and depend on the experts' personal situations. For instance, social pressures, wishful thinking, self-interest as well as organizational contexts can trigger this type of biases [284]. Given that motivational biases are not different for univariate and multivariate uncertainty assessments we will not consider them in our review in section 4.3.2.

Table 4.1: Main biases and heuristics for dependence elicitation

	Name	Reference(s)[a]	Description	Originates with	Suggested Remedies
conditional	Confusion of the inverse	[272], [127], [106], [193]	Experts confuse conditional probabilities of $\mathsf{P(X\|Y)}$ with its inverse $\mathsf{P(Y\|X)}$	representativeness heuristic, causal interpretation, "non-natural" base-rates	elicit frequency formats (if possible) [306, 270], structure rationale/relationships, include graphical aids (see [151])
	Causality heuristic	[8], [391]	Experts overestimate $\mathsf{P(X\|Y)}$ when perceiving causal relationship, i.e. Y *causing* X	causal interpretation, base-rate neglect	avoid single, focal scenarios as experts' rationale, evoke alternative scenarios, use experts with different backgrounds [430]
	Insufficiently regressive prediction	[219]	Experts *translate* one scale to the other, not adjusting for imperfect association	representativeness heuristic, predictive interpretation	specify reference class with central tendency and variability → assess individual case → adjust/calibrate [306]
	Bayesian likelihood bias	[132], [122]	Experts are more conservative than Bayes' Theorem implies	representativeness heuristic, base-rate neglect	decompose into assessing priors (odds) and likelihoods (ratios) [284]
conjunction	Confusion of joint and conditional probabilities	[134]	Experts confuse joint and conditional probabilities	causal/temporal interpretation	address semantic misunderstandings in training, structure rationale/scenarios/ functional relationships
	Conjunction fallacy	[8], [391]	the conjunction of X and Y is judged as more probable than X and Y individually	causal interpretation, base-rate neglect	demonstrate probability logic [284]
concordance	Cell A strategy	[371], [13], [221]	Experts overvalue joint presence of variables (in bivariate assessment)	predictive interpretation	clarify underlying assumptions, such as rarity assumption
general	Illusory correlation	[66], [131]	Experts base assessment on false (pre-existing) belief about relationships	availability bias, causal interpretation	as for availability: provide probability training, counter-examples, relevant statistics (if available) [284]

[a] only main and/or original references listed

Regarding the mitigation of biases, a motivational bias can be addressed in a technical way by introducing (strictly proper) scoring rules or as well by the direct influence of a facilitator who encourages truthful answers. A cognitive bias is mainly counteracted through training of experts, decomposing and/or structuring the experts' knowledge prior to the quantitative elicitation as well as a sensible framing of the elicitation question(s). The latter also entails the choice of the elicited form.
Over the last 40 years, the number of newly identified heuristics and biases has increased tremendously. Nevertheless, only a few findings are available for the case of assessing dependence. We present these findings in the remainder of this section and Table 4.1 provides an overview. For discussions on some main univariate biases, we refer to [245] and [284].
As can be seen in Table 4.1, most identified heuristic and biases that are applicable for the case of multivariate uncertainty concern conditional assessments, such as conditional probabilities. While conditionality is a common way to conceptualise probabilistic dependence, it is shown that in addition to the explicit fallacies (as introduced in the following), understanding and interpreting conditional forms remains a challenge in today's statistics and probability education [114]. An explanation for this difficulty comes from [63] who note that a main focus of probability education is on frequentist approaches to probability together with (idealised) random experiments, such as coin tosses. Regarding conditional probabilities, such a position is however problematic as with equally likely cases, reducing the subspace has no clear impact on the equal probabilities. With a subjective view on probability on the other hand, a conditional probability is more intuitive as one simply revises judgements given new information that has become available [53].

4.3.1 Causal reasoning and inference

Before we address in detail the biases from Table 4.1, recall that we are interested in the experts' ability to assess dependence in accordance with the subjective dependence definition presented in the introduction. Usually this is done through specifying a dependence parameter and we address the choice of an elicited form in section 4.4.3. While emphasizing that assessing dependence, e.g. as a correlation, is not the same as claiming a causal relationship, we consider experts' mental models about causal relationships as a main determinant for their assessments (despite the missing statistical noise). Therefore, we briefly address findings of behavioural studies on causal reasoning and inference first.
The concept of causation itself is highly debated[3] and its discussion is out of scope here, yet it is proposed that in most situations people believe that events actually have causes. In other words, their belief is that events mainly occur due to causal relationships rather than due to pure randomness or chance [192]. Moreover, it is argued that people have systematic rules for inferring such causal relationships based on their subjective perception [134]. They then update their mental models of causal relationships continuously and might express summaries of causal beliefs in various forms, such as serial narratives, conceptual networks or images of (mechanical) systems [192].

[3]There has been ongoing philosophical debate about the meaning of causation. While some refuted the concept of causation in science altogether [352], others focused on specific aspects. For us, probabilistic causation [384] and its perception/inference are of interest. [203] proposes one of the most established accounts for that. He proposes a (unobservable) causal mechanism which is inferred through the regularity of an effect following a cause.

Due to incomplete knowledge and imperfect mental models, we emphasize the concept of probabilistic causation [384]. A formal framework that has been used widely for representing probable causes in fields such as statistics, artificial intelligence, as well as philosophy of science and psychology, is a probabilistic (causal) network. The topic of causation within probabilistic networks is however not without criticism and generates debate. Extensive coverage of this topic is given in [376], [319] and [350].
A first type of information for inferring a probabilistic causal relationship is the set of necessary and sufficient conditions that constitute a presumed background of no (or only little) causal relevance (i.e. they are not informative for inference), but which need to be in place for an effect to happen. These conditions are known as *causal field*. For instance, when inferring the cause(s) of someone's death, being born is a necessary and sufficient condition, nevertheless it is of little relevance for establishing a causal explanation [134]. The causal field is a key consideration when structuring experts' beliefs about relationships as it relates to model boundaries and determines which events should be included in a graphical (or any other) representation of the system of interest. We discuss structuring beliefs in section 4.5.1.
Another type of information that is assumed to be in place for making causal inferences is summarised as *cues-to-causality*. Most of these origin with [203] and comprise temporal order, contiguity in time and space, similarity, covariation, counterfactual dependence and beliefs about the underlying causal mechanism as seen by events' positions in causal networks [192]. Generally, the presence of multiple cues decreases the overall uncertainty, even though conflicting cues increase it. The way in which these cues are embedded in the causal field and how both types of information together shape one's causal belief is shown by [134] with the following example:

> "Imagine that a watch face has been hit by a hammer and the glass breaks. How likely was the force of the hammer the cause of the breakage? Because no explicit context is given, an implicitly assumed neutral context is invoked in which the cues-to-causality point strongly to a causal relation; that is, the force of the hammer precedes the breakage in time, there is high covariation between glass breaking (or not) with the force of solid objects, contiguity in time and space is high, and there is congruity (similarity) between the length and strength of cause and effect. Moreover, it is difficult to discount the causal link because there are few alternative explanations to consider. Now imagine that the same event occurred during a testing procedure in a watch factory. In this context, the cause of the breakage is more often judged to be a defect in the glass."

This simple example shows that by changing the contextual factors while keeping the cues constant, someone's causal belief can change rather dramatically.
The ways in which these types of information influence a causal perception are important for the remainder of this section as experts' causal beliefs and inferences often serve as candidate sources for several biases.

4.3.2 Biased Dependence Elicitation: an Overview

In the following, the main cognitive fallacies that can occur when eliciting dependence, as shown in Table 4.1, are presented in more detail. In addition to

introducing the examples that the original researchers of the different biases propose, we illustrate each bias with a simplified example from the area of project risk assessment. Explaining all biases with the same example allows for a better comparison between their relevance and the context in which they apply.
Suppose, we manage a project with an associated overall cost. The project's overall cost is determined by various individual activities which are essential for the project completion and which each have their own cost. We denote the cost of an individual activity by c_a and when we distinguish explicitly between two different activities, we do so by indexing them as 1 and 2, so as c_{a_1} and c_{a_2}. It follows that we are interested in modelling and quantifying the dependence between the individual activities' costs and the dependence's impact on the overall cost. Note that assuming independence between the activities might severely underestimate the likelihood of exceeding some planned overall cost. In order to better understand the dependence relationships, we take for instance into account how the individual activities can be jointly influenced by environmental and systemic uncertainties. In this simple example, we consider whether (and if yes, how) such uncertainties impact the activities' costs, e.g. due to affecting the durations of certain activities. The duration or time an activity takes is represented by t_a. A main area of research in PRA that focuses on modelling implicit uncertainties, which have a joint effect on the model outcome but that are not well enough understood to consider these factors explicitly, is common cause modelling. For an introduction, see [84].

Confusion of the inverse

A common way of eliciting dependence is in form of conditional judgements, such as conditional probabilities (section 4.4.2). A main bias for conditional forms of query variables is the *confusion of the inverse* [272, 127, 106, 193]. [402] provide a list of alternative names proposed in the literature. For that, a conditional probability $P(X|Y)$ is confused with $P(Y|X)$. In our project risk example, this might happen when considering the time that an activity takes and whether this influences its own (but also other activities') cost. When eliciting the conditional probability $P(c_a \geq v|t_a \geq w)$ where v and w are specific values, an expert might confuse this with its inverse, $P(t_a \geq w|c_a \geq v)$.
An empirical research area in which this fallacy has been studied more extensively is medical decision making. It is shown that medical experts often confuse conditional probabilities of the form $P(\text{test result}|\text{disease})$ and $P(\text{disease}|\text{test result})$. In a pioneering study, [127] reports this confusion for cancer and positive X-ray results. More recently, [393] lists the confusion of the inverse among the main misunderstanding that "educated citizens" have when making sense of probabilistic or statistical data. Further, [393] outlines several cases in which being prone to this fallacy has led to false reporting about risk in the media.
One explanation for confusing the inverse is attributed to the similarity of X and Y. Therefore, some researchers suggest that this bias is linked to the better known *representativeness heuristic* [218, 217]. For that, people assess the probability of an event with respect to essential characteristics of the population which it resembles. For dependence assessments this implies that experts regard $P(X|Y) = P(Y|X)$ due to the resemblance or representativeness of X for Y and vice versa [306]. For instance a time-intensive project activity might resemble a cost-intensive one and vice versa.
Another explanation for this fallacy is related to neglecting (or undervaluing) base-rate information [234, 146]. Generally, the *base-rate neglect* [219, 27] states

that people attribute too much weight to case-specific information and too little (or no) to underlying base-rates, i.e. the more generic information. With regards to confusing the inverse, [168] distinguish between *natural* and *non-natural* sampling spaces. A natural sampling space is one that is accessed more easily in one's memory (this may or may not be the sample space as prescribed by probability theory). In the fallacy's classical example of $P(\text{test result}|\text{disease})$ for instance, the sample space of "people with a disease" often comes to mind easier than that of "people with a certain test result", such as "positive", given that the latter can span over several types of diseases. Similarly in our project risk example, for $P(c_a \geq v | t_a \geq w)$ an expert ought to regard the activities exceeding a certain duration before thinking of the activities within this subspace that exceed a certain cost. However, the sample space of activities exceeding a specified cost might be more readily available so that from this the proportion of the activities exceeding a certain time is considered.

A last suggested source for the inverse fallacy stems from experts' (potentially) perceived causation between X and Y. [325] attribute a potential confusion between conditionality and causation to similar wordings such as "given that" or "if". Remember that temporal order is important for determining the cause(s) and the effect(s) of two or more events. For instance, [29] show how causal beliefs influence the inference of their temporal order and vice versa, i.e. how temporal order informs causal beliefs. Thus, when eliciting the dependence between two activities' durations, experts might confuse $P(t_{a_1} \geq w | t_{a_2} \geq w)$ with its inverse if the durations are not easily observed, e.g. due to lagging processes, and the first completed activity is seen as causing the other.

In the medical domain, in which this confusion has been observed most often, we note that for $P(\text{test result}|\text{disease})$ the test result is observed first (in a temporal order) even though the outbreak of the disease clearly preceedes in time. Therefore, the cause is inferred from the effect. This is a situation in which [134] see the confusion of the inverse very likely to occur, even though temporal order has no role in probability theory. By some researchers, this is called the *time axis fallacy* or *Falk phenomenon* [141]. Another interesting example from medical research concerns the early days of cancer research and the association between smoking and lung cancer. While it is now established that smoking causes lung cancer, some researchers have also proposed the inverse [267]. Indeed, the question of whether a certain behaviour leads to a disease or whether a disease leads to a certain behaviour can be less clear. A potential confusion of the inverse is then subject to the expert's belief on the candidate cause.

Causality heuristic

The close connection between conditional assessments and causal beliefs can be the source of another cognitive fallacy. In a pioneering study, [8] coined the term *causality heuristic*, claiming that people prefer causal information and therefore disregard non-causal information, such as base-rates with no causal implication. Other researchers (e.g. [40]) have since then confirmed this preference for causal information. At a general level, the causality heuristic relates to causal induction theories in contrast to similarity-based induction [370]. For instance, [271] found that people regarded the statement "bananas contain retinum, therefore monkeys do" as more convincing than "mice contain retinum, therefore monkeys do" which shows that the plausibility of a causal explanation can outweigh a similar reference class.

In the context of conditional assessments, it is noteworthy that people assess a higher probability for $P(X|Y)$ when a causal relation is perceived between X and Y, even though according to probability theory, a causal explanation should make no difference in the assessment [141]. This is shown further by people's preference to reason from causes to effects rather than from to effects to causes [192]. As a result, causal relations described as the former are judged as more likely than the latter even though both relations should be equally probable. For our example of assessing $P(c_a \geq v|t_a \geq w)$, we therefore need to consider whether experts perceive a causal explanation and how it influences the assessment outcome.

In an experimental study, [391] asked subjects whether it is more probable that (a) a girl has blue eyes if her mother has blue eyes?, (b) a mother has blue eyes if her daughter has blue eyes?, or (c) whether both events have equal probability? While most participants (75) chose the correct answer (c), 69 participants responded (a) compared to 21 that chose (b). Whether this result can be fully attributed to the role of participants' perception of causation is however questionable given other possible influences on the assessments such as semantic difficulties [134]. Nevertheless, it is an indicator for how experts are led by preferences about perceiving a conditional relation (which might contradict the elicited one) once they regard the variables as causes and effects.

While sometimes being regarded as a different bias, the *simulation heuristic* [228] affects judgements in a very similar manner. Here, the premise is that conditional probability judgements are based on the consideration of if-then statements. This is an idea originating with [334] and his "degree of belief in p given q", roughly expressing the odds one would bet on p, the bet only being valid if q is true. Hence, it is proposed that for assessing a conditional probability, $P(X|Y)$, one first considers a world in which Y is certain before assessing the probability of X being in this world. The simulation heuristic states then that the ease with which one mentally simulates these situations affects the probability judgement. People often compare causal scenarios and tend to be most convinced by the story that is most easily imaginable, most causally coherent and easiest to follow. However, they then neglect other types of relevant information together with causal scenarios that are not readily available for their conception.

Insufficiently regressive prediction

A fallacy that might occur when people interpret a conditional form as a predictive relation is *insufficiently regressive prediction.* [219] show that when assessing predictive relationships, people do not follow normative principles of statistical prediction. Instead, they "merely translate the variable from one scale to another" [219]. In the project risk example, when predicting an activity's cost from its duration, e.g. through conditional quantiles, experts might simply choose the value of the cost's i^{th} quantile based on the time's i^{th} quantile. This is problematic as typically there is no perfect association between the variables. Hence, people do not adjust their assessment for a less than perfect association between the variables. [306] give an example of predicting the height of males from their weight while assuming a correlation of 0.5 between the variables. Then, for a male who is one standard deviation above the mean weight, the best prediction for his height should only be 0.5 standard deviations above the mean height. However, people tend to assess the prediction too close to one standard deviation above the mean height.

A common explanation for this fallacy is again the representativeness heuristic. Regarding one variable representative for the other, e.g. viewing tall as representative for being heavy or a time-intensive project activity as representative for a cost-intensive one, experts disregard the aforementioned imperfect association. As shown in section 4.4.2, eliciting conditional quantiles is one common way to elicit dependence information.

Bayesian likelihood bias

Research investigating experts' conditional assessments in the context of intuitively using Bayes' Theorem[4] formulated what is named (by some) the *Bayesian likelihood bias* [122]. Bayes' Theorem is proposed as a normative rule for revising probabilities given new evidence. The fallacy is that people are too conservative in their assessment [132], at least for certain framings (see [245] for a critical discussion on this fallacy). The univariate equivalent is the *conservatism bias*. It refers to the finding that higher probabilities are underestimated while lower ones are overestimated, i.e. assessments vary less from the mean and avoid extreme values. For $P(c_{a_1} \geq v | c_{a_2} \geq v)$, experts might make too conservative assessments in light of new information about another activity's cost. In a pioneering study by [122], participants assessing the probability of a person's gender given the height, $P(\text{gender}|\text{height})$, tended to underestimate the number of tall men and overestimate the number of tall women.

Confusion of joint and conditional probabilities

A cognitive fallacy that might be present when assessing dependence for events occurring together, i.e. the conjunction of events, such as in a joint probability assessment is the *confusion of joint and conditional probabilities*.

Consider the framing of the elicitation question: "What is the probability of $c_{a_1} \geq v$ and $c_{a_2} \geq v$?" While a more precise framing (specifying that we elicit the joint probability) or eliciting a joint probability still framed differently (see section 4.4.3) would be helpful, it is important to note that from the view of probability theory, when using the word "and", we would expect the expert to assess $P(c_{a_1} \geq v \cap c_{a_2} \geq v)$, i.e. the conjunction of the events. However, it is shown that this is often interpreted differently. For some people "and" implies a temporal order (which has no role in probability theory), so they assess the conditional probability of $P(c_{a_1} \geq v | c_{a_2} \geq v)$ instead [134]. This fallacy is closely related to the confusion of the inverse for which one explanation is based as well on an implicit influence of temporal order.

Conjunction fallacy

A more extensively studied bias that is relevant when eliciting the conjunction of events is the *conjunction fallacy* [391]. In experiments, subjects assessed the probability of a conjunction of events $P(X \cap Y)$ as more probable than its separate components, i.e. $P(X)$ or $P(Y)$, despite its contradiction to probability theory. For instance, when [250] asked participants which of the following two statements is more likely: (a) a randomly selected male has had more than one heart attack, and (b) a randomly selected male has had more than one heart attack and he is over 55 years old, (b) was judged more probable than (a) by

[4]Bayes' Theorem is named after Thomas Bayes (1701-1761) who first proposed it. Since then it has been further developed and had its impact in a variety of problem contexts (see [267] for a historical overview). In its simplest form, for events X and Y, it is defined as $P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$ whereas $P(Y) \neq 0$.

most participants. Similarly, experts in our project risk example might assess $P(c_a \geq v \cap t_a \geq w)$ as more probable than $P(c_a \geq v)$ or $P(t_a \geq w)$ separately. As with the confusion of the inverse, a suggested source for the conjunction fallacy is the representativeness heuristic. However, while this is the most common explanation, it is not without criticism and numerous other candidate sources for this fallacy exist [90, 388]. For example, another explanation is the aforementioned causality heuristic. Hence, the constituent events are related through a causal explanatory variable. The additional information that constitutes the subset is then judged as causally relevant, as e.g. in our earlier examples being over the age of 55 is seen as causally relevant for having a heart attack, and an activity exceeding a certain duration for exceeding a certain cost.
In the context of assessing conditional probabilities, [249] discuss the conjunction fallacy through the related concept of *disjunction errors*. People assess the conditional probabilities through subordinate and superordinate categories. For example in their example, a subordinate category, Asian flu, was regularly judged as more probable than its superordinate category, flu, given a set of symptoms. A possible explanation is based on a predictive interpretation for the conditional probability. Participants view the symptoms as more predictive for the subordinate category and base their likelihood judgement on it.

Cell A strategy

Some research focuses on interpreting and assessing dependence as the concordance of events whereas this is based on a frequency (or cross-sectional) interpretation for the event pairs. In other words, it explicitly requires a population to draw from. At the most general level, this relates to people's ability to assess dependence in form of the "perhaps simplest measure of association" [241], the quadrant association measure. It gives the probability that the deviations of two random variables from (for instance) their medians have the same signs, i.e. positive or negative. This is closely related to assessing a concordance probability which is introduced in section 4.4.3.
In some situations this is the way how people perceive association between (binary) variables and a research stream that investigates this form of dependence perception is *associative learning* [279]. A common cognitive fallacy is the *cell A strategy* [221] which is named like this for reasons that will become apparent. While certain activities are highly standardised and performed similarly across numerous projects, it is still rather an idealised case to serially observe whether or not the duration of an activity exceeded a certain value for j projects with $j = 1, 2, ..., J$, i.e. whether $t_{a,j} \geq w$ or $t_{a,j} < w$, before obtaining this information for its cost. Despite its idealisation, this is how experts would perceive dependence in this case. Similarly in his pioneering study, [371] worked with medical experts and the variables referred to symptoms and diseases. The experts were given information about the presence or absence of a disease following information on the presence or absence of a symptom and then assessed its correlation. This information can be ordered within four quadrants. The upper left corresponds to the presence of both variables, the lower right shows the joint absence and the remaining two quadrants relate to one variable being present while the other is absent. Whereas in normative theory, all four quadrants should be equally informative, it is found that people focus on the joint presence of both variables disproportionally in relation to the observed frequencies, so that this quadrant has a larger impact on the assessment. This quadrant

has also been called cell A when labelling the four quadrants from A to D[5] which explains the name of this fallacy. It suggests that subjects fail to use all relevant information available and in fact, a preference order exists in form of $(X_+, Y_+) > (X_+, Y_-) \approx (X_-, Y_+) > (X_-, Y_-)$ [268]. [264] offer two explanations. The first considers the frequencies (or observations) per quadrant as a sample from a larger population and assumes presence is rare ($P < 0.5$) while absence is common ($P > 0.5$). Then a joint presence is more informative to judge a positive relationship in contrast to joint absence. In other words, it would be more surprising to observe a joint presence rather than a joint absence. The second explanation relates to hypothesis testing and since the quadrant of joint presence is evidence in favour of the hypothesis, this is again (typically) more informative in contrast to both non-joint quadrants that are evidence against it.

Illusory correlation

A cognitive fallacy that is not subject to the specific form of an elicited variable but applies at a general level is known as *illusory correlation*. For this, experts assess that two uncorrelated events show a (statistical) dependence or the correlation is (at least) overestimated. Note that this bias is a systematic deviation that experts may make consistently and not simply a false belief that one expert has but not another. Illusory correlation can be present due to prior beliefs that people have about the co-occurrence of events so that a statistical dependence is expected even though actual observations/data do not confirm this.
In their pioneering research in psychodiagnostics, a field of psychology studying the evaluation of personality, [66] found that medical experts assessed an illusory correlation for the relation of symptoms and personality characteristics. The phenomenon of assuming a correlation where in fact no exists was since then confirmed in different settings and experiments [131] and explains various social behaviours, such as the persistence of stereotypes [185].
One explanation for the (false) expectation of a correlation is that it is triggered by the availability bias. This bias implies that people are influenced considerably by recent experiences and information that can be recalled more easily [390]. For instance, one might be overvaluing the recent observation of a co-occurrence of two events by regarding it as a commonly observed co-occurrence. In our project risk example, this could apply when having recently observed a project delay before seeing its cost exceeding a certain value and regarding this co-occurrence as a frequent observation for similar type of projects. Another source of this fallacy is attributed to pre-existing causal beliefs [40]. In this regard, the prior belief about the correlation stems simply from a false belief about an underlying causal mechanism, as shown in the causality bias.

4.3.3 Implications of biases for the elicitation process

After having presented the main biases that are relevant for eliciting dependence from experts in various forms, we briefly outline the implications that these findings have for the design of the elicitation process.
One finding is that various biases are triggered from the different possible ways that experts might interpret a dependence relationship. In particular, for conditional forms of elicitation, such as conditional probabilities, it is crucial for

[5] When, $+$ indicates the presence of variables X and Y, and $-$ their absence, the quadrants can be presented as:

$A : (X_+, Y_+)$	$B : (X_+, Y_-)$
$C : (X_-, Y_+)$	$D : (X_-, Y_-)$

a facilitator to understand whether the experts might assess the conditional relationships based on similarity/representativeness, causation (e.g. temporal order), or predictive power. As shown, each of these different interpretations can have an effect on the amount and type of information that experts take into consideration when making assessments. In other words, each of the interpretations biases the outcome of an elicitation in a certain way. While more research is necessary to understand how different interpretations are triggered and affect an assessment, we highlight the importance of structuring experts' knowledge and beliefs about a dependence relationship qualitatively, prior to the quantitative elicitation. This ensures that the decision maker and the experts have the same understanding about the dependent variables and more insight about experts' interpretation might be provided. Further, it helps experts to clarify their own understanding and interpretation. This is essential for ensuring confidence in the resulting elicitation outcome as well as for supporting transparency and reproducibility of the expert judgement process.
In addition, the different interpretations and their implications should be addressed in a training session for the experts, in which misunderstandings, such as semantic ones, are resolved. Then, common pitfalls, such as confusing conditional statements and conjunction of events, can be avoided.
Another finding is that several of the presented fallacies originate with (and are closely linked to) more common biases that are not only observed when assessing dependence, e.g. the representativeness heuristic, base-rate neglect and availability bias. For these, research has addressed debiasing methods through alternative framing of elicitation questions, eliciting variables in various forms and training. [284] discuss and give an overview to debiasing methods. Further, Table 4.1 lists specific debiasing techniques for the discussed biases.

4.4 Elicitation process: Preparation/Pre-elicitation

As can be seen in Fig. 4.1, the elicitation process starts already before actually interacting with any experts. The different elements of the preparation (or pre-elicitation) phase ensure that the decision maker's problem is addressed properly and in accordance with the underlying model for which the right variables need to be quantified by suitable experts. In addition, the choices made in this phase allow the experts to assess the uncertain variables as intuitively as possible. In the following, we present the various elements of the this part in more detail.

4.4.1 Problem identification and modelling context

The first step in an elicitation process is the identification of the actual problem at hand in accordance with the decision maker or stakeholder. This step has been termed for instance *background* [75] or *preparation* [306] and includes typically not just the definition of the elicitation's objective but also the identification of the variables of interest.
When drawing conclusions from one of the earliest experiences on formal processes for probability elicitation, [375] referred to this step as the *deterministic phase*. They describe it as the part of the modelling process in which relevant variables are identified and their relationships are determined before uncertainty assessment is considered (in the *probabilistic phase*).
Likewise for dependence elicitation, a main consideration during this part of the process is to design the elicitation in accordance with the underlying dependence

model. A multivariate stochastic model might be pre-determined by the decision maker or is decided upon at this point in accordance with the analyst. In this regard, a broad variety of dependence models exists and their applicability is subject to particular problem situations as they serve different purposes and allow for varying degrees of scrutiny. [412] review the elicitation for several dependence models and discuss how decisions in the modelling context are related to the elicitation by outlining elicitation strategies for three different, broad dependence modelling situations which were introduced in the previous chapter.
The implication for the remainder of the process is that the choices in the different modelling contexts are determined by the level of understanding about the dependencies to be modelled and therefore formulate our variables of interest. These in turn, define the applicability of elicited forms for a satisfactory representation of the experts' information in the model. Therefore, decisions on the model strongly affect the choice of which dependence parameter to elicit as discussed next.

4.4.2 Choice of elicited parameters

The next step in the preparation phase is the choice of an appropriate elicited form for the dependence information. [412] review commonly elicited dependence parameters extensively with regards to the modelling context (section 4.4.1) as well as the assessment burden for experts. These two considerations for choosing an elicited form formulate already main desiderata for this choice, however more are worth discussing.
While some desiderata are the same as for eliciting univariate uncertainty, others are of particular concern when eliciting multivariate quantities. Two desiderata that stem from the univariate case, are: 1.) a foundation in probability theory, and 2.) the elicitation of observable quantities. A foundation in probability theory ensures a robust operational definition when representing uncertainty. Observable quantities are physically measurable, and having this property may increase the credibility and defensibility of the assessments [79]. Moreover, the form of the elicited variable should allow for a low assessment burden. [214] emphasise practicality in this regard. The elicited variables should be formulated so that experts feel comfortable assessing them while their beliefs are captured to a satisfactory degree. For the former, the elicited parameter should be kept intuitively understandable and for the latter, the information given by the experts should be linked (as directly as possible) to the corresponding model. When eliciting dependence, it might be preferred (for instance due to a potential reduction in the assessment burden) to elicit a variable in a different form than the one needed as model input, in which case we need to transform the elicited variable. Then, it is important to measure and control the degree of resemblance between the resulting assessments (through the model) and the dependence information as specified by the expert [238]. The transformation of dependence parameters is typically based on assumptions about their underlying bivariate distribution. For instance, when transforming a product moment correlation into a rank correlation, the most common way assumes bivariate normality [241]. Similarly, when transforming a conditional probability into a product moment correlation, we might assume an underlying normal copula [287]. A potential issue is that positive definiteness is not guaranteed [238], leading to the next desideratum which is coherence. Coherence means that the outcome should be within mathematically feasible bounds. If it is not, it might need to be adjusted such that it still reflects the expert's opinion (as good as possible). Another solution to inco-

herence is to fix possible bounds for the assessment a priori, even though this can severely decrease the intuitiveness of the assessment. Both solutions are rather pragmatic and show why forms of elicited parameters that result in coherent assessments while being intuitive should be preferred. A last desideratum relates to the (mathematical) aggregation of numerous expert judgements (section 4.6.1). When combining expert judgements, it is desirable to base this combination on the accuracy of experts' assessments measured by performance against empirical data. Therefore, the availability of related historical data based on which we can measure such performance is preferred. While there is no query variable that fulfils all of these desirable properties, the desiderata serve as guidance for which elicited parameter to choose under certain circumstances. For instance, an analyst might chose an elicited form that corresponds directly to the model input given a familiarity of the experts with the dependence parameter, therefore having intuitiveness ensured.

At a broad level, most elicited forms can be categorised into *probabilistic* and *statistical* representations. Table 4.2 outlines some main elicited forms in more detail.

We note that the majority of approaches for eliciting dependence fall under the probabilistic umbrella. Probabilistic forms have two main advantages: they (usually) elicit observable quantities and they are rooted in probability theory. Moreover, they are the direct input into various popular models, such as discrete BNs [319, 317] and its continuous alternative [188]. For instance, [412] found in a review of the literature on dependence elicitation and modelling that 61% of case studies, in which dependence was elicited, a BN was used for modelling the dependence. The predominant form for the elicited parameter was a conditional probability (point estimates and quantile estimates).

A potential issue with the forms elicited in the probabilistic approaches, such as conditional and joint probabilities, is that they are regarded as non-intuitive and cognitively difficult to assess. [73] compare their assessment with other approaches, such as the direct assessment of a correlation coefficient, and found that conditional and joint probabilities were among the worst performances for coherence and in terms of accuracy against empirical data, i.e. not well-calibrated. In particular, joint probability assessments seem cognitively complex.

Table 4.2: Overview of elicited forms

	Name	Definition	Framing	Assessment		
			"Consider the variable pair, X and Y. [...]"	independence	positive	negative
probabilistic	Conditional Probability	$\mathsf{P}(X > x_i \mid Y \geq y_i)^a$	"[...] Suppose now that Y is observed above your/its median value. What is the probability that X lies also above your/its median value?"	$\mathsf{P}(X > x_i)$	$\in (\mathsf{P}(X > x_i), 1]$	$\in [0, \mathsf{P}(X > x_i))$
	for higher dimensions	$\mathsf{P}(X > x_i \mid Y_1 > y_{1,i}, Y_2 > y_{2,i})$	"[...] Suppose that not only Y_1 but also Y_2 is observed above your/its median value. What is now your probability that X is above your/its median value?"	$\mathsf{P}(X > x_i \mid Y_1 > y_{1,i})$	as above	as above
	Joint Probability	$\mathsf{P}(X \leq x, Y \leq y)$	"[...] What is the probability that both are within the lower (upper) k^{th} percentage of their respective distributions?"	$F_X(x)F_Y(y)^b$	$F_X(x)$ or $F_Y(y)$	towards 0
	Concordance Probability	$\frac{\sum_{a=1}^{n-1}\sum_{b=a+1}^{n} 1_{C*}((x_a,y_a)(x_b,y_b))}{\binom{n}{2}}$ $C* = (x_a - x_b)(y_a - y_b) > 0$	"Consider two independent draws, (x_a,y_a) from population a and (x_b,y_b) from population b. Given $x_a > y_a$ holds for a, what is your probability that $x_b > y_b$ holds for b?"	0.5	towards 1	towards 0
	Expected Conditional Quantiles	$\mathsf{E}(F_X(x) \mid Y = y_i)$	"[...] Given the value for Y has been observed at its i^{th} quantile, y_i. What is X's value in terms of its quantile?"	0.5	towards max $[F_Y(y), 1 - F_Y(y)]$	towards min $[F_Y(y), 1 - F_Y(y)]^c$
verbal/statistical	Direct Correlation	e.g. Kendall's τ, Spearman's ρ or Pearson's ρ^*	"[...] What is the (rank) correlation between them?"	$\rho = 0$	$\rho = 1$	$\rho = -1$
	Verbal	e.g. $\rho = \frac{S_{X,Y} - 4}{3}$	"[...] Assess the strength of their relationship as: *strong positive* , *positive*, *slightly positive*, *neutral*, *slightly negative*, *negative* or *strongly negative*."	$S_{X,Y} = 4$	$S_{X,Y} = 7$	$S_{X,Y} = 1$

[a] i refers to i^{th} quantile, e.g. $i = 0.5$ for the median; [b] i.e. the product of the cumulative distributions; [c] if $F_Y(y)$ is above its median

This is even true for independence assessments which are (typically) among the easier judgements to express. A further concern is the assessment of a conditional probability with a higher dimensional conditioning set, as discussed in [287] and [289]. The growing conditioning set poses a challenge for experts and this method is (in its current form) difficult to implement. Similarly, expected conditional quantiles (percentiles) are difficult to assess as they require the understanding of location properties for distributions together with the notion of regression towards the mean [74].

As a more accurate and intuitive probabilistic way to assess dependence, concordance probabilities have been proposed [172, 73, 165]. A requirement, which may restrict the variables of interest that can be elicited in this way, is the existence of a population to draw from and a certain familiarity with the population.

Alternatively to eliciting probabilistic forms, we can ask experts to assess dependence through statistical dependence measures. While theoretical objections, such as non-observability [214], persist for the elicitation of moments and similarly cross-moments, they seem to perform well with respect to various desiderata (other than theoretical feasibility). For instance, the direct elicitation of a (rank) correlation coefficient is shown to be accurate and intuitive in some studies [74, 73, 341, 295], even though some research is not in agreement with this finding [172, 214, 176]. The contrasting opinions may arise from the difference in normative expertise that the experts in the studies have or as well from the difference in the complexity of the assessed relationships. For example, in the studies which conclude that eliciting a correlation coefficient is accurate and intuitive, the assessed correlations are on rather simple relationships, such as height-weight, or as well on relationships between stocks and stock market indices. This suggests that regarding relationships for which experts have a certain familiarity and maybe even some knowledge about historical data, the direct statistical method is indeed advantageous. Support for this conclusion comes from findings of weather forecasting. Here, experts obtained frequent feedback on correlations which allowed them to become accurate assessors [51]. Neurological research concludes similar findings after evaluating the cognitive activity in a simulation game where participants obtained regular feedback on correlation assessments [432].

An indirect statistical approach is the assessment of dependence through a verbal scale that corresponds to correlation coefficients (or other dependence parameters). [73] for example provide a scale with seven verbal classifiers. Generally, verbal assessment is seen as intuitive, directly applicable and has therefore enjoyed further consideration. [386] introduce the Technique for Human Error Rate Prediction (THERP) which uses a verbal scale for assigning multivariate uncertainty between human errors. Since its introduction, THERP has been developed extensively in the field of human reliability analysis (HRA) and it has been applied in various industries (see [280] for a review on modelling and eliciting dependence in HRA).

Further, some BN modelling techniques, originating with noisy-OR methods [317], make use of verbal scales. For instance, in the ranked nodes approach, random variables with discretised ordinal scales are assessed by experts through verbal descriptors of the scale [145].

While these are the main approaches for eliciting a dependence parameter, note that when quantifying some models, such as parametric multivariate distributions and regression models, more commonly so called *hyperparameters* are elicited. They allow (through restructuring) for eliciting (mainly) univariate vari-

ables.
For a more detailed and comprehensive review of the elicitation methods and elicited forms mentioned above together with some additional ones, see [412].

4.4.3 Specification of marginal distributions

Before dependence can be elicited, the marginal distributions for the variables of interest need to be specified. In some situations, this information is available from historical data and we can simply provide the experts with this data (if they do not know it already). If this is not the case however, we need to elicit the information on the marginal distributions prior to eliciting dependence. This is important as otherwise the experts base their dependence assessments on different beliefs.
Consider for instance, we elicit dependence from experts in a conditional form. If the marginal distributions have not been specified formally, each expert will base their assessment on their own implicit judgement and as a result each assessment will be conditional on different marginal probabilities. While this leads to dependence assessments which are not comparable and therefore cannot be combined for model input, the implicitly specified marginal probabilities are also likely to lack the scrutiny that a formal elicitation process would allow for. In other words, even if eliciting multivariate uncertainty only from a single expert, a formal process for specifying the marginal distributions is still highly encouraged to ensure less biased and better calibrated assessments. Note that if we omit the specification of the marginal distributions, experts might even refuse to assess dependence as they regard the process as flawed.
Various expert judgement methods exist to elicit univariate quantities (as presented elsewhere in this book) and the process is similarly complex as the one presented here. This is an important remark as we need to decide whether all (univariate together with multivariate) variables are elicited in the same session or whether this is done separately. Eliciting all variables in one session is likely to be tiring for the experts while arranging two separate elicitation workshops might be challenging in terms of availability of experts and organisational costs.

4.4.4 Training and motivation

Training and motivating are likely to improve elicitation outcomes for various reasons, one of which being the effort to mitigate motivational and cognitive biases [199]. Recall from section 4.3.2 that although it is possible for experts to have an intuitive understanding of probabilistic and/or statistical dependence parameters, psychological research shows that interpreting and assessing dependence is often cognitively difficult and results may be distorted. Therefore, we try to counteract the influence of biases and a main approach to achieve this is to train and motivate experts. As aforementioned, motivational biases are not specific to quantifying multivariate uncertainty and are therefore not discussed in this chapter. Consequently, we will further consider only training (not motivating) experts.
Generally, a training session serves to familiarise the experts with the form in which the query variables are elicited by clarifying its interpretation. For univariate quantities this (typically) includes introducing the experts to particular location parameters, such as the quantiles of a marginal distribution. This ensures that these are meaningful to the experts and they feel comfortable assessing them. Further, experts are made aware of the main cognitive fallacies

that might affect their assessments so that they can reflect on them and make a well-reasoned judgement by taking a critical stance. While this ability is an important characteristic of someone's statistical literacy [158], we emphasise a pragmatic approach to training experts as even experienced statisticians often have difficulties with such critical examining and reasoning.

For assessing multivariate uncertainty, the objectives are similar. As concluded in section 4.3.3, main determinants of cognitive biases when assessing dependence are the different interpretations of the elicited forms (in particular of the conditional form). Recall that causal, predictive as well as similarity-based interpretations have a misleading influence on assessments. Therefore, a first focus of an effective training is on explaining the correct interpretation of the dependence parameter to be elicited. This involves an emphasis on the probabilistic and statistical features, such as randomness, in contrast to causal, predictive as well as similarity-based relationships. For instance, causal relationships are often regarded as deterministic, i.e. if Y is understood as the cause of X, then it follows that $P(X|Y) = 1$ as X is always present when Y is present. However, $P(X|Y) = 1$ is not claiming a causal relationship and we might need to account for other factors that affect X and Y [114]. As aforementioned, the confusion of the inverse as well as the causality heuristic (section 4.3.2) are two main biases that can be explained by such a misleading interpretation. In this regard, some researchers have mentioned their concern about the language that is used in many statistics textbooks to teach fundamental concepts such as independence [114]. For instance, the phrase "whenever Y has no effect on X" is used to explain that two variables, X and Y, are independent and their joint distribution is simply the product of their margins. However, for many experts, the term "effect" might imply a causal relationship. This shows that training on the elicited form should also address any semantic misunderstandings at this step of the elicitation process.

In the same manner, we can address the other misinterpretations. For example, in order to avoid that conditional assessments are based on similarity, i.e. resemblance of X for Y, we should stress that the assessments might also be influenced by other factors. As such, a specific outcome, such as a certain diagnosis, can be *typical* for a certain disease but still unlikely [306].

While probabilistic reasoning is commonly included in school curricula, its teaching is often done through formula-based approaches and neglects real-world random phenomena [28]. Therefore, it is common that experts hold misconceptions on probabilistic/statistical reasoning which are hard to eradicate. In fact, they might even consider this kind of reasoning as counterintuitive. A possibility to enhance a better understanding of these concepts might be to complement the practice of forming probability judgements and providing feedback on training questions (as commonly done before elicitations) with simulation-based approaches. There is empirical evidence that multimedia supported learning environments successfully support students in building adequate mental models when teaching the concepts of correlation [260] and conditional probability [133].

Once the experts are familiar with the elicited form and its correct interpretation, an additional focus of the training session is on outlining the common biases as identified in section 4.3.2. This allows the experts to obtain a better conceptual understanding and we can address potential issues more specifically, such as recognising that a conditional probability involves a restriction in the sample space, distinguishing joint and conditional probabilities or as well distinguishing the inverses.

4.5 Elicitation process: Elicitation

After the preparation/pre-elicitation phase is concluded, the actual elicitation starts. Note that this is the phase in the overall process in which the facilitator works interactively with the experts, first when supporting experts to structure their knowledge and beliefs (or rationale), and second when eliciting the uncertain variables quantitatively. We will explain both steps in more detail below.

4.5.1 Knowledge and belief structuring

Neglecting existing knowledge and data that can be relevant for an assessment is another reason for biased elicitation outcomes in addition to misinterpreting the elicited form (section 4.3.3). However, experts often have cognitive difficulties in *exploring* the underlying sample space to a satisfactory degree. Therefore, they need support for making better use of their knowledge and beliefs, a procedure we call *structuring* or which is also known as *knowledge evocation* [59]. Apart from mitigating biases, structuring experts' knowledge and beliefs about a joint distribution prior to eliciting dependence quantitatively is essential for ensuring confidence in the later assessment as well as for supporting transparency and reproducibility of the expert judgement process. In fact, when quantifying multivariate uncertainties, identifying the factors that are relevant to the particular problem is a main outcome of the structured expert judgement process. In other words, knowledge structuring allows for obtaining an insight into the details of experts' understanding about the dependence relationships, thus their *rationale*. [201] views this step of probability elicitation as the most challenging one in the process. This is due to people possessing knowledge about uncertain events or variables which is composed of many fragmented pieces of information, often all being of high relevance. Further, people typically know more than they think, therefore neglecting this step could result in less informative judgements.
Structuring knowledge might be part of a hybrid approach to dependence modelling in which qualitative, structural information about dependence relationships is specified first, before probabilistic quantification is considered. Typically, graphical models are used to reduce the cognitive load on experts' short term memory, even though other structuring methods, such as directed questions (checklist-based approaches) have been proposed [59]. Some commonly used graphical models are *knowledge maps* [201], *event and fault trees* [84], *influence diagrams*[6] [358, 202] and BNs (section 4.4.1). Note that we can nevertheless also include a structuring part when quantifying a dependence model with experts which offers no such a graphical representation. In this case, rather than including the result of knowledge structuring in the actual model, we use it solely for supporting the experts. That being said, when reviewing the literature on eliciting dependence in probabilistic modelling, [412] found that the dependence model, which is used most often together with expert judgement, is in fact a BN. A reason for its popularity is likely that it allows for an intuitive graphical representation. According to [438], deriving the structure of a BN can be achieved in four ways. First, the structure can be specified through transforming existing probabilistic models of the problem, such as event and fault trees. Such a transformation is straightforward as the necessary structural information is already given in the existing models and it can be sensible as BNs are more

[6] In the literature on event trees and influence diagrams, the idea of *decomposition* is often mentioned as it describes a "divide and conquer" technique [199] that allows to ease the assessment in particular of conditional probabilities (see e.g. [232]).

flexible. Second, a BN structure can be inferred from some empirical or physical model. Third, the structure can be built based on existing historical data and fourth, it can be elicited from experts. The last way is of most interest for us as it is a common situation that not only the probabilistic information needs to be elicited from experts, but also the qualitative relationships [326, 149]. Further, it corresponds directly to the knowledge structuring part of the process.
[438] propose to begin the structural elicitation with identifying the relevant variables and to achieve this, they refer for instance to organized interviews [190]. Then, the actual arcs are elicited, either interactively (as we describe below) or through reusable patterns of structures [144]. Last, they deal with unquantifiable variables (e.g. through proxies).
As mentioned before, one way to derive the graphical structure is by eliciting the experts' input on these interactively [303]. One advantage of such an interactive procedure is that it allows (typically) for discussion among experts about the justification of nodes and arcs. In other words, pre-existing knowledge is challenged and elaborated on if necessary. Further, experts obtain a greater ownership of the model which they structured themselves so that they are more comfortable in quantifying it later on. A potential difficulty, which needs to be considered, is that the consensus on the final model structure might have been achieved by a dominating expert who dictated the result or due to group-think, i.e. without critical evaluation. Regarding these potential issues, [407] suggest to elicit a structure from each individual expert, whenever there is a concern about not capturing the opinion of less confident experts. Aggregating diverse structural information coherently through rules (as opposed to consensus) is discussed in [55]. While for hybrid dependence models a combined graphical structure is necessary, in terms of knowledge structuring it is also of interest how sharing knowledge and rationales among experts affects a later assessment. For instance, [187] integrate group interaction in a structured protocol for quantitative elicitation as it is shown to be beneficial in assessment tasks.
Besides the initial structuring step, [197] mentions the potential necessity to refine a model structure during the actual quantification. In particular, the violation of conditional independence is of concern. By definition of a BN, the successor nodes (children) are conditionally independent given their parents. If this is not the case when observing the final model, an additional node is required. [317] regards conditional independence therefore as a guiding principle as where it fails, further clarification about an assumed, hidden variable is needed.

4.5.2 Quantitative elicitation

After structuring experts' knowledge and beliefs about the factors that influence the variable(s) of interest, the quantitative assessment follows. This step of the process is also named *encoding* [375]. In this step, experts assess the variable(s) of interest in the form that was chosen to be appropriate with respect to various desiderata (section 4.4.3).
The main considerations herewith are similar to those of eliciting univariate uncertainty. Likewise, we need to decide on how much interaction between the experts we allow for (we address the aggregation of assessments in section 4.6.1). Further, at least one facilitator is present to answer questions regarding the understanding of the query variables. Prior to the session, experts should have received a briefing document which helps them to familiarise themselves with the purpose and structure of the elicitation [85].
As there are no differences to univariate uncertainty elicitation in this part, we

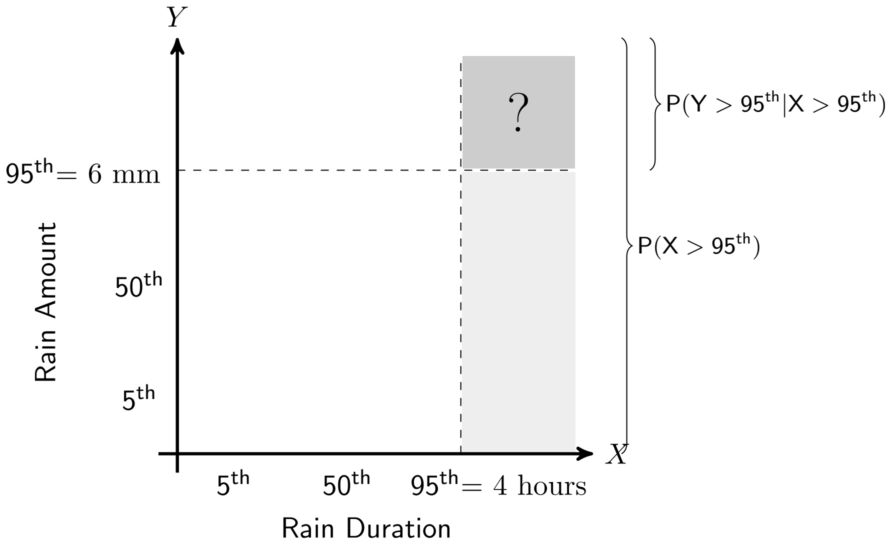


Figure 4.2: Exemplary elicitation question with visualisation

devote the remainder of this sub-section to illustrating an exemplary assessment which has been used similarly in an actual dependence elicitation problem. [295] and [292] elicit and quantify dependence between rain amount and rain duration in the Netherlands through conditional exceedance probabilities. The elicited results are used as model input for quantifying parametric copulas. Modelling dependence in this way informs resilience analysis for critical components of road networks, such as tunnels and road sections. The aim of this analysis is to improve the understanding about the effects of extreme rainfall for the development of probabilistic models in reliable infrastructure risk analysis.
Fig. 4.2 shows a way of presenting experts with the elicitation question:

> For Rotterdam, NL, consider all samples for which the rain duration in *hours* (X) is larger than its 95^{th} quantile (4 hours). What is the percentage of this set of samples, for which the rain amount in *mm* (Y) is also larger than its 95^{th} quantile (6 mm)?
> This can be expressed as $P(Y \geq 95^{th}$ quantile$|X \geq 95^{th}$ quantile) or likewise as $P(Y \geq 6\, mm|X \geq 4\, hours)$.
> Please provide your assessment: ________________

The inclusion of a visualisation can be helpful for experts to obtain a better understanding about the framing of the elicitation question.

4.6 Elicitation process: Post-Elicitation

The last phase in the overall elicitation process (Fig. 4.1) is the post-elicitation part. The two main steps that are of importance here are aggregating the assessments of various experts and providing feedback to the experts. We address both steps in more detail below.

4.6.1 Aggregation of expert judgements

In order to capture a broad perspective on the uncertainties that we model and quantify, we (usually) elicit judgements from a variety of experts. Therefore, a

main aspect of the post-elicitation phase is the aggregation (or combination) of the assessments from several experts.
As in the univariate case, a distinction at a broad level is made between *behavioural* and *mathematical* (or algorithmic) aggregation methods. The first type aims at reaching consensus so that the outcome is a single assessment upon which the group of experts has agreed. This might be achieved within a group elicitation session or through methods, such as Delphi [351]. Given that these methods are the same as for univariate elicitation, they are not further discussed here. Recall however that a potential shortcoming of these methods (in the univariate as well as multivariate case) is that the consensus might be reached through one expert dominating the elicitation discussion or even dictating the elicitation's outcome [155].
For aggregating judgements mathematically, in particular two approaches are common. The first is the *Bayesian approach* which allows for modelling quality aspects of individual expert distributions, for example overconfidence. The second approach is a *pooling function* which is typically seen as more robust and easier to use [200].
For Bayesian aggregation, we apply Bayes' Theorem (section 4.3.2) while regarding the expert judgements as data. If we are interested in an event or unknown quantity x, we elicit its probability or set of quantiles and obtain the experts' individual prior opinions, $f_{0,e}(x)$ for experts $e = 1, 2, ..., E$. We denote the set of elicited distributions as $\underline{D} = (f_{0,1}(x), ..., f_{0,E}(x))$, and get the combined posterior distribution for x, $f_{1,DM}(x|\underline{D})$ through $f_{1,DM}(x|\underline{D}) \propto f_{0,DM}(x)L_{DM}(\underline{D}|x)$. It is then necessary to elicit the likelihood function of observing $\underline{D}$ given x, i.e. $L_{DM}(\underline{D}|x)$ [422]. A Bayesian aggregation model which has been used more commonly is [300].
A pooling function on the other hand assigns weights to individual assessments to derive a weighted combination of the experts' judgements. The weights are either equal for each expert or they reflect an expert's competence or performance (in terms of statistical accuracy, if empirical data can be used for measuring this). For equal as well as performance-based weighting, all weights are non-negative and sum to one. A commonly used pooling function is linear averaging, for which the combined assessment is $DM_{(f_1(x),...,f_n(x))} = \sum_{e=1}^{E} w_e f_e(x)$, with w_e being the weight of expert e. Alternatively, other pooling methods exist, such as logarithmic pooling, for which the combined assessment is defined as $DM_{(f_1(x),...,f_n(x))} = k \prod_{e=1}^{E} f_e(x)^{w_e}$ where k is a normalising constant.
Linear pooling functions originate with [383] and [110] and the legitimacy of their application from an axiomatic perspective is primarily based on *event-wise independence* (or the weak set-wise function) and *unanimity preservation* [7, 266, 116]. The first axiom implies that the collective probability of an event is only determined by the individual probabilities for that specific event (and not that of other ones). Unanimity preservation holds that if all experts give the same assessment, then this will be the collective one.
For aggregating dependence assessments, mainly linear pooling functions have been used [412], which is why we address them in more detail. Before we discuss these however, note that a possible concern with mathematical aggregation in the multivariate case is that not all dependence assessments are preserved. For instance, a linear combination of correlation matrices is still a correlation matrix, however conditional independencies such as in a BN are not preserved. Further, an axiomatic issue might be that of preserving *probabilistic independence* which ensures that if all experts regard two variables as (conditionally) independent,

then this is preserved in the combined assessment. For several pooling functions (e.g. linear as well as logarithmic ones) this is problematic. However, it might be argued that unless independence assessments are also based on structural judgements (section 4.5.1), i.e. they are not purely accidental, this normative constraint is questionable [55]. Note that this is a question of whether one regards dependence information as fully represented by probabilistic (un-)conditional dependence or only in addition to structural judgements in form of graphical representations (such as in BNs). As we have emphasised in section 4.5.1 that structural information should be elicited either within the same modelling framework or separately, the independence axiom is not of concern and we regard linear pooling methods as applicable for dependence information.

Equal weighting

One option to set weights in a linear pooling function is by equally weighting all assessments (simple average). When eliciting correlation parameters directly, overall accuracy improved in that way through adding experts [424]. The authors tested the robustness by removing/adding experts and found that the mean absolute error (MAE) decreased when the number of experts increased.

Performance-based weighting

Alternatively, [424] also showed that taking the average of only the top performing cohort of experts (in terms of lowest MAE) instead of the whole set of experts reduces the overall MAE further. This finding is consistent with expert judgement studies for univariate quantities [86] and therefore motivated the idea of using a measure of calibration to assess experts' performance in terms of statistical accuracy as a score for multivariate assessments. Before we introduce this score, note that there is an indication that a common calibration method for univariate expert judgements [79] might not be feasible for aggregating dependence assessments [289].
The first and only calibration score for multivariate assessments (according to the authors' knowledge) is the dependence calibration score introduced in [294] which is based on the *Hellinger distance*. In order to assess this score (similar to *Cooke's Classical* model [79]) seed variables known to the facilitator but not the experts are elicited in addition to the target variables. Then, two bivariate copulas f_C (a copula model used for calibration purposes) and f_E (a copula estimated by expert opinions) are used to derive the Hellinger distance, H, which is defined as:

$$H(f_C, f_E) = \iint_{[0,1]^2} \sqrt{\frac{1}{\sqrt{2}}(\sqrt{f_C(u,v)} - \sqrt{f_E(u,v)})^2 \mathrm{dudv}}$$

In [3] an overview of different distances between distributions is given. If the distributions are Gaussian, these distances can be written in terms of the parameters of the Gaussian distributions (i.e. the mean and covariance matrix). Under the Gaussian copula assumption, H may be parametrised by two correlation matrices:

$$H_G(\Sigma_C, \Sigma_E) = \sqrt{1 - \frac{\det(\Sigma_C)^{1/4}\det(\Sigma_E)^{1/4}}{(\frac{1}{2}\det(\Sigma_C) + \frac{1}{2}\det(\Sigma_E))^{1/2}}}$$

Here Σ_C is a correlation matrix used for calibration purposes and Σ_E the one estimated by experts. The *d-calibration* or dependence calibration score is:

$$D = 1 - H$$

The score is 1 if an expert's assessments correspond to the calibration model exactly. Conversely, it differs from 1 as the expert's opinion differs from the calibration model. Under the Gaussian assumption, i.e. when using H_G, the score approaches 1 as Σ_E approximates Σ_C element-wise and it decreases as H_G differs from H_C element-wise. A score equal to zero means that at least two variables are linearly dependent in the correlation matrix used for calibration purposes and the expert fails to express this. Or contrary to this, an expert expresses perfect linear dependence between two variables when this is not the case. For more details, see [293]. In the same paper [293], the method discussed in [294] is extended by using the Hellinger distance to compare a Gumbel copula generated from precipitation data with a copula constructed from experts' assessments of tail dependence between rain amount and duration in Rotterdam and De Bilt, in the Netherlands. The experts' assessments are obtained by a similar framing as shown in section 4.5.2 and varying the elicited quantiles, e.g. 50^{th} and 95^{th} (see [287] for more details). An overview of the results in given in Table 7.1.

In this study, the combination of expert opinions based on the dependence calibration score outperforms individual expert opinions as well as weighting experts equally. In fact, the equal weights approach does not give satisfactory results. We observe that the performance-based aggregation is much closer to the actual empirical rank correlation. Further, it was noticed that experts with highest calibration scores for univariate assessments are not necessarily the experts with the highest dependence calibration score.
In order to combine dependence assessments, experts are weighted according to their dependence calibration score. Similar to the univariate case, a cut-off level is established, either chosen by the facilitator or by optimising the performance of the combination. If an individual expert falls below this level, their score will be unweighted for the pooling function.

4.6.2 Feedback and robustness analysis

Similar to eliciting univariate uncertainty, one of the final steps of the dependence elicitation process is testing the robustness of elicited results and providing feedback to the experts after a combined assessment has been constructed. While this procedure is not much different for the multivariate case, it should be noted that many dependence models produce graphical outputs, such as scatter plots. Depending on the experts' understanding of the graphical output and their willingness to examine such outputs, it might be possible to feedback such a visualisation and assess their agreement with it.

Table 4.3: Dependence calibration results based on rank correlation, Gaussian (H_G) and Hellinger (H) distance (Morales-Nápoles et al. 2016b)

Name	Rotterdam $X > 0.95$	De Bilt $X > 0.95$	Rotterdam $X > 0.5$	De Bilt $X > 0.5$
$1 - H_G$				
Expert 1	0.809	0.812	0.894	0.897
Expert 2	0.889	0.892	0.766	0.769
Expert 3	0.960	0.963	0.853	0.856
Expert 4	0.746	0.769	0.960	0.963
Expert 5	0.832	0.812	0.979	0.982
Expert 6	0.733	0.736	0.730	0.733
Expert 7	0.787	0.790	0.730	0.733
Expert 8	0.809	0.812	0.894	0.897
$1 - H$				
Expert 1	0.822	0.825	0.900	0.903
Expert 2	0.895	0.899	0.784	0.787
Expert 3	0.962	0.965	0.862	0.865
Expert 4	0.767	0.787	0.962	0.965
Expert 5	0.843	0.825	0.980	0.983
Expert 6	0.756	0.759	0.753	0.756
Expert 7	0.802	0.805	0.753	0.756
Expert 8	0.822	0.825	0.900	0.903
Calibration Score:				
Equal Weighting	0.814	0.817	0.837	0.841
Performance-Based Weighting	0.960	0.963	0.979	0.982
Rank Correlation (Result):				
Equal Weighting	0.264	0.264	0.326	0.326
Performance-Based Weighting	0.578	0.578	0.608	0.608
Realisation	0.622	0.617	0.622	0.617

4.7 Chapter conclusions

In this chapter, we have presented the main considerations for eliciting multivariate uncertainty from experts. As shown, there are several important adjustments that are necessary when eliciting dependence given that many of the findings from expert judgement processes for univariate quantities are not readily applicable.

A first remark for concluding this chapter is that a few areas still lack insight to a considerable extent. For instance, we have discussed that the biases and heuristics which influence dependence assessments might be mitigated by training and knowledge structuring. In particular, experts' potential misinterpretations of dependence parameters need to be corrected and ways to do so might be informed by the educational literature on teaching concepts such as conditional and joint probabilities. Nevertheless, we need to acknowledge that experiences here might not be directly transferable to designing experts' training due to a different understanding of that of students and therefore further research in training design is necessary.

Further, more insight is needed on the exact triggers of the potential biases and their relative influence on judgements. It would be desirable for behavioural researchers to take a similar interest in this field as they do with the more common (typically univariate probability) heuristics and biases. This would allow developing the various (undeveloped) steps in the pre-elicitation phase, e.g. format choices.

In the elicitation phase, in particular the topic of structuring knowledge is identified as a key area for which further research is necessary. For instance, the graphical representation of BNs offers a way to incorporate qualitative dependence information. However issues still remain such as eliciting the structure of highly complex BNs as well as eliciting tail dependencies graphically. Therefore, again, we need to obtain more experiences for this part of the elicitation process.

Lastly, we have discussed that when combining assessments mathematically, more research is necessary for addressing some common desiderata for this step, such as performance-based as well as mathematically coherent aggregation.

Part III

Original Research

Chapter 5

Mapping conditional scenarios for knowledge structuring in (tail) dependence elicitation

This chapter[1] *addresses the challenge that guidance for eliciting dependence is sparse whereas particularly little research addresses the structuring of experts' knowledge about dependence relationships prior to a quantitative elicitation. However, such preparation is crucial for developing confidence in the resulting judgements, especially when assessing tail dependence. Therefore, we introduce a scenario mapping technique that structures experts' knowledge about (tail) dependence.Further, we show with an illustrative example how to elicit conditional scenarios that support assessing a quantitative model for the complex risks of the UK higher education sector.*

5.1 Chapter introduction

Structuring experts' knowledge about a joint distribution before its quantitative elicitation is one of the most neglected parts of research in formal elicitation processes for multivariate uncertainty. Nevertheless, it is essential for ensuring confidence in the elicitation, supporting transparency and reproducibility of the expert judgement study as well as mitigating experts' potential cognitive fallacies.

The few methods which specifically structure knowledge about joint distributions are part of hybrid approaches to dependence modelling by specifying qualitative relationships first, before assessing them probabilistically. These approaches comprise Knowledge Maps [201], Event and Fault Trees [84], Influence Diagrams [358, 202] and Bayesian Belief Networks [317]. While these methods enjoy popularity in the decision and risk analysis literature due to their convenient graphical representation, they are not suitable in various modelling contexts which might be of interest. For example, they do not address potential tail dependencies as these models in themselves do not capture main characteristics of systems where extreme value distributions emerge, such as underlying vicious cycles (or

[1] Based on: Werner, C., Bedford, T. and Quigley, J. (under review). Mapping Conditional Scenarios for Knowledge Structuring in (Tail) Dependence Elicitation, Journal of the Operational Research Society

reinforcing loops). Moreover, due to their hybrid nature, these models require a specific (often large) number of assessments for quantifying the underlying joint distribution.

In order to address knowledge structuring for joint distributions in a more flexible manner, i.e. for various models, including those suitable for capturing tail dependence, and separate from quantification (i.e. without the need to quantify all underlying relationships), we present our scenario mapping method. It builds on findings and approaches from *Probabilistic Risk Analysis* (PRA) [84] together with research in *risk perception* [338], *Problem Structuring* (PSM or "soft OR") [152, 347, 278], *Systems Thinking* (ST) [251, 255, 324, 67] and *Scenario Planning* (SP) [54]. By doing so, it is in line with the arguments in favour of "sense-making" through qualitative scenarios when modelling uncertainty [156], whereas such a representation of a simplified part of reality has also been termed "small world" [357].

The remainder of this chapter is as follows. First (in section 5.2), we define the variables of interest for a quantitative dependence elicitation and in section 5.3 present a way to elicit them for a common model that captures tail dependence if applicable. This shows the assessments that experts are required to make. In section 5.4, we propose our scenario definition and derive the main desiderata that determine the type of information we want to capture from experts. Next (in section 5.5), we outline the features of existing structuring/system analysis methods on which our scenario mapping method is based and introduce our method in section 5.6. After that we present in section 5.7 an illustrative example which shows how the method was applied to structure and model tail dependencies in the UK higher education sector. Finally, we conclude the chapter in section 5.8 with a reflective discussion on the method's achievements (for instance, mitigating common biases of dependence elicitation), the validation of the resulting scenario models and the method's current limitations.

5.2 Tail dependence models and the resulting variables of interest for the elicitation

Tail dependence can be modelled with a copula. For an introduction and discussion on the topic, see [208]. Recall, we can decompose any multivariate distribution function into its univariate margins and a copula. This can be reversed in order to construct new multivariate distribution functions with a given copula, so that a convenient modelling feature is the separate treatment of the marginal distributions and the dependence relationship. Various common parametric copulas can be grouped into classes. For instance, *Elliptical* copulas are radially symmetric, i.e. their upper and lower tail dependence is the same, whereas *Archimedean* copulas do not show this symmetry. This is an important modelling property as for the former, large losses always occur together with large gains which might not be a realistic dependence characteristic. For example, [99] show how copulas can be used to appropriately model asymmetric dependence of joint high default rates in a credit card portfolio.

Formally, lower tail dependence (which is of interest in our illustrative case study) for the distribution functions F_X and F_Y of random variables X and Y is defined as:

$$\lambda_L(X, Y) = \lim_{u \to 0} P(Y \leq F_Y^{-1}(u) | X \leq F_X^{-1}(u))$$

when a limit $\lambda_L \in [0,1]$ exists. Whenever $\lambda_L > 0$, X and Y are dependent in the lower tail whereas whenever $\lambda_L = 0$ they are tail independent. In other words, in the tail dependent case one is more likely to observe low values for Y given low values for X. From that, we can distinguish various copula types through their lower tail dependence coefficient $\lim_{u\to 0} \frac{C(u,u)}{u}$ (see [208] for its derivation) of which we make use in the elicitation (as shown in the next section).
In this chapter, we are focusing on bivariate dependence as this is already cognitively complex for experts to assess. Hence, our variables of interest are denoted as X and Y and correspond to a risk characteristic (monetary losses in the later case study) for which we are particularly interested in potential tail dependence. Note that similarly to the bivariate case for which we structure conditional scenarios corresponding to the elicitation of $P(Y \leq u^{th} quantile | X \leq u^{th} quantile)$, our method can be extended to structure conditional scenarios of larger conditioning sets, such as $P(Y \leq u^{th} quantile | X_1 \leq u^{th} quantile, X_2 \leq u^{th} quantile, \ldots, X_n \leq u^{th} quantile)$.

5.3 Tail dependence elicitation

While an elicitation can be designed in various ways, we briefly present a method that allows for the explicit consideration of tail dependencies as these are often of interest for a decision maker. The assessment of tail dependence highlights the need for a structuring process due its low intuitiveness. In particular, we re-emphasise the importance of a *formal* approach to structuring experts' knowledge as a necessary further development of previous approaches that evoke experts' narratives (or rationales) within univariate uncertainty elicitation [85, 87]. The process below is a pragmatic solution to eliciting dependence information for choosing a copula that represents an experts belief, yet it allows for distinguishing main parametric forms. Together with other elicitation methods it is discussed in [412] in more detail. Note, this is only a brief description focussing on the actual elicitation while neglecting elements of pre- and post-elicitation, such as training experts and aggregating judgements:

1. The marginal distributions are specified either through historical data or an expert judgement method for univariate quantities.

2. We elicit the conditional median in the form of $P(Y \leq 50^{th} quantile | X \leq 50^{th} quantile)$ for the variables of interest X and Y. This can be framed as:"Given that X is below your *median* for it, what is the probability that Y is also below your *median*?" (see Figure 5.1 on the left).

3. We elicit another quantile, one that corresponds to the (lower) distribution tail (e.g. the 5^{th}), i.e. $P(Y \leq 5^{th} quantile | X \leq 5^{th} quantile)$ which can be framed as:"Given that X is below your $5^{th} quantile$ for it, what is the probability that Y is also below your $5^{th} quantile$?" (see Figure 5.1 on the right).

4. With the assessments of 2.) and 3.) in place, we can compare an expert's judgements with different parametric copula forms. This is done by plotting the assessments against the converging conditional exceedance probabilities for selected parametric copulas simulated at the u^{th} quantile from 0 to 0.5 through the tail concentration function. Figure 5.2 shows the comparison of parametric copulas at the 50^{th} and 5^{th} quantile for a rank correlation of 0.3 and 0.7. The copula choices and rank correlations

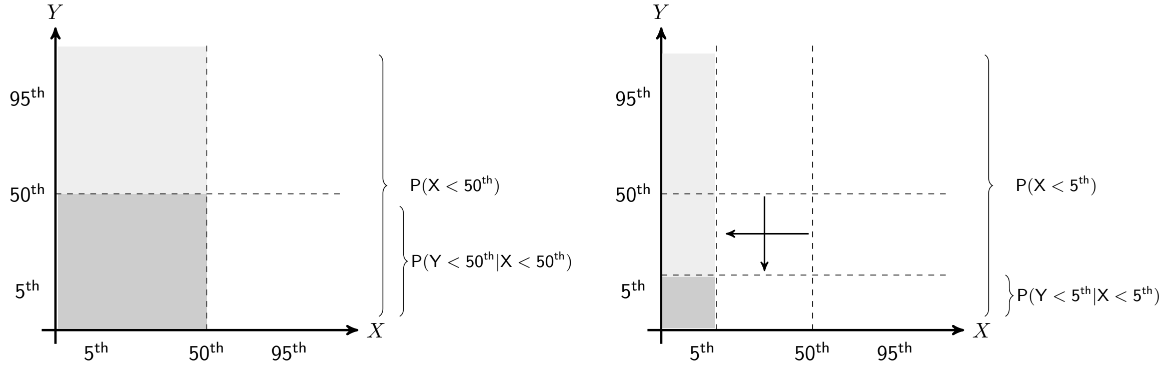


Figure 5.1: Schematic representation of eliciting the conditional 50^{th} and 5^{th} quantile.

can be varied for approximating the assessments better (see illustrative case-study).

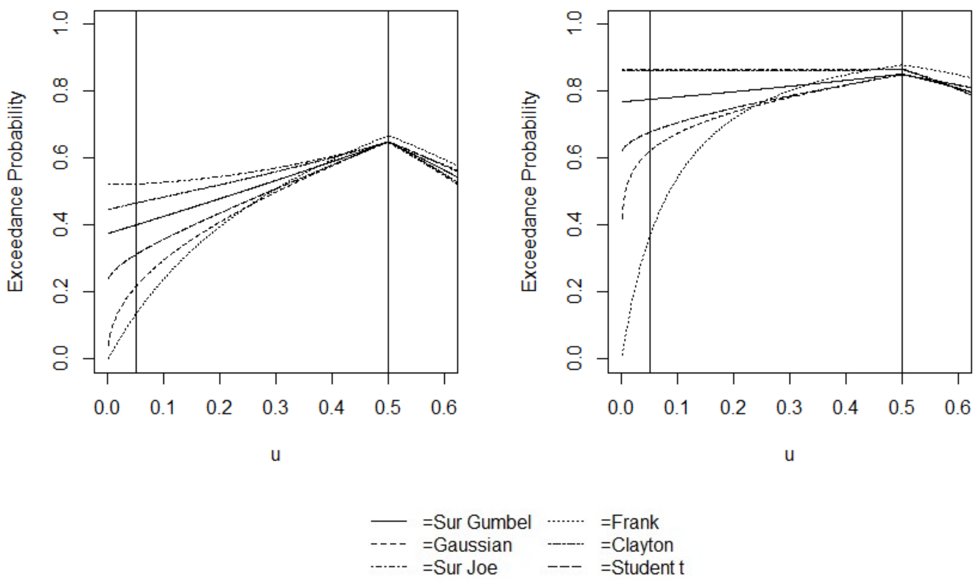


Figure 5.2: Convergence of exceedance probabilities for (selected) parametric copula families (rank correlation 0.3 (left) and 0.7 (right)).

5. With a first idea of which copula represents the expert's information reasonably well given a specific rank correlation, we can test the robustness of that choice, e.g. by "feeding back" the probabilities for non-elicited quantiles and check an expert's agreement for it.

Alternatively to 3.), we might elicit the conditional median for various quantiles. Thus, we elicit $P(Y < 50^{th}quantile|X < u^{th}quantile)$, solely varying the u^{th} quantile for X. Both ways of eliciting dependence information allow for deriving a copula that represents an expert's input satisfactorily and both alternatives relate to our scenario mapping method.

5.4 Scenario definition and desiderata of structuring methods

In this section, we propose a definition of the underlying scenarios that allow experts to express a rationale for their quantitative assessment. From this definition, we derive the main desiderata that a method, which structures experts' knowledge through conditional scenarios, should possess.
The term *scenario* is used differently in operational research, risk analysis and related fields. Purposes for which scenarios are formulated include forecasting [61], strategic planning [396], multi-criteria decision making [382], as well as decision analysis, e.g. through decision tree modelling [69, 157], and risk analysis by identifying hazards and vulnerabilities [222, 223]. Therefore, in the literature scenarios are used variously to support predictions of *what will happen*, exploration of *what can happen* (through hypothetical futures) and *what happens if* (in stress testing). The latter is also known as "wind-tunnelling" [396]. Depending on the purpose, scenarios can be qualitative, quantitative or both. For instance, the Scenario Planning literature spans various qualitative techniques to develop scenarios as narratives [15, 399], whereas in PRA quantitative scenarios are more common [223] (though there is also first a qualitative step followed by a quantification step).
While scenario thinking has been around most arguably since Plato in form of treatises on utopias and dystopias [54], [216] are regarded as pioneers in establishing it in the Scenario Planning literature. In other areas, such as PRA, Raiffa and his work on decision trees pioneers the use of scenarios [332]. The former regard a scenario as "a hypothetical sequence of events constructed for the purpose of focussing attention on causal processes and decision points". Similarly, [130] describe scenarios as "hypothetical sequences of events constructed as causal chains of argumentation for the purpose of focussing attention on alternative futures". Building onto these descriptions, we define scenarios with probability space notation [235] and cylinder set theory. It defines the larger world in which triggering events result in (potentially adverse) consequences. Experts state which elements the sample space contains. From that, the *state of the world* and its *future path* is defined as $\omega_i = x_0, x_1, \ldots, x_k$ with ω_i denoting one possible outcome from the entire sample space and x_0 being the *current* state of the world. Hence, x_1 is the state of the world *one* time unit into the future up to k time units for x_k. Further, ω_i is contained by the cylinder set of order k, $(x_0, x_1, \ldots, x_k | x'_0, \ldots, x_k \in \mathcal{A}_u)$ where x'_0 describes the *specified/known* current state and $\mathcal{A}_u$ represents a certain aspect that is defined for the state of the world at time k, in our case being below the u^{th} quantile. With a similar definition for another scenario set y in place, the set of (triggering) events is contained in today's state and determines alternative future states $x_1, y_1, \ldots, x_k, y_k$ all satisfying $\mathcal{A}_u$ for the state of the world at time k.
From that, we define a scenario as a *sequence that links triggering events to specified consequences (or final states) through intermediate conditions.* This definition builds on the literature of defining risk, decomposed into *hazards* (or threats), *vulnerabilities* and *outcomes*, the first being the "source of risk" [84]. This decomposition corresponds to the quantitative definition of risk in form of the triplet $\langle s_i, p_i, x_i \rangle$ [223] with s_i as the scenario, p_i the probability of scenario i, and x_i as its consequence. Further, it is in line in systems-based definitions of risk that highlight the importance of the concepts of vulnerability and resilience [182].

From our scenario definition, we identify two main desiderata which a method that structures scenarios for dependence elicitation should possess. First, it should only evoke scenarios that are *relevant* by resulting in a future consequence with the specified condition $\mathcal{A}_u$, i.e. below the specified u^{th} quantile. For the second desideratum, a method should identify the systemic impact of threats (through determining the interconnectivity of the intermediate conditions). In risk analysis, failures can be attributed very rarely to a single cause but rather chains of events that combine to produce the outcome. This is reflected in the alternative future states $x_1, y_1, \ldots, x_k, y_k$ that result from the set of triggering events. In particular, we are interested in the interconnectivity of intermediate conditions for understanding the difference between the unconditional and conditional distributions of X and Y.

5.5 Applicability and features of existing structuring methods

With respect to the main desiderata, we reviewed the literature on potentially applicable techniques, such as the ones used in Systems Thinking, Problem Structuring, PRA and Scenario Planning.
In order to identify the features of existing methods that we can use for our purposes, we examined how they perform in terms of three different properties, 1.) understanding severities in an anticipatory way, 2.) modelling the dynamic complexity of an underlying system and 3.) capturing how common causes propagate through systems. The first property relates to the desideratum of identifying relevant scenarios, i.e. regarding severities through a specific quantile. The other two properties relate to the second desideratum of understanding the impact on both, the unconditional and conditional distribution. Table 5.1 below shows the features of various methods.

Table 5.1: Features of existing qualitative structuring methods

	Technique	Reference	Technique models/captures...		
			anticipatory severities	dynamic complexity	system-propagation
PRA[a]	Fault Tree Analysis	Bedford and Cooke [84]	✓	✗	✗
	Event Tree Analysis	Bedford and Cooke [84]	✗	✗	✗
	FMEA	Ben-Daya [38]	✓	✗	✗
	HAZOP	Dunjó et al. [124]	✗	✗	✗
	Root Cause Analysis	Ben-Daya [38]	✓	✗	✗
	Influence Diagrams	Howard and Matheson [202]	✗	✓	✗
	Bow-Tie Approaches	e.g. Ale et al. [12]	✓	✗	✓
ST/PSM[b]	Causal Loop Diagrams	Morecroft [296]	✗	✓	✗
	Soft System Methodology	Checkland and Poulter [68]	✗	✓	✗
	Causal/Cognitive Mapping	Ackerman and Eden [4]	✗	✓	✗
	Knowledge Maps	Howard [201]	✗	✓	✗
SP[c]	Intuitive Logics	Van Der Heijden [396]	✗	✓	✗
	Backward-Logic SP	Wright and Goodwin [430]	✓	✓	✗
	Horizon Mission Methodology	Anderson [16]	✓	✗	✗
	Impact of Future Technologies	Bishop et al. [44]	✓	✗	✗
	Trend Extrapolation: Manoa	Curry and Schultz [103]	✗	✓	✓

[a] Probabilistic Risk Analysis; [b] Systems Thinking/Problem Structuring; [c] Scenario Planning

For the first property, applicable methods identify events that lead to a particular type and level of severity through *backwards logic*. This differs from forward logic approaches by not considering the possible development of scenarios from a fixed starting point (such as a threat) but by determining threats from given outcomes. In the SP literature, the terms *forecasting* and *backcasting* have been introduced for that [44] and [123] distinguish *exploratory* and *anticipatory* scenarios. [430] apply backwards logic to enhance traditional SP methods, such as *Intuitive Logics* [396], with a way to focus particularly on rare events with low predictability, i.e. ones which implicitly are assigned a probability close to zero. Their method is motivated by crisis management approaches which aim at preparing organizations for high impact/low probability catastrophes. An alternative SP method is the *Horizon Mission method* (HM) [16]. HM originated within NASA to support engineers' decision-making for research and development pathways as their scenarios often led to recommending incremental rather than breakthrough research. For enhancing scenarios, in the HM method engineers first envision a horizon mission (infeasible given today's technology) and then identify the new capabilities needed. The *Impact of Future Technologies* method serves the same purpose in IBM [44]. Similarly to SP's backwards logic, in PRA in particular *Fault Tree* and *Root Cause* analysis methods [84] investigate how specific failure events can occur. In PRA however, scenarios are captured by event sequences rather than narrative SP approaches.

Regarding the second property, we observe from Table 5.1 that most PRA and many SP methods do not allow for modelling the dynamic complexity of a system. Yet, capturing this is important for experts to understand how interdependent components of systems interact over time and across different systems. [314] emphasises the need for traditional PRA methods, which often model engineered systems, to apply more holistic forms of analysis in order to address the challenge of more complex risks. Likewise, [416] examines the deficiency of PRA methods with regard to systems thinking and [6, 5] highlight the need for a comprehensive, holistic and systemic approach to risk analysis to account for "risk systemicity". It refers to the idea that "the effect of two risks might be more than the sum of the two individual effects thus reflecting systemicity" [420]. While there is no agreement on the definition of a dynamic and complex system (see [247] for a discussion), a commonly mentioned characteristic is *non-linearity* due to reinforcing (or vicious) feedback loops. Their identification is hence crucial when analysing a system. Various methods, summarised under the umbrella of systems thinking, together with "soft OR"/PSMs are analytical approaches for understanding a system holistically rather than through its separate parts. Common methods that identify feedback loops by graphical representation of influences are *Causal Loop Diagrams* [296] and *Cognitive Maps* [328, 4]. Both methods allow a participatory approach to modelling complex problems [102] and have been used in mixed-method approaches together with cybernetics for analysing structures of systems [261].

Related to understanding dynamic and complex systems is the assessment of common cause propagation through *distinct* systems. This is the third property and it is not clear how most methods distinguish between (what is perceived/defined as) different systems (Table 5.1). It draws on a fundamental aspect of systems thinking/PSM, the idea of a *system boundary* [71]. For ST/PSMs, identifying what lies inside a system and hence which factors are included in a model requires experience and judgement, which is why the modelling process is usually iterative and circular rather than linear [296]. Once a first model version

is constructed, experts might refine the model by re-assessing which factors to include (or exclude) based on a reflective understanding. Emphasising the importance of the modelling process and the judgemental nature of a model boundary is in agreement with [277] who discusses the definition of a system's boundary from the viewpoint of *Critical System Heuristics* [392]. This is a framework for participatory and reflective practice on boundary judgements which requires system thinkers to consider critically what a system includes and to examine it from multiple perspectives through a checklist/question-based approach. More generally, for decision models this is related to the issue of infinite regress when modelling [154], and for which [323] introduced the term of a *requisite model* that results from a circular and interactive modelling process.
In order to understand better the difference of the unconditional and conditional sub-systems that determine our distributions, it is important to derive the overall system boundaries as well as the boundaries of the sub-systems. Therefore, we use a technique from PRA, a *bow-tie* model for which the idea of system propagation of events can be illustrated by the bow-tie logic [12] in the sense of assessing how a hazard is caused by threats and at the same time is the cause of a consequence.

5.6 Mapping conditional scenarios

After having identified some useful features of existing methods that comply with our desiderata, we introduce our method which synthesizes specific elements of some of the methods discussed above for our purpose of structuring knowledge for (tail) dependence elicitation.

5.6.1 Overview of the mapping process

For mapping conditional scenarios, we propose an iterative process which is facilitated with the experts individually, and scenarios are shared only as a final step of knowledge sharing. Figure 5.3 provides an overview of the overall mapping process. In the *first step*, the facilitator ensures the expert's familiarity with the set-up, i.e. the different steps together with the tasks at each of them. Further, common expert judgement formalities, such as confidentiality of personal information and usage of mapping session results, are clarified.
In the *second step*, the expert is introduced to the first variable of interest, X, which concerns the unconditional distribution. Further, we clarify the specified time-frame in which scenarios lie, e.g. the next five years (we regard our method more suitable for shorter rather than longer time frames due to the focus on tail dependencies which might be not recognisable for events too far into the future). Then, the expert is presented with the final condition of the required scenarios, $\mathcal{A}_u$, which states that the unconditional distribution is above (or below) a certain quantile.
In *step three*, the expert is given time to brainstorm and note the different *reasons* why the variable of interest lies above (or below) the specified quantile. We emphasise reason here as this is an unstructured part of the process and an expert might express these in own words. Further, note that this step employs backward logic by reasoning from a specific consequence to potential causes.

Fourth, the facilitator together with the expert classifies the reasons, identified in the previous step, into two different event types according to their causal logic. For this, note that an expert's dependence assessment might be based on

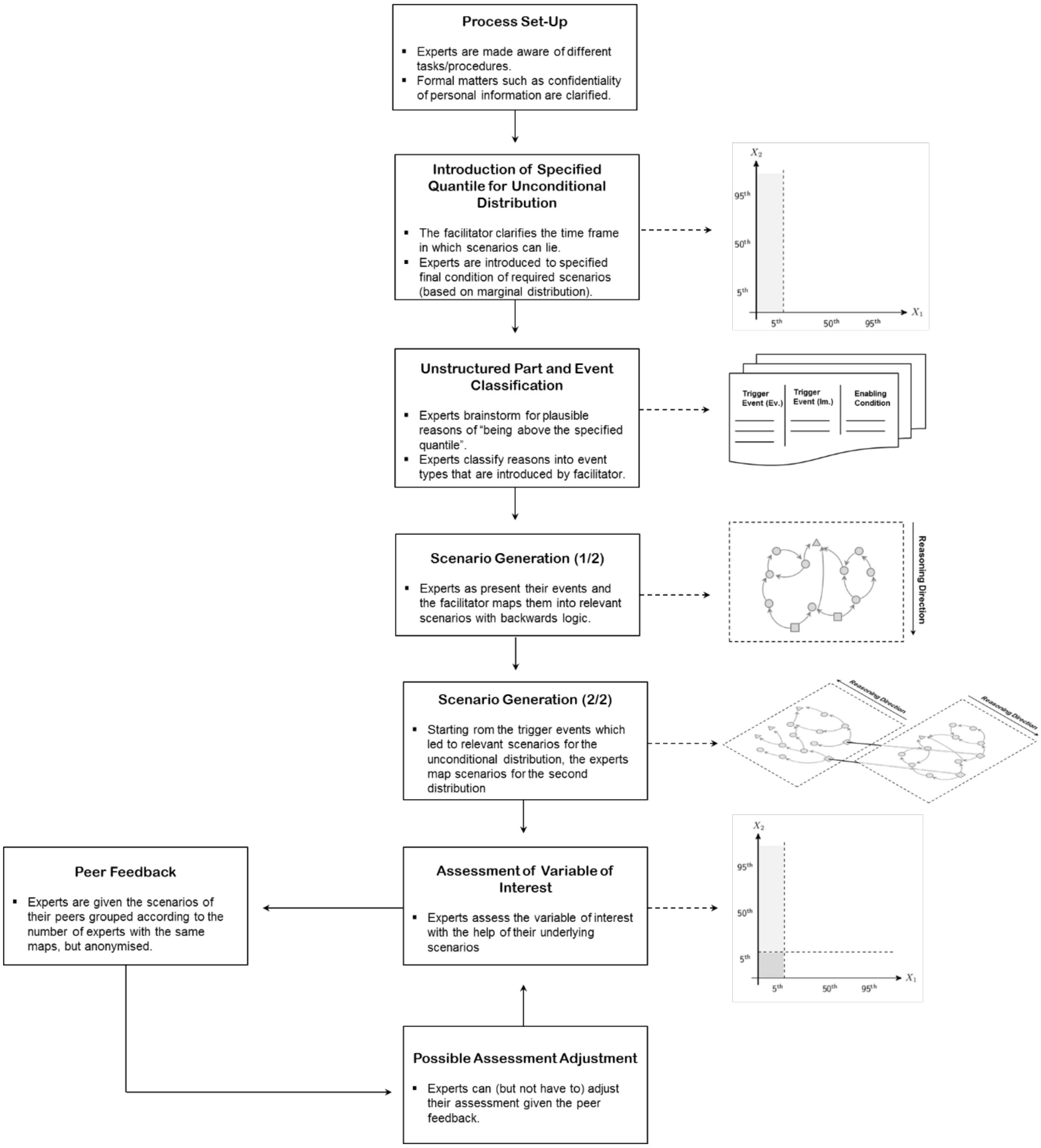


Figure 5.3: Overview of Scenario Mapping Process.

various sources of information. However, we regard in particular mental models about causal relationships as main determinants for their assessments. This is in agreement with the behavioural judgement and decision making literature which proposes that people believe that most events have causes (rather than happening due to pure randomness) and further, that they use systematic rules to derive causal inferences [134]. Yet, as these rules are often incomplete or imperfect, we adopt a probabilistic view on causation [384]. Probable causes are invariably linked to linguistic expressions of causal relationships, such as *cause*, *enable*, *prevent* [369]. In accordance with the former two, the event types used for classification are:

Trigger Event (immediate) A trigger event is a plausible initiator of a scenario contained in the current state of the world and it may or may not be (fully) observable. For clarification we might add words like "start", "outbreak", "attack", "eruption", "shock" etc., e.g. "disease outbreak", "terrorist attack", "volcanic eruption", "oil price shock". For observable trigger events, it is possible to neglect any preceding events as we condition on the them knowingly. However, for trigger events that are only partly observable, we need to include immediate preceding events (which led to the trigger event) for ensuring a richer set of scenarios. Suppose for instance that an expert identified "oil price shock" as a trigger event. In this case, the facilitator might have to clarify whether this is due to geo-political risks involving OPEC countries or due to a change in usage of alternative energy sources, as for both versions, very different scenarios unfold in the future.
Another remark on correctly identifying a trigger event is that experts cannot position a trigger event in the future. If this occurs, it is important that the facilitator supports the expert in re-considering why such an event will happen in future in order to identify its corresponding trigger event in the current state of the world.

Trigger Event (evolving) Evolving trigger events are similar to immediate ones, a difference is however that they capture a longer development of an event which can be seen as an initial cause. For these, it is possible to insert words like "development" in a sensible manner, e.g. "development of (long lasting) rain showers" as a trigger event for a certain flood severity.

Enabling Conditions Complementary to both types of trigger events, an expert should also identify enabling conditions. These follow from the trigger events and capture evolving trends in a system by constituting the conditions that need to be in place for a trigger event to reach a certain consequence. They might labelled as "higher"/"lower", e.g. "lower economic growth", "higher risk of infection", "higher migration" for clarification. Enabling conditions are directly related to the ideas of system vulnerability and resilience in PRA [182].
The categorisation into the different event types is important for elaborating conditional scenarios from the current unconditional ones as shown in the subsequent steps as well as their comparison across experts. Note that risk perception research [338] identifies similar event types whereas trigger events are called *hazard events* which result in *secondary effects* that affect more people than the ones affected by the original trigger event, e.g. economic impacts or social and political pressures. Thus, our definition of event types is not only justified from the natural language of probabilistic causation but should also correspond to

people's experience with unfolding risks.
In the *fifth step* of the process, experts determine the links between trigger events and enabling conditions in order to map out the path to the specified final condition. This part makes use of cognitive mapping and causal loop diagram methods as we can identify feedback loops and the overall interconnectivity of the events. Links are set according to an expert's belief while we omit the assignment of polarities as these are already captured in the enabling conditions. In fact, it serves as a robustness test for the enabling conditions as experts might change their labelling based on the links. Testing robustness in graphical models is commonly embedded in the modelling process [252, 283]. This part can be supported by mapping software (such as the one developed within this project [411]) to allow for a direct visualisation of the unfolding scenarios. Once an expert is satisfied with the resulting set of scenarios, the final picture can be captured.
Next (in *step six*), the expert maps how (and if at all) the triggering events that are relevant for the unconditional distribution propagate in the conditional one. For that, we simply import the previous trigger events (e.g. automatically with the developed software) onto a blank screen which represents the conditional scenario space of Y. Then, an expert identifies the relevant enabling conditions (in forward logic) from the imported trigger events with respect to the specified time frame. This part of the process corresponds to the "bow-tie" element discussed earlier.
Now, an expert has a thorough understanding of how both variables of interest are affected by the same events and we can proceed with the quantitative assessment in the form introduced earlier.
The *last step* of the process allows for (anonymously) sharing experts' scenarios. By facilitating the structuring/scenario mapping sessions with experts first individually, before providing each experts with the scenarios of other experts, this process shows similarities with elicitation processes, such as Delphi [351]. A difference is however that we do not seek consensus. Rather, our process builds on findings that the accuracy of individual assessments improves upon receiving feedback about other people's judgement [425, 433]. An expert judgement process for univariate quantities which encourages a second round of assessments after individual assessments have been shared and discussed with other experts, is the IDEA protocol [187].
The overall process of mapping scenarios is repeated for all quantiles of interest.

5.7 An illustrative example: Managing risk in the UK Higher Education sector

The higher education (HE) sector in the United Kingdom (UK) has been frequently in focus for applying operational research techniques, mainly for problems of performance measurement and resource allocation [210, 109, 265]. Less experience is available for assessing and managing risk in this sector. This is the case, even though the general management of HE in the UK has been studied and is well-understood [195, 194]. As factors that might pose a risk to an HE institution, [195] outlines variable tuition fees, which increase competition and change students' expectations, the increased exposure to and reliance on overseas markets, large investments in infrastructures to facilitate institutional expansion as well as potential loss of market share due to new technologies. Likewise, [14] view tuition fee income as a main driver for internationalisation whereas some uncertainties affecting its development are political realities together with

national security concerns, such as changing visa requirements in the face of international terrorism, government policies influencing the cost of studies, the potential expansion of domestic capacity for sending countries to meet education demands, the increasing importance of English as lingua franca, the alignment and accreditation of degrees and the future impact of e-learning offerings.
In order to better understand such complex uncertainties, our method has been applied to support decision-makers in charge of managing the postgraduate taught course portfolio at the authors' home institution and department (for the full details of the case-study, see [411]). Our method was used to map the interdependence of future scenarios that might affect the tuition fee income of the established *MSc Business Analysis and Consulting* (BAC) course and the newly introduced *MSc Data Science* course within the next four years. After that, the dependence between the courses was assessed quantitatively. We applied our method with five experts who are in charge of managing postgraduate taught courses (and implementing the new course) and our variables of interest are defined as the generated income from each course through tuition fees. After having specified the marginal distributions for our variables of interest with a structured expert judgement process for univariate uncertainties [79], we introduced our experts to the elicitation questions for the dependence assessments[2]:

1. *"Given that the generated income of the MSc BAC is below its median in the academic year 2020/21, what is the probability that the MSc Data Science is also below its median?"*

2. *"Given that the generated income of the MSc BAC is below its* 5^{th} *quantile in the academic year 2020/21, what is the probability that the MSc Data Science is also below its* 5^{th} *quantile?"*

Then, we started structuring the scenarios with each expert individually as described in the previous section, first the unconditional and conditional scenarios for being below the median, then the same process for being lower than the 5^{th} quantile. Figures 5.5 and 5.6 show the scenarios of one of the experts that were elaborated within approximately an hour.
As we can see in Figures 5.5 and 5.6, for the 50^{th} quantile the expert believes that most trigger events of the unconditional scenarios will affect the conditional ones similarly. However, a slight difference relates to the future demand of MSc Data Science graduates (in particular) due to a more important data science market. For the 5^{th} quantile, this expert considered the trigger events that result from the unconditional scenarios to have the same impact on the conditional ones. In other words, once the income generated from tuition fees by the MSc BAC is below a certain threshold, the scenarios must be relevant on a more global level, so that the MSc Data Science will be affected similarly. The idea of such tail dependence corresponds to most experts' scenarios and assessments as they view the risk of being below the 5^{th} quantile as a result of events that affect the UK HE sector more broadly rather than the different courses individually. Due to such similar beliefs, no expert changed her/his assessment after reviewing the other experts' scenarios in the last round of the process (Figure 5.3). As a result, when aggregating the experts' assessments, both, an equal weighting combination as well as a performance-based one (based on the statistical accuracy of experts for the marginal distributions), indicate the fit of a tail dependent copula. Figure 5.4 shows how both combinations fit well with a Clayton and

[2]The experts were given the corresponding monetary values for the specific quantiles.

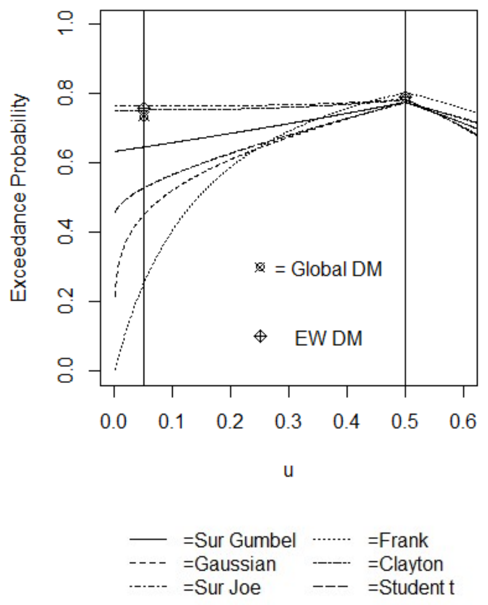


Figure 5.4: Fitting parametric copula forms to combined assessments (rank correlation=0.55).

Survival Joe copula of rank correlation of 0.55). Note that this is a pragmatic way of combining experts' assessments and the performance-based aggregation of dependence assessment is a topic of ongoing research [411].

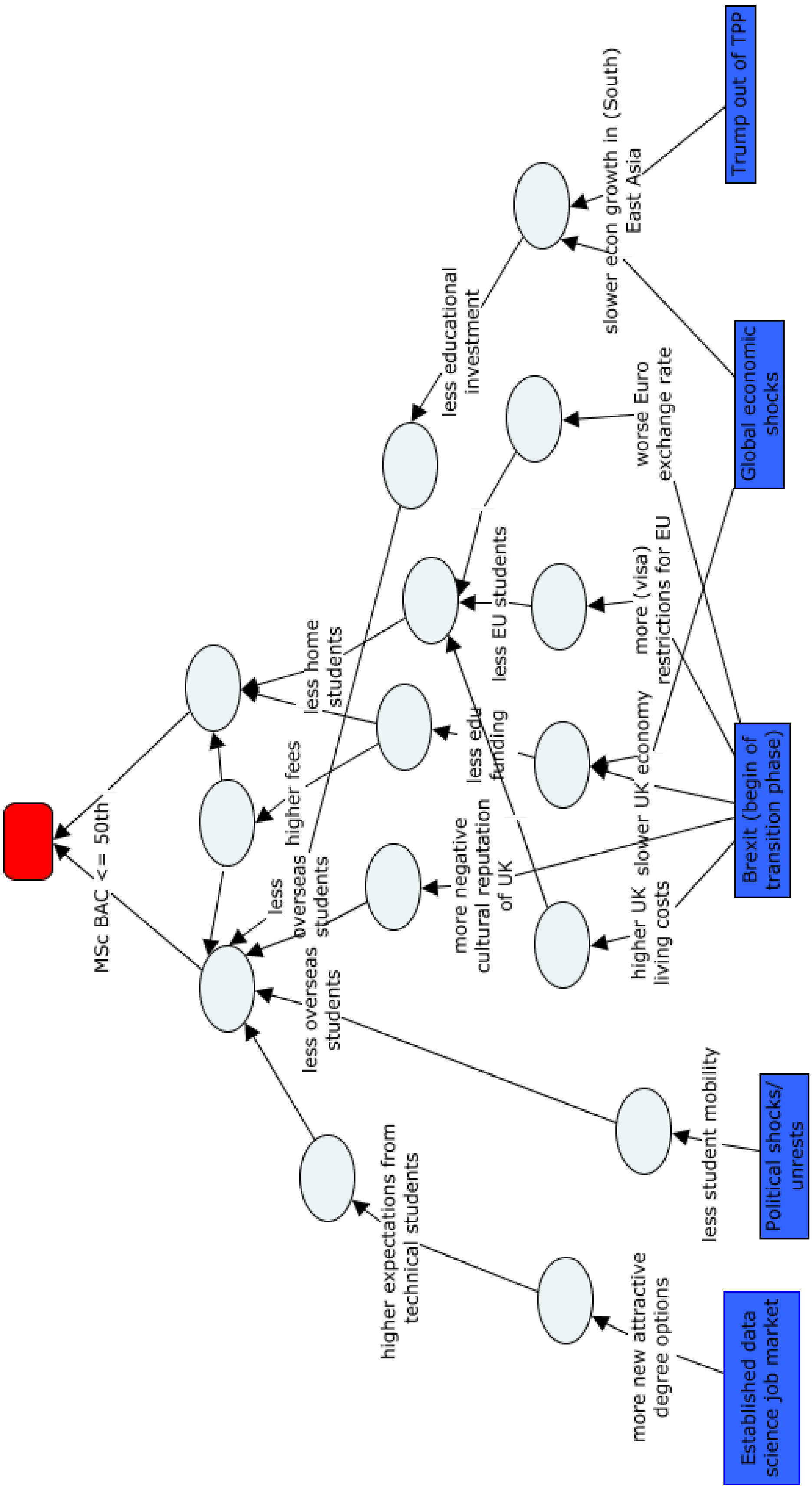


Figure 5.5: Unconditional scenarios for 50th quantile.

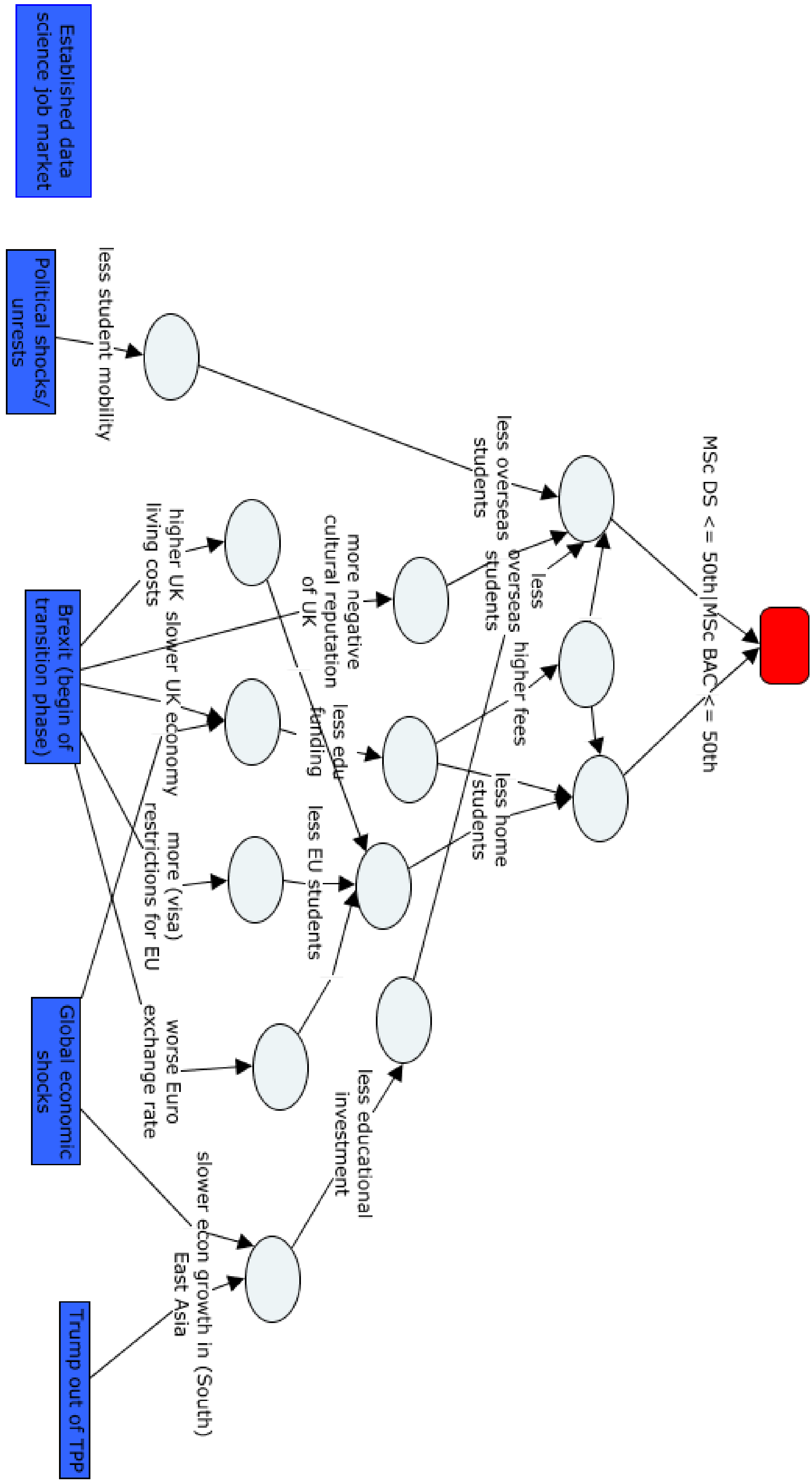


Figure 5.6: Conditional scenarios for 50th quantile.

5.8 Discussion and chapter conclusions

In this chapter, we have proposed a scenario mapping method for structuring experts' knowledge about dependence relationships due to the high cognitive complexity that experts face when assessing joint distributions. The aim is to offer a method that can be used together with various dependence models, including ones for tail dependence.
Structured approaches for supporting experts in expressing their rationale have already been recommended for eliciting univariate uncertainty [85, 87] and (as mentioned in the introduction) they (1) increase confidence for the quantitative assessment, (2) mitigate potential cognitive fallacies and (3) support transparency as well as reproducibility of the expert judgement study. In this conclusion we briefly discuss how our method achieves the above.
With respect to (1), increasing confidence for the later quantitative assessment, the experts' feedback showed that they regard our method as helpful for obtaining a better understanding of the dependence relationships and hence have more confidence to assess dependence quantitatively after using it. This is due to the possibility to express their thinking in natural language whereas the classification into the different event types is regarded as intuitive. Further, the decomposition of the dependence relationships, which allows for seeing how the trigger events that are elaborated in backwards logic (for the unconditional distribution) are relevant when thinking about the conditional distribution, is appreciated by the experts as a structured way to think about complex scenarios and the influencing factors of dependence relationships.
With regards to (2), the mitigation of biases, we first discuss briefly how our method addresses two main cognitive fallacies of (conditional) dependence assessment (see [415] for an overview) before we discuss mitigating fallacies when assessing extreme probabilities, such as in tail dependence assessment, with our method.
When eliciting dependence in conditional form, a common fallacy is the *confusion of the inverse* [193, 106]. A conditional probability, such as $P(X|Y)$, is confused with $P(Y|X)$. One explanation states that this bias is related to the better known *representativeness heuristic* [217] as people assess how similar or representative characteristics of X are for Y (rather than the conditional probability). This can lead to getting confused between the inverses. Another explanation refers to the perception of a causal relationship between X and Y. People might have incorrect preconceptions or observations on information that is important for causal inferences, such as the temporal order of events [29, 134], which can enhance the confusion. Our method decomposes the underlying factors of both variables of interest and by doing so, it challenges the representativeness heuristic through alternative scenarios and clarifies the perceived causal and temporal orders.
In other cases, people *confuse conditional and joint probabilities* [134]. The main candidate explanations attribute this confusion to linguistic ambiguities introduced by describing conditional dependence relationships through "given that" or "if" [134]. Our method - and even though this is only anecdotal evidence - helped one of the experts to clarify the meaning of "given that" (i.e. that we elicit conditional probabilities in the case-study) when thinking out loud about the probability of a "perfect storm". She first thought about the probability of a "perfect storm" event happening as being extremely small before realising that this is not applicable for conditional assessments and the conditional probability

is higher after reflecting on her scenario map (see [314] on the idea of a "perfect storm" in risk analysis).

When eliciting tail dependencies (i.e. extreme events), the applied backwards logic, which is used to structure scenarios for the unconditional distribution, allows for mitigating some main issues that can arise due to cognitive complexity. While the advantages of applying backwards logic are similarly discussed by [430] (in a different decision analysis context in which extreme scenarios are important), for us it is specifically useful as it addresses the following cognitive challenges.

The first is *frame blindness*, i.e. forecasting the wrong event. With backwards logic, experts elaborate scenarios by starting from the final condition, so that they will not explore scenarios that are not relevant. However, a potential difficulty might be that backwards reasoning can be unintuitive [192].

Another challenge is that low probability events (in the tail of a distribution) by definition lack a reference class for similar events. Therefore, we cannot assess whether experts have well-calibrated assessments or are (for example) frequently overconfident. Through outlining experts' mental models on underlying causal processes we can however go beyond the historical data. A main part of that is to see which experts' scenarios are (mostly) coherent. Through backwards logic experts can explore how final conditions can (or cannot) be reached from the current state of the world and if not discard them. A potential issue with relying on causal models is that a plausible representation of causal processes might increase the associated likelihood for a scenario. This has been called *simulation heuristic* [220]. Therefore, we need to ensure that a rich set of scenarios is developed and these are shared, so that alternative scenarios are considered and challenge existing presumptions.

The graphical representation of the experts' rationales is also important for (3), ensuring transparency of the quantitative assessment results. Experts' scenario maps can be associated with later assessments which makes the outcome of the elicitation more transparent for anyone not involved in it. However, not only transparency is of concern for anyone not involved in an elicitation, but they also need to have confidence that the scenario maps are the correct qualitative models for the assessments. Thus, we need to ensure the validity of the scenarios. To do so, we clarify that validity does not mean that the conditional scenarios should be compared against the real-world events that have happened by the time this can be done. Rather than basing validation on *prediction*, we emphasize *explanation*. This has been called *white box* validation in the operational research literature [231] and can be achieved by a process by which experts (not involved model development) determine, with some level of confidence, whether a model is suitable for decision-making [167, 166]. This is usually determined by common-sense and comparison to the real world [167].

Chapter 6

Sequential Refined Partitioning for probabilistic dependence assessment

This chapter[1] *addresses the challenge that, whenever relevant data for quantifying and modelling dependence between uncertain variables is lacking so that expert judgement might be sought to assess a joint distribution, without restrictive parametric assumptions, a model is underspecified while experts' assessments may also be easily overspecified, particularly when making several, detailed assessments. Underspecification means that we cannot determine a unique distribution as various alternatives are compatible with the given (partial) information. For overspecification, experts' assessments on a distribution's related parts are contradictory and infeasible. The sequential refined partitioning method addresses under- and overspecification whilst allowing for flexibility about which part of a joint distribution is assessed and its level of detail. Potential overspecification is avoided by ensuring low cognitive complexity for experts through eliciting single conditioning sets and by offering feasible assessment ranges. The feasible range of any (sequential) assessment can be derived by solving a linear programming problem. Underspecification is addressed by modelling the density of directly and indirectly assessed distribution parts as minimally informative given their constraints. Hence, our method allows for modelling the whole distribution feasibly and in accordance with experts' information. A non-parametric way of assessing and modelling dependence flexibly in such detail has not been presented in the expert judgement literature for probabilistic dependence models so far. We provide an example of assessing terrorism risk in insurance underwriting.*

6.1 Chapter introduction

In this chapter, we address the problem that experts can only ever assess certain aspects of a joint distribution whereas a decision-maker might desire these assessments to be made at a detailed level. The former implies that we have a partially unknown distribution for which various alternatives fit the given in-

[1]Based on: Werner, C., Bedford, T. and Quigley, J. (2018). Sequential Refined Partitioning for probabilistic dependence assessment, Risk Analysis, doi: 10.1111/risa.13162

formation. This is known as model *underspecification*. More specifically, we are only ever given the probability mass (or density) within some distribution parts, either through their direct assessment or (in parts which are never assessed) through the indirect result of these parts together with related assessed parts having to comply with the marginals. However, we can model these probability masses in various forms which all have the right amount (i.e. are feasible). Of course, we might elicit additional information from experts to distinguish between distributions, yet we need to acknowledge the impossibility of ever eliciting enough information to single out a unique distribution. This is unless adopting a low-dimensional parametric model early on in the modelling process[2]. Such parametric assumptions nevertheless restrict the obtained knowledge on dependencies and we might miss potentially important model aspects, such as random variables' behaviour in the extreme parts (tails) of a joint distribution. Hence, it is often desirable to avoid distributional assumptions which might exclude phenomena that the expert thinks are important.

Within a non-parametric setting, an elicitation should capture detailed distribution features, e.g. the probability mass within narrowly defined parts of the distribution, such as the tails to determine tail dependence, as they result in a more specific distribution, thus making the model more valuable for a decision-maker. Nevertheless, while detailed assessments might be desired by decision-makers, they increase the experts' cognitive burden, potentially resulting in inconsistent and infeasible assessments. This is termed *overspecification*, the second modelling challenge that we encounter[3].

As a non-parametric approach, addressing under- and overspecification, we present the *sequential refined partitioning* (SRP) method for assessments that can be made to any level of detail for any part of a joint distribution. In the SRP method, we address overspecification through an elicitation procedure which never increases the conditioning set to more than one condition and thus maintains a low cognitive complexity. Further, the procedure ensures consistent and feasible assessments through explicit guidance on assessments' feasibility ranges. Underspecifcation is dealt with by allowing the expert to specify as much detail as is desired and by then determining the density form of directly and indirectly assessed parts of the distribution through the unique copula distribution that is minimally informative with respect to the independent copula and that corresponds to the elicited information. Hence, we do not introduce any unspecified assumptions. This ensures that the whole distribution is in agreement with the experts belief. The minimum information approach offers a recognised approach to incomplete knowledge[206]. Further, it allows us to stop the elicitation process at any time and still derive a unique distribution (in contrast to common probabilistic dependence models for which a full conditional probability table is required, e.g. Bayesian (Belief) nets (BNs) [317]). In the context of dependence elicitation, minimum information methods (and related approaches) have been used before, for instance in probabilistic inversion (PI) methods [35, 239, 273, 80], vine-copula quantification [32, 31], or as well joint distributions more generally within decision analysis contexts [286, 285, 1, 42]. However, these previous methods do not consider flexible nor detailed dependence assessments and their impact on potential overspecification of experts' judgements and on the minimum

[2]Under low-dimensional parametric assumptions, it suffices to assess a chosen form's main parameters. E.g. eliciting the mean vector and the covariance matrix quantifies a multivariate Gaussian distribution sufficiently.

[3]Overspecification can also occur with parametric models, e.g. if assessed covariances jointly do not result in a positive definite matrix.

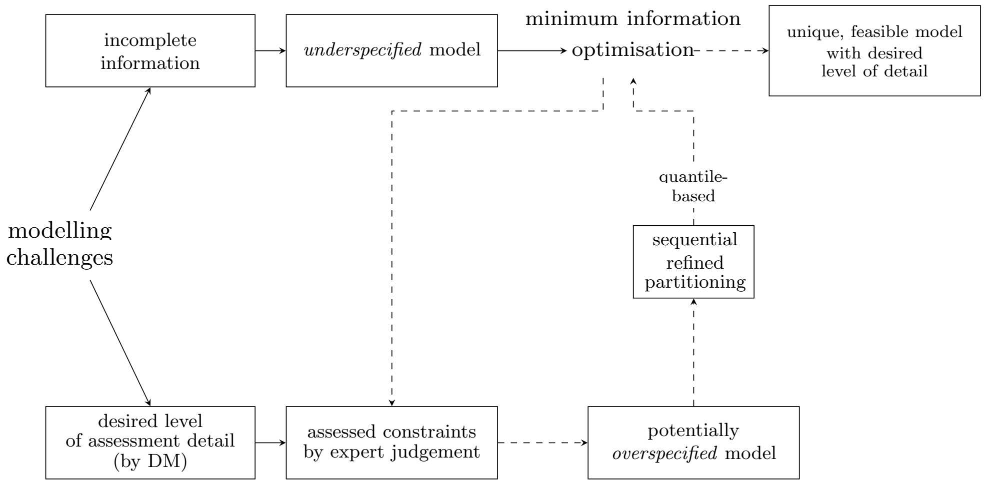


Figure 6.1: Modelling context of the SRP method.

information solution to underspecification. For example, [35] explicitly provide guidance on feasibility constraints. Yet, they consider dependence elicitation at a rather broad level, eliciting only a small number of assessments. This restricts the information to be obtained already early on in the modelling process and thus neglects focusing on specific parts of a distribution more exclusively. The SRP method's contribution is therefore that we provide an elicitation procedure to assess any part of a distribution to any desired level of detail while maintaining low cognitive complexity[4] and avoiding infeasible expert judgements. Similarly, the SRP method's approach to underspecification is more detailed than in previous research. Figure 6.1 illustrates our method's modelling context schematically.

In the upper part, we observe that incomplete knowledge leads inevitably to an underspecified model. This is solved by a minimum information approach. In order to derive a model that is valuable for a decision-maker, the modelling process deviates along the dashed lines to the lower part. Here, the constraints of the minimum information problem determined by the experts' judgements are assessed as detailed as desired. As these might be overspecified, we the use an elicitation process that leads to feasible assessments. In the remainder of this chapter, this is presented in section 6.2, introducing the elicitation procedure, and section 6.3, outlining the optimisation problem. Section 6.4 shows how our method has been used in an insurance underwriting risk assessment of political violence/terrorism in which a detailed and flexible method is of particular interest for stress-testing a model. Finally, section 6.5 concludes the chapter.

6.2 Eliciting detailed dependence information feasibly and consistently through *sequential refined partitioning*

In this section, we introduce our sequential elicitation procedure which addresses the potential issue of overspecification by providing explicit guidance on making feasible and consistent assessments. In the expert judgement literature, several approaches to ensuring feasibility and consistency are proposed, each with different implications on the robustness of the final assessment result. As such,

[4] As such, it also contributes to expert judgement methods for dependence in which increasing conditioning sets pose a concern (see [412] for a discussion).

some methods (always) allow for an assessment within the elicited forms' standard ranges (for correlation coefficients $\in [-1, 1]$ and for conditional and joint probabilities $\in [0, 1]$). However, this might jeopardise experts' commitment and confidence in the elicitation method if assessments are adjusted afterwards (for ensuring feasibility). While other methods do not modify assessments, they might increase experts' cognitive complexity. For instance, by limiting assessment ranges (away from the aforementioned standard ones), or by imposing unrealistic assumptions onto experts' understanding of elicited forms, e.g. when eliciting conditional judgements with large conditioning sets. For the latter, we might expect an expert to include and equally consider all the information given by a large conditioning set so that common cognitive fallacies, such as the conjunction fallacy and its conditional version (see [415] for an overview on heuristic and biases in dependence assessment), should be (ideally) avoided and hence feasibility is given. Yet, this might not be guaranteed.
In our method, we do not impose such unrealistic assumptions on experts' cognitive capabilities, nor do we modify assessments after they have been given. Rather, we only ever elicit single conditioning sets and give guidance on possible feasible assessment ranges. This includes not only providing the corresponding upper and lower bounds but also explaining their interpretation.
Mathematically, the feasibility range for any sequential assessment procedure is derived by solving a linear programming (LP) problem (see [398] for an introduction to LP). The number of constraints is restricted to a maximum of nine, irrespective of the number of elicitations. In the remainder of this section, we first present the general set-up together with the relevant proofs before we outline some specific elicitation sequences, which we regard as of interest for several practical applications.

6.2.1 General set-up of sequentially refined partitioning

We shall start by introducing some definitions. The unit square is here defined as the product of $(0, 1] \times (0, 1]$. Given values $u_0 = 0 < u_1 < \cdots < u_n < 1 = u_{n+1}$, and $v_0 = 0 < v_1 < \cdots < v_m < 1 = v_{m+1}$, we define the associated *quantile partition* of the unit square as the set of rectangles of the form $(u_i, u_{i+1}] \times (v_i, v_{i+1}]$. We call this set of rectangles $QP(u, v)$.
Given (p, q) with p different to the u_i and q different to the v_j, the (p, q)-refinement of $QP(u, v)$, denoted $QP(u, v; p, q)$, is the quantile partition obtained by including p and q in the values for u and v respectively. All rectangles in the old partition are either in the new partition or are a union of two or four rectangles of the old partition. Figure 6.2 shows two partitioned example distributions which result from any number of previously elicited quantiles (solid lines) in addition to new ones (dashed lines).

A probability distribution on a quantile partition $QP(u, v)$ simply assigns a probability value to each rectangle of the quantile partition. A (p, q)-*refinement* of such a probability distribution is a probability distribution on $QP(u, v; p, q)$ such that the probability of a rectangle in $QP(u, v)$ is either the same as it is in the (p, q)-refinement of $QP(u, v)$, or it equals the sum of the probabilities of the rectangles that make it up.
A *merging* of a quantile partition $QP(u; v)$ is obtained by merging together some of the partition rectangles in such a way that we still have a quantile partition. This can also be obtained by taking a subsequence of the u's and v's and building the corresponding quantile partition. A *merged probability distribution* on the refined quantile partition is obtained by adding together the probabilities of the

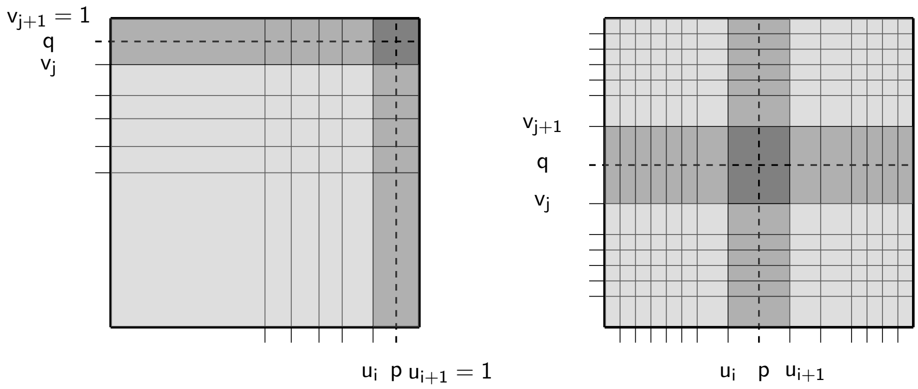


Figure 6.2: Partition example $QP(\tilde{u}, \tilde{v}; p, q)$ with solid lines for previously elicited quantiles and dashed lines for new ones.

rectangles in each refined rectangle.
We always work with *discrete copula distributions*, which are probability distributions on a quantile partition that have the additional property that (for any k) the sum of probabilities of rectangles $(u_i, u_{i+1}] \times (v_i, v_{i+1}]$ with $u_{i+1} \leq u_k$ is equal to u_k, and similarly, the sum of all probabilities of rectangles $(u_i, u_{i+1}] \times (v_i, v_{i+1}]$ with $v_{i+1} \leq v_k$ is equal to v_k. For a general introduction to copula theory, see [302],[208] and [125]. However, note that most theory is on continuous copulas with marginals being continuous uniform distributions. For an overview on elicitation methods for copulas, see [412].

Proposition 1. *Suppose we are given values $u_0 = 0 < u_1 < \cdots < u_n < 1 = u_{n+1}$, and $v_0 = 0 < v_1 < \cdots < v_m < 1 = v_{m+1}$ (where $n, m > 0$), $0 < p, q < 1$, with p different to the u_i and q different to the v_j. Then a copula distribution on $QP(u, v)$ can be refined to a copula distribution on $QP(u, v; p, q)$.*

The proof of proposition 1 is found in the Appendix.
Having shown that we can always refine a copula distribution as above, we now wish to establish the possible range of values that can be taken by the rectangle $(p, 1] \times (q, 1]$ in a refined copula distribution. That is, we depart from the specific copula refinement defined in the Proof of Proposition 1, and ask what range of values can be allocated as the probability of $(p, 1] \times (q, 1]$ in some copula refinement.
Suppose that i and j are chosen such that u_i is the largest of the u-quantiles that is smaller than p, and v_j is the largest of the v-quantiles that is smaller than q (this includes the possibility that u_i or v_j is 0, or that u_{i+1} or v_{j+1} is 1). Define $\tilde{u}_1 = u_i$, $\tilde{u}_2 = u_{i+1}$, $\tilde{v}_1 = v_j$ and $\tilde{v}_2 = v_{j+1}$. The quantile partition $QP(\tilde{u}, \tilde{v})$ is a merging of $QP(u, v)$, and we can merge the copula distribution on $QP(u, v)$ to get one on $QP(\tilde{u}, \tilde{v})$.
Furthermore $QP(\tilde{u}, \tilde{v}; p, q)$ is a merging of $QP(u, v; p, q)$. Note that $QP(\tilde{u}, \tilde{v})$ has at most 9 rectangles and that $QP(\tilde{u}, \tilde{v}; p, q)$ has at most 16 rectangles - see Figure 6.3.
For convenience we shall now consider only the case of 16 rectangles, which occurs when $u_i, v_j \neq 0$ and $u_{i+1}, v_{j+1} \neq 1$, as shown on the right of Figure 6.3. Other cases are simplifications of the one we consider here and can be dealt with in the same way.
We label the 16 rectangles of $QP(\tilde{u}, \tilde{v}; p, q)$ as $R_{11}, \ldots, R_{44}$ as shown in the right hand of Figure 6.3.

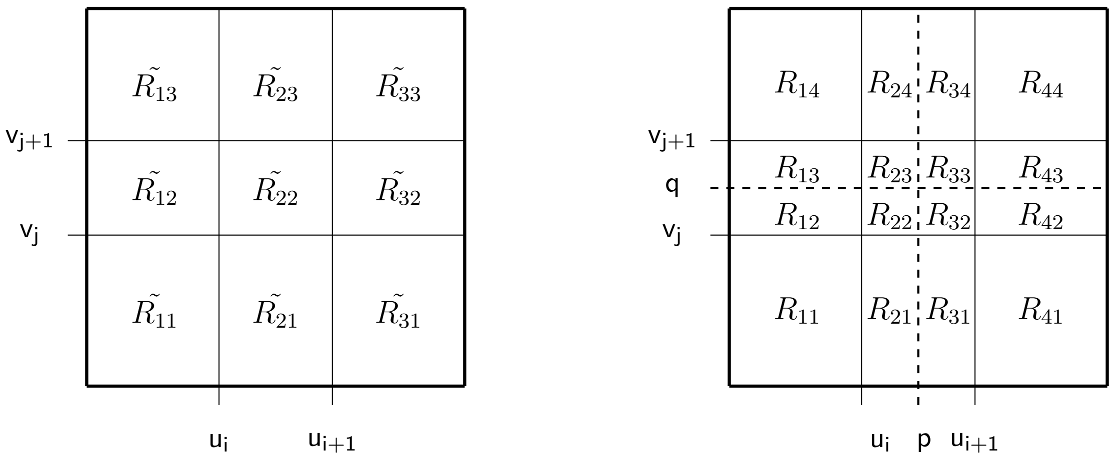


Figure 6.3: Maximum case of 16 partitions (right) resulting from partitioning 9 rectangles (left).

The 9 rectangles of $QP(\tilde{u}, \tilde{v})$ are labelled as $\tilde{R}_{11}, \ldots, \tilde{R}_{3,3}$ as shown in the left hand of Figure 6.3. Clearly $R_{11}, \ldots, R_{4,4}$ are each unions of rectangles in $QP(u, v)$, and furthermore,

$$\begin{aligned}
R_{12} \cup R_{13} &= \tilde{R}_{12} \\
R_{42} \cup R_{43} &= \tilde{R}_{32} \\
R_{21} \cup R_{31} &= \tilde{R}_{21} \\
R_{24} \cup R_{34} &= \tilde{R}_{23} \\
R_{22} \cup R_{23} \cup R_{33} \cup R_{32} &= \tilde{R}_{22}.
\end{aligned}$$

Suppose we are given a copula distribution on $QP(\tilde{u}, \tilde{v})$, for which $\tilde{p}_{st}$ is the probability of $\tilde{R}_{st}$ $(s, t = 1, 2, 3)$. We wish to assign copula probabilities p_{st} to the rectangles R_{st} $(s, t = 1, 2, 3, 4)$ so that the new distribution merges to p on $QP(\tilde{u}, \tilde{v})$.
For the merging we simply require,

- for the corner rectangles of $QP(\tilde{u}, \tilde{v})$: $p_{11} = \tilde{p}_{11}$, $p_{14} = \tilde{p}_{13}$, $p_{41} = \tilde{p}_{31}$, $p_{44} = \tilde{p}_{33}$,
- for the central rectangle in $QP(\tilde{u}, \tilde{v})$: $p_{22} + p_{32} + p_{23} + p_{33} = \tilde{p}_{22}$,
- for the remaining rectangles

$$\begin{aligned}
p_{12} + p_{13} &= \tilde{p}_{12} \\
p_{42} + p_{43} &= \tilde{p}_{32} \\
p_{21} + p_{31} &= \tilde{p}_{21} \\
p_{24} + p_{34} &= \tilde{p}_{23}.
\end{aligned}$$

To ensure that the new distribution is a copula we also need to impose two constraints corresponding to a row and a column:

$$\begin{aligned}
p_{21} + p_{22} + p_{23} + p_{24} &= p - \tilde{u}_1 \\
p_{12} + p_{22} + p_{32} + p_{42} &= q - \tilde{v}_1.
\end{aligned}$$

(Note that these constraints correspond to row 2 and column 2 of the right hand of Figure 6.3. We could also have specified similar constraints on row 3

and column 3, but it straightforward to see that these are redundant).
Now define,

$$f(p_{11}, ..., p_{44}) = p_{33} + p_{43} + p_{34} + p_{44}$$

to be the total probability in the square $(p, 1] \times (q, 1]$. This in a linear function of the p_{st} and we are free to choose it to take any value subject to the constraints listed above. As all these are linear, we immediately see that we have the form of a linear programming problem, and so the range of allowable values is an interval whose maximum and minimum values can be found be solving 2 LP problems. The cases in which $QP(\tilde{u}, \tilde{v}; p, q)$ has fewer than 16 rectangles work similarly. The above discussion (with minor adaptations to the other cases by removing further redundant constraints) can be summarized in the following Proposition:

Proposition 2. *The range of feasible values for the probability of* $(p, 1] \times (q, 1]$ *in any copula refinement of the copula distribution on* $QP(\tilde{u}, \tilde{v})$ *is given by the interval:*

$$[\min f, \max f]\,,$$

given the corresponding constraint sets.
We can obtain $\min f$ *and* $\max f$ *by solving feasible LP problems with at most* 12 *variables and* 9 *constraints.*

This now allows us to construct an algorithm for assessing copulas with expert judgements for quantile exceedance probabilities of the form:

$$P(Y > y_q | X > x_p)$$

where x_p and y_q index the p^{th} and q^{th} quantile for X and Y accordingly. For example, $p = 0.5$ and $q = 0.5$ correspond to the medians of X and Y. Other distribution areas can then be derived. Given a number of such coherent elicitations at quantile pairs $(u_1, v_1), \ldots, (u_n, v_n)$ we can calculate the copula distribution on the copula partition $QP(u, v)$.
For a new quantile pair (p, q), we then solve the LP problem to obtain the exact feasible range for the probability of $(p, 1] \times (q, 1]$. Note that this does not fully specify the distribution on all elements of the refined partition $QP(u, v; p, q)$. To achieve this, either

(a) we can carry out further elicitations at corner points in $QP(u, v; p, q)$ using proposition 2 repeatedly for obtaining feasible ranges from the expert; or

(b) we can make assumptions, such as minimally informative probabilities to restrict the number of elicitations required.

In the next section, we give a simple example of making assessments in the tail of the distribution along the lines of (a) but carried out in a slightly different order as there are few constraints in this case.

6.2.2 Commonly assessed quantile partition sequences

After having presented the mathematical set-up of refined partitioning generally, we now discuss some partitions that might be commonly assessed in practice.
One recurrent way of refining a joint distribution's assessments is by sequentially choosing a quantile for p and/or q that is either higher or lower than any previously assessed value. Then, we elicit the corresponding area above it for a new maximum or below it for a new minimum. Such sequences assess in particular

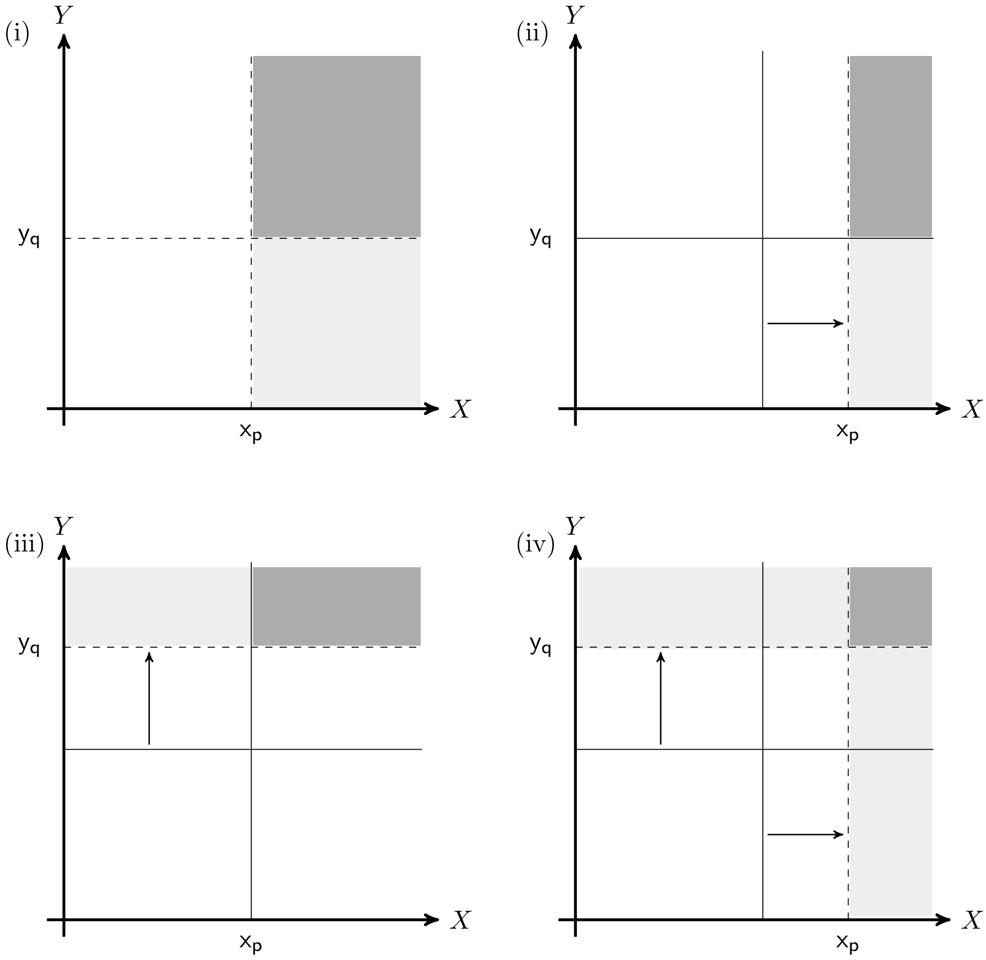


Figure 6.4: Example of a quantile partition for assessing the upper tail.

the distribution tails more explicitly. Figure 6.4 illustrates a sequence of quantile partitions on the upper tail constructed through setting new quantile maxima in (ii) to (iv) following an initial assessment (i) (note that this carries out the option (a) described in the previous section). We consider the procedure of Figure 6.4, i.e. further partitioning that probability mass which has been assessed directly in step (i) as most intuitive and practically useful. Nevertheless, the initial assessment also determines the probability mass in areas of the joint distribution which are not assessed further, $P(Y > y_q|X \leq x_p)$, $P(Y \leq y_q|X > x_p)$ and $P(Y \leq y_q|X \leq x_p)$, meaning we can also use a similar procedure to refine these.

First (in (i)), we elicit the overall probability mass and then subsequently refine the assessment. Suppose we first elicit $P(Y > y_{0.5}|X > x_{0.5})$. Alternatively, we might choose to elicit specific values, e.g. $1, 10, \ldots, 100$, rather than common quantiles, such as the median. This relates to the choice of whether to frame the elicitation question in terms of quantiles or values. Both have been suggested (as P- and V-methods) since the pioneering probability elicitations by the Stanford Research Institute in the 1970s [375].
Following (i), we elicit a refined quantile partition as determined by a new x_p in (ii). A common choice here might be the 90^{th} or 95^{th} quantile in order to assess the probability mass in the joint distribution's extreme (tail) region. Thus, we elicit for instance $P(Y > y_{0.5}|X > x_{0.95})$. In the illustrative case-study of

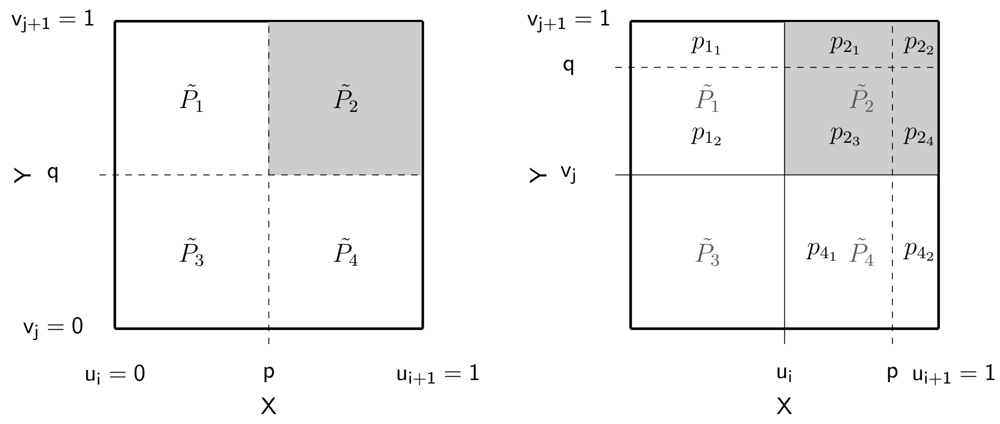


Figure 6.5: Quantile partition of the joint distribution from (i) to (iv).

Section 6.4, we use a scenario mapping method [414] prior to the elicitations in order to gauge experts' familiarity with such tail judgements and decide on a quantile for which experts are comfortable to make assessments.

In (iii), we condition on Y and the new y_q is chosen to assess the tail region. With x_p being the median, we thus elicit $P(X > x_{0.5}|Y > y_{0.95})$. Depending on the underlying meaning of the variables, and knowledge about causal or probabilistic relationships (see e.g. [350, 415]), the expert might find it easier to condition on one variable than the other. Our method is flexible enough to allow for this.

In the last step of this quantile partition sequence, experts assess either $P(Y > y_{0.95}|X > x_{0.95})$ or $P(X > x_{0.95}|Y > y_{0.95})$, depending on case-specific interest, whereas p and q are the ones from the previous two rounds. Thus we further explore the joint tail region. Figure 6.5 displays the refinement in the quantile partition from the first to the latest assessment.

The assessments' feasibility ranges are as follows. The assessment in (i) is unrestricted, meaning experts can assess any value between $[0, 1]$. If the expert believes the variables are independent, the assessment is equal to $P(Y > y_q)$, that is learning about X does not change experts' belief. For negative dependence, the assessment is between $[0, P(Y > y_q))$ and for positive dependence, it is within $(P(Y > y_q), 1]$.

All following assessments on the other hand are restricted and only feasible if the assessed value falls within the range which is determined by solving the LP problem of minimising and maximising the possible values of the assessed area subject to the constraints that any new partition simply adds up to their previous assessments (see medium and dark grey areas $\tilde{P}_k$ in Figure 6.2) while areas which have not been newly partitioned do not change (see light grey areas $\tilde{P}_k$ in Figure 6.2). Consider for example the assessment in (iv). It is only feasible within the range that is determined by solving the following LP problem (with regards to Figure 6.2 on the right):

$$\left.\begin{matrix}\min\\ \max\end{matrix}\right\} p_{2_2} \tag{2.1.1}$$

subject to

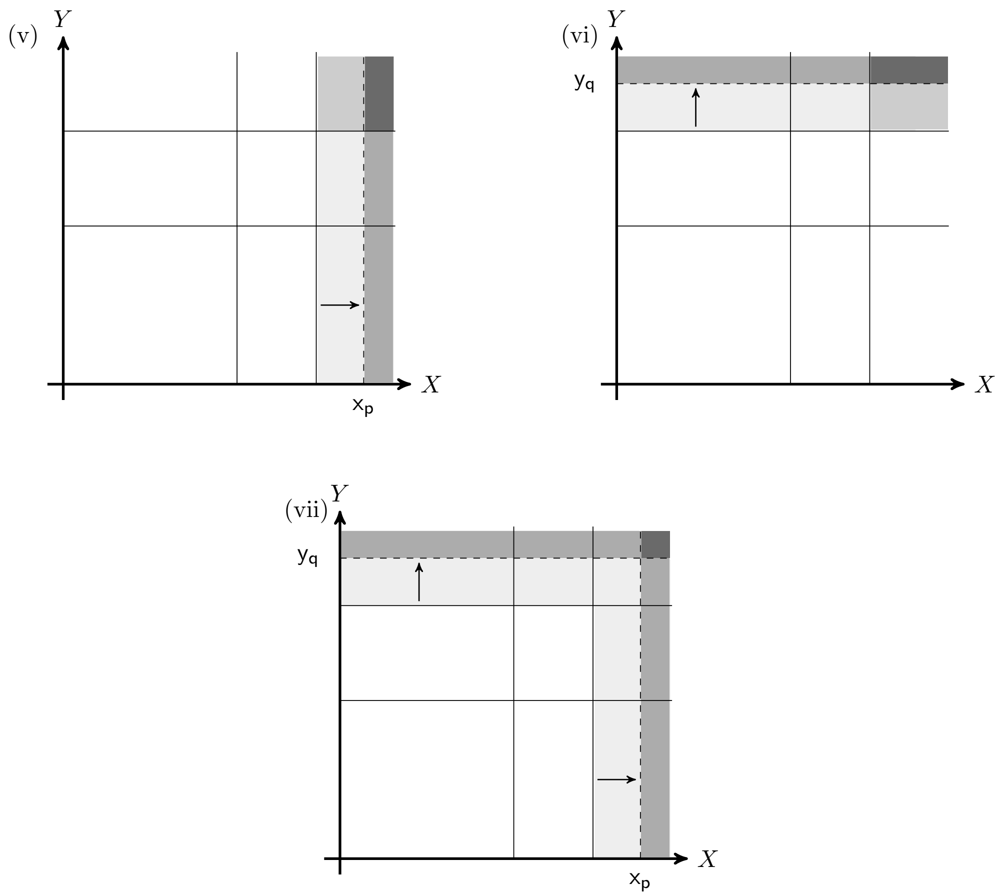


Figure 6.6: Further refining the assessment on the joint upper distribution tail.

$$p_{1_1} + p_{1_2} = \tilde{P}_1 \tag{2.1.2}$$
$$p_{2_1} + p_{2_2} + p_{2_3} + p_{2_4} = \tilde{P}_2 \tag{2.1.3}$$
$$p_3 = \tilde{P}_3 \tag{2.1.4}$$

and

$$p_{4_1} + p_{4_2} = \tilde{P}_4 \tag{2.1.5}$$

Experts express negative dependence again through a judgement close or equal to the lower bound, positive dependence is expressed by judgements close or equal to the upper bound and independence is assessed as before. As the upper and/or lower bounds deviate from the standard range of $[0, 1]$, it is necessary to communicate these restricted feasibility bounds to an expert and explain their interpretation.
The procedure for assessments (ii) to (iv) is repeated as often as necessary (with appropriate modifications) to obtain a desired level of detail (see assessments (v) to (vii) in Figure 6.6 for the next round of three assessments). Having assessed previously the 90^{th} or the 95^{th} quantile of X and Y, we now might consider the 99^{th} quantile. This allows for "zooming in" on the joint distribution's tail even further.

The resulting quantile partitions are illustrated in Figure 6.7.

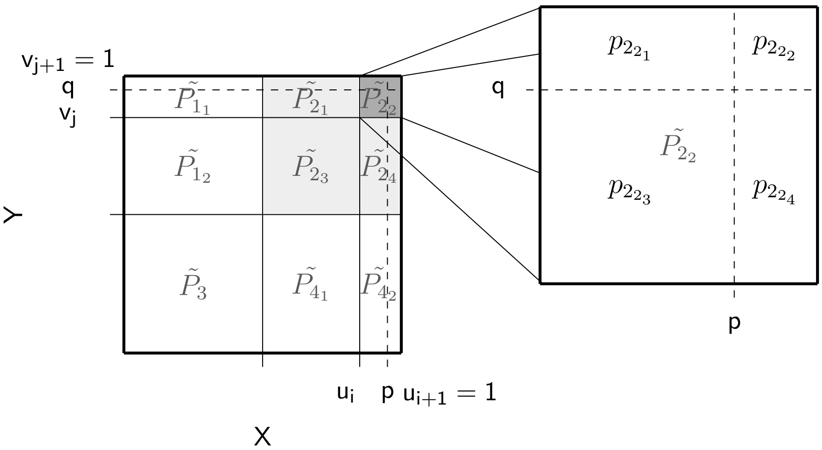


Figure 6.7: Resulting quantile partitions after further refining the previous assessments.

While this section presents an example with a focus on refining the upper distribution tail, remember that the generality of the method (as introduced in Section 6.2.1) allows for any further refinement of the distribution, such as for instance shown in Figure 6.2 (on the right).

6.3 Modelling the form of directly and indirectly assessed probability masses through minimum information

After having presented the elicitation procedure, which allows for feasibly assessing the probability mass within any part of the joint distribution, in this section we outline how we model the form of directly and indirectly assessed parts as minimally informative.
The reason for a minimum information approach is to address the modelling issue of underspecification. We do not have enough information for choosing a distribution that fits the experts' assessments uniquely but we wish to find the simplest distribution that matches them. This approach allows us to derive a unique distribution regardless the quantile partition's level of detail. As such, it does not restrict the flexibility of the assessment procedure from section 6.2.
Formally, we aim for modelling dependence through that copula which is chosen to have minimum information (also called *Kullback-Leibler divergence* [242]) with respect to the uniform copula given the quantile constraints. The resulting distribution is considered the most independent copula satisfying the constraints.
Consider the joint distribution $g(x, y)$ with marginal densities $g_1(x)$ and $g_2(y)$. Whenever g_1 and g_2 are not independent, i.e. $g(x, y) \neq g_1(x)g_2(y)$, we need to model the dependence between them. To do so, we introduce the concept of relative information $I(g; h)$ which is a measure of similarity between the two distributions and it is defined for $g(x)$ with respect to $h(x)$ as:

$$I(g; h) = \int g(x) \log \left(\frac{g(x)}{h(x)} \right) \mathsf{d}x \tag{6.3.1}$$

Whenever $g(x) = h(x)$, it follows that $I(g; h) = 0$. A higher value of $I(g_1; g_2)$ corresponds to less similarity. We consider $h(x)$ a background distribution, com-

monly chosen as uniform or log-uniform. Alternatively, we use sensitivity analysis for selecting an appropriate form [35]. Together with the constraints, this choice determines the form of $g(x)$ in absence of further information [32].
Information is invariant under monotone transformations. Therefore, if c_g and c_h are copula densities associated with the previous densities g and h, we have $I(c_g; c_h) = I(g; h)$. In particular if h is the joint independent distribution with the same marginals as g (g_1 and g_2), so that $h = g_1 g_2$ then $I(g; g_1 g_2) = I(c; \mathbf{uniform})$ where h is the uniform copula. This gives the interpretation of our minimum information copula as the most independent copula given the constraints.
See [34] for a detailed derivation on how a minimum information distribution can be approximated by the equivalent distribution of maximum entropy[5] [360]. For an extensive discussion on obtaining a minimum information copula through the convex optimisation problem, we refer to [32, 34, 35]. Here, it suffices to say that the conditional density within each rectangle is uniform. As discussed in Section 6.2, when we stop eliciting information from experts, some rectangles' density has been directly assessed by an expert while for other rectangles the mass is given indirectly through related assessment and the marginals. In order to obtain a unique solution for the whole distribution, we hence need to solve the minimisation problem of equation for directly and indirectly assessed parts. We refer to [34] and [273] for the corresponding proofs that such a minimum information distribution exists and is unique. Furthermore, [35] and [42] discuss and apply a Lagrangian dual for a minimum information problem to show a way for obtaining more insight on the optimal solution.

6.4 An illustrative case-study: assessing spatial dependence of political violence/terrorism risk in insurance underwriting

Given the flexibility and detail that the SRP method allows for when modelling dependence, we regard it as of particular interest for application areas in which common simplifying assumptions, such as bivariate normality, are not justified. Rather, different kinds of tail dependencies which potentially induce extreme impact scenarios are prevalent. For these, we often assess and model upper and lower tail dependence exclusively (similarly to testing the goodness of fit for asymmetric, Archimedean copulas to historical data when available) given that e.g. joint large losses are typically not observed together with joint large gains[237, 291].
As such, we consider (re-)insurance as an industry in which rigorous dependence modelling approaches are of particular interest. Due to the increasing complexity of (re-)insurance products, new (holistic) modelling approaches, such as *dynamic financial analysis* (DFA) (a Monte Carlo simulation-based method to model risks jointly), have become popular among actuaries to better understand the risks an insurer underwrites [137]. For these new approaches, flexible and detailed assessments of dependencies under a specific probability model are required. Exemplary for a DFA application, [136] present how various parametric copulas can be used for stress-testing an insurer's risk management strategies together with the implication on stakeholders, such as regulators and rating agencies. The

[5] In the context of expert judgement, an invariance approach to encoding information probabilistically is considered a main justification for maximum entropy methods [304].

DFA model inputs, the perils (or risks) covered by an insurer, are informed by a *catastrophe model*. The components of catastrophe models are a hazard, inventory, vulnerability and loss estimation module. The loss estimation output is usually an exceedance probability curve specifying probabilistically the severity levels of a certain hazard in a region. Capturing relevant dependencies between severity levels is crucial for a more robust output. See [179] for an introduction to catastrophe models.
We have already established that a common challenge is lacking relevant historical data for quantifying dependence relationships serving as model input. In actuarial risk assessment, *non-life* insurance underwriting is particularly challenged. So called *low frequency-high severity* perils, natural and man-made, are by definition not frequently observed but cannot be ignored. Therefore, we require structured expert judgement to model their uncertainty. In this illustrative case-study, we apply the SRP method to elicit and model the spatial dependence of the man-made peril of terrorism. Terrorism attacks are not only often low frequency-high severity catastrophes but pose an additional challenge due to intelligent adversaries which further inhibit the use of historical data. Better understanding the dependence between terrorism attacks' frequencies in different regions globally is nevertheless key for an insurer to quantify and price this peril's risk when managing a portfolio of (global) clients[6].

6.4.1 Pricing terrorism risk in insurance

Traditionally, pricing of terrorism risk in insurance has not been evaluated from actuarial principles, but rather covered by the balance of supply and demand in the insurance market together with some less formal risk selection from site surveys [427]. Terrorism coverage (e.g. in the United States) had been included in standard commercial insurance policies as an unnamed peril on all-risk commercial and home owners coverages for property and contents [276]. More recent loss developments though have highlighted the necessity of treating its risk assessment more rigorously. A major turning point for dealing with terrorism risk in insurance was the attacks of September 11^{th}, 2001 (9/11) on the United States. The attacks incurred an estimated monetary loss up to 60 billion US dollars, distributed among various lines of business, such as property insurance, business interruption insurance and workers' compensation [108]. Globally, the worst 15 terrorist attacks in terms of casualty numbers have occurred since 1982 with many more near-miss events [276]. Mathematically, the relationship between the frequency of more recent attacks and their severity can be described by a power law, i.e. attack severities that are orders of magnitude larger than the mean can be common [72]. The changing nature of its risk through an increasing number of frequencies and severities in multiple regions globally underlines the urgent need for improved assessment.

6.4.2 Expert judgement for adversarial risk

A specific aspect of assessing terrorism risk is the role of intelligent adversaries. Their impact is thus included in recent discussions on risk definitions [22, 24, 21, 94]. In fact, 9/11 led many researchers to propose modified risk

[6]As [427] emphasises, we must not confuse quantifying terrorism risk with predicting a next attack. This is similar to natural catastrophes, such as earthquakes, for which we cannot determine the time, location and severity of the next event, but the aim is rather to evaluate the annual exceedance probability of loss, for instance to inform a property insurance portfolio.

definitions [182]. For instance, the triplet definition by [223] is extended to include adversaries in [160] and [161] by considering the likelihood of a hazard as the conditional probability of a successful attack given that an attack is planned. Models addressing adversarial risk are typically of game-theoretic nature [139, 95, 315] whereas the area of *adversarial risk analysis* comprises decision-analytic approaches combining traditional *probabilistic risk analysis* (PRA) methods with game theory [331, 343, 344]. Nevertheless, there is some debate on (traditional) PRA's effectiveness for adversarial problems (see [139] defending its usefulness and [58] and [96] arguing against it). A main argument against PRA approaches for adversaries is the dynamic attacker's decision rule for choosing a target as this choice might be based on the anticipated defender's assessment of targets' likelihoods. In other words, a defender's PRA might inform the attacker's choice and hence override its purpose as the previously most likely target has now zero probability of being attacked (closely related in terrorism risk analysis are decision on allocating defensive resources [43]). Experts quantifying adversarial risk should therefore decompose their judgement in accordance with adversarial risk definitions, so that we understand experts' beliefs about attackers' choices. When doing so, assessments of an attack choice might be based on attackers' motivations, resources and capabilities together with defenders' vulnerabilities. In that way, expert judgement is used in the Probabilistic Terrorism Model by Risk Management Solutions Inc. (RMS[7]) for assessing likelihoods on target selection, capabilities of attack modes and an attack's overall likelihood. However, dependence between targets is neglected [421]. In other approaches, event trees are used to reason from an attacker's capabilities through a defender's counter-measures [348, 406]. In addition, several qualitative approaches for structuring the available knowledge on terrorists' objectives and motivations exist in the risk and decision analysis literature [363, 226, 160].

6.4.3 Expert judgement for spatial dependence of terrorism attacks

Knowledge and beliefs on terrorists' motivations, resources and capabilities together with defender's vulnerabilities inform experts directly about the spatial dependence between attack frequencies. Terrorist groups, such as the Irish Republican Army (IRA), Basque Separatist Group (ETA) or as well Hamas and the Palestine Liberation Organization (PLO), had and have specific geographical foci with a politically motivated attack purpose. Their goals are formulated and self-proclaimed as separatism or liberation. The attacks' geographical impact is identified straightforwardly. Based on the number of active terrorist groups per region plus their resources and capabilities relative to counter-measures, an expert assesses either positive or negative dependence. While positive dependence might not seem intuitive at first due to different local foci and typically a lack of collaboration between these groups, learning and encouragement by another groups' successes can still occur. [428] regards learning of optimal behaviours beyond the own organisation as a main strength of some well-known terrorist groups. Other scenarios for positive dependence can be due to defenders' collaboration, joint counter-terrorism activities and sharing of intelligence resources.
In contrast to terrorists motivated by self-proclaimed liberatism and separatism, other groups derive their goals from religious ideology. These groups are often

[7] RMS, founded at Stanford University in 1989, provides services in the area of catastrophe modelling for (re-)insurers.

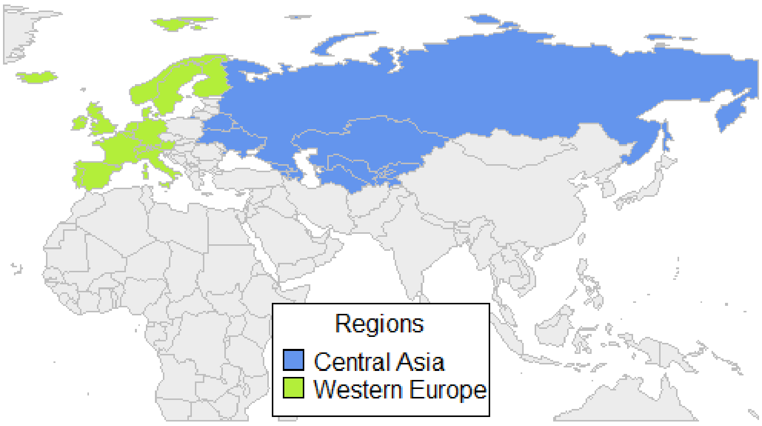


Figure 6.8: Regions of interest for dependence assessment.

globally active. Their members are organised as multiple independent hubs with satellite cells. Al-Qaeda and the Islamic State of Iraq and Syria/the Levant (ISIS/ISIL) are typical examples of such network-based organisations [427, 429]. Models from swarm intelligence and statistical network analyses are used to evaluate the effectiveness of counter-terrorism measures and understand the attackers' capabilities. It is understood that organisations like Al-Qaeda and ISIS/ISIL are more resilient and capable of more severe attacks than (hierarchical) army-like structured groups [428]. For dependence assessments, understanding the global presence of members and sympathisers (potentially future recruits) together with the functioning of the network structure is crucial. For instance, scenarios of positive dependence can occur when a terrorist group obtains more power and resources to extend globally or when new attack types are used for which little intelligence or counter-measures exists. Scenarios of negative dependence might describe attackers' scarce resources, e.g. lacking financial support for regional hubs, so the target focus shifts towards a certain region. The latter also depends on vulnerabilities of target countries, desired attention through media or as well a planned revenge, e.g. for a country's military actions.
While these are only brief considerations for scenarios that can influence the assessment of dependence between the number of terrorist attacks in different regions (see [428] for a more extensive discussion on regional and global terrorism), it shows the complexity of factors to be thought of. In this illustrative case-study, we focus on the geographical regions of Central Asia (CA) and Western Europe (WE) which are shown in Figure 6.8 (see the Appendix for a full list of the countries included per region).

6.4.4 Eliciting the marginal probabilities

Before eliciting dependence assessments from experts, we need to specify the marginal distributions for the variables of interest. Otherwise, the experts condition their judgements on different marginal probabilities and their assessments cannot be sensibly aggregated. The specification is done either through historical data (if available) or another, prior elicitation with a structured expert judgement method for univariate uncertainty, such as [330, 175, 189]. A structured elicitation for the marginal distributions is also encouraged when eliciting

dependence only from one expert, i.e. without aggregation, as this mitigates potential biases of the marginals and ensures transparency[415].
In our case-study, the marginal distributions have been assessed by 17 experts[8]. The experts are involved in analysing and pricing the peril of terrorism and other armed conflict categories. They work for different (re-)insurers, catastrophe modellers and related service providers. The elicitation session was organised as part of the European Cooperation in Science and Technology, COST Action IS1304 - Expert Judgement Network, which aims at stimulating the emergence and spread of high quality evidence-based decision support approaches through structured expert judgement methods. The marginal distributions $F_X(x)$ and $F_Y(y)$ are defined as the number of terrorist attacks in Central Asia (x) and in Western Europe (y), both in 2017. We define a terrorist attack in accordance with common global data-bases on the topic (see [380]). Thus, for an attack to be recorded as such there must be evidence of an intention to coerce, intimidate, or convey some other message to a larger audience (or audiences) than the immediate victims. In this regard, any perpetrator group, any weapon type (e.g. biological, chemical, explosive, firearms etc.), any attack type (e.g. armed assault, bombing, facility/infrastructure attack, hostage taking etc.), any target apart from private persons (i.e. business, infrastructure, military, educational/religious institutions etc.) is included.
We elicited $F_X(x)$ and $F_Y(y)$ through the so called *Classical Model* [**?**, 79]. Experts provide various quantile assessments for a continuous quantity rather than point estimates. Usually (and in our case), we elicit the 5^{th}, 50^{th} and 95^{th} quantile. The experts answer two types of questions. The first questions are about so called seed or calibration variables. For these, the true value is known to the analyst but not the experts at the time of the elicitation (or they will be known later and within the time frame of the study). The second question type is about the actual target value or variable of interest, i.e. the uncertainties we intend to include in the model. Based on each expert's assessments of the seed variables, the experts are aggregated. For that, two performance measures are derived, the calibration and information score. Loosely, the calibration score measures the statistical accuracy of the experts whose assessments are treated as statistical hypotheses. The information score measures the assessments' concentrations relative to a background distribution. Good expertise is shown by a high calibration and information score (see [79] for a more detailed introduction). Figure 6.9 shows each experts' individual assessment for the target variables' marginal distributions together with the aggregated assessments of equal weighting (EW) and the classical method (DM global).

We observe in Figure 6.9 that the marginal distribution assessments are similar for both regions whereas most of the experts provide narrow uncertainty bounds. The experts who are more uncertain are so for both assessments. Hence, the performance-based and the equally weighted combination show no major difference for either region. As commonly observed with the classical method, the performance-based aggregation is more informative even if both combinations lead to similar median assessments.

[8] The 17 experts are from a first elicitation round from a currently ongoing study that aims to include more experts.

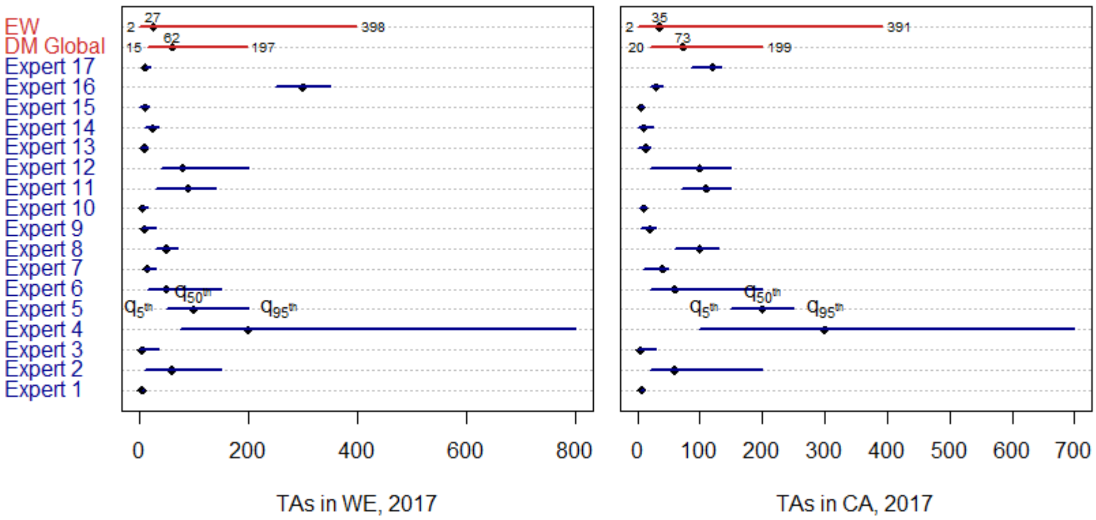


Figure 6.9: Outcome of eliciting the marginal distribution for each region.

6.4.5 Applying the SRP method for quantifying spatial dependence of terrorism risk

Once the marginal distributions had been elicited, we proceeded with eliciting and modelling dependence through the SRP method. This elicitation was done with a single expert who is a professional in the area of terrorism catastrophe modelling within (re-)insurance (as well) and who subscribed to the aggregated results for the marginal distributions. In total, we elicited six dependence judgements in addition to one further marginal assessment. The latter was required as we had not considered the 99^{th} quantiles previously. As outlined in the initial exemplary procedure in the previous chapter, we started by first eliciting an overall probability mass which was later partitioned to further explore the joint upper distribution tail. The first elicitation is therefore on the probability of the terrorist attack frequency in Western Europe (y) being above its 50^{th} quantile, 62 attacks, given that we observe more than 73 attacks in Central Asia (x) (again the corresponding 50^{th} quantile), both in the year 2017. All judgements were conditional probabilities given the expert's familiarity with its interpretation. Table 6.1 summarises the dependence assessments by showing the results together with the framing of the questions.

As second part of the SRP method, we then modelled the overall joint distribution for the spatial dependence through solving the minimum information minimisation problem based on the above assessments. The result can be seen in Figure 6.10.

We see that the expert's distribution indicates a slight negative dependence relationship between the spatial terrorism risk of both regions which is however close to independence. This is particularly driven by the first assessment being equal to 0.5 which indicates independence for a broad area of the joint distribution. In more detail, the difference between assessment *ii.*) and *iii.*) shows that in the joint tail, the expert assesses that an extreme year in terms of number of attacks for WE affects CA more than vice versa. The slight negative dependence (close to independence) corresponds to the expert's rationale which has been formally facilitated in order to support the expert with structuring his/her knowledge about the spatial dependence between both regions. For that, we used a conditional scenario mapping method [414]. In addition to mitigating some

Table 6.1: Overview of dependence elicitation procedure and results.

	Framing "Given that we observe [...]"	**Conditional Probability**	**Assessment**
(*i*)	"[...] more than **73** terrorist attacks in CA, what is your probability that we observe more than **62** terrorist attacks in WE?"	$P(Y > y_{0.5} \| X > x_{0.5})$	**0.5**
(*ii*)	"[...] more than **199** terrorist attacks in CA, what is your probability that we observe more than **62** terrorist attacks in WE?"	$P(Y > y_{0.5} \| X > x_{0.95})$	**0.03**
(*iii*)	"[...] more than **197** terrorist attacks in WE, what is your probability that we observe more than **73** terrorist attacks in CA?"	$P(X > x_{0.5} \| Y > y_{0.95})$	**0.045**
(*iv*)	"[...] more than **199** terrorist attacks in CA, what is your probability that we observe more than **197** terrorist attacks in WE?"	$P(Y > y_{0.95} \| X > x_{0.95})$	**0.025**
(*v*)	"[...] more than **199** terrorist attacks in CA, what is your probability that we observe more than **225** terrorist attacks in WE?"	$P(Y > y_{0.99} \| X > x_{0.95})$	**0.04**
(*vi*)	"[...] more than **225** terrorist attacks in WE, what is your probability that we observe more than **199** terrorist attacks in CA?"	$P(X > x_{0.95} \| Y > y_{0.99})$	**0.01**

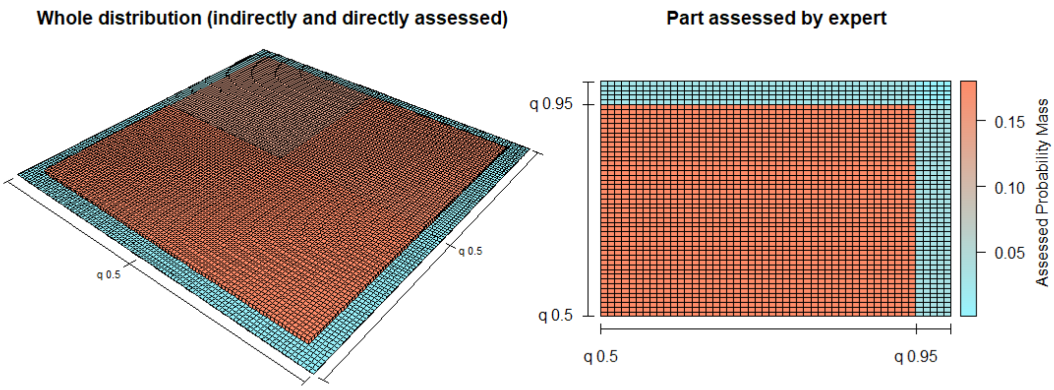


Figure 6.10: The experts joint distribution: overall (left) and assessed upper quadrant (right).

prevalent cognitive fallacies of assessing dependence, such as the *confusion of the inverse* or *confusing joint and conditional probabilities*(see also [415] for an overview), this method allows for considering and reflecting explicitly which scenarios affect the probability spaces of both regions (in a conditional sense). Scenarios are defined as "sequences that link triggering events to specified consequences (or final states) through intermediate conditions" [414]. For the example shown in Figures 6.11 and 6.12, the expert first reasoned through backwards logic, i.e. starting from the specified consequence, about observing more than 199 in Central Asia until the end of 2017. Then, based on the initiating events that might cause Central Asia to experience more than 199 attacks and which are (at least partly) observable today, the expert reasoned (in forward logic) how these same initiating events affect the development of the number of terrorist attacks in Western Europe until the end of 2017. Based on the the number and plausibility of these conditional scenarios causing more than 255 attacks, the expert could then make a dependence assessment in a more informed and confident manner. [414] presents the structured process of generating such conditional scenarios in more detail.
As can be seen in Figures 6.11 and 6.12, the expert considers both regions to be slightly negatively dependent (close to independence) due to the consideration that the active terrorist groups in both regions are different. In Central Asia, local separatists have political and regional motivations while in Western Europe mainly islamist groups are prevalent despite e.g. Russia's military involvement in the Middle East. Furthermore, the expert considers both regions to be different with regards to their vulnerability given not only the types of active terrorist groups but also the varying counter-terrorism and intelligence capabilities which drive the negative dependence relationship.
Before concluding this illustrative example, a first remark is that for quantifying the spatial dependence of terrorism attacks the definition used in this example is rather broad by including all attack types. Thus, the consideration of specific attack types might have very particular effects on the geographical interdependencies. As such, of growing interest in the adversarial risk literature have been biological attacks [139] and cyber attacks [362]. For these, it can be informative to assess the dependence between variables of interest, such as casualties or monetary losses.
Further, we understand that an elicitation considering more explicitly the geographical interdependencies of critical infrastructure can be informative for insurers, for instance when offering business interruption coverage. Our method could hence build upon some modelling approaches that have ranked the susceptibility of critical infrastructures targeted by attacks [316].
Lastly, we acknowledge the inherent difficulties particular when considering attacks, such as 9/11, which some might title "black swans". For dependencies, the term "perfect storms" appeared (see [314] for a discussion on the use of these terms in risk analysis and management). However, even for such events, structured assessment through experts can be informative and it is interesting that e.g. Zelikow (as director of the 9/11 Commission) called the misreading of precursors to these events as "failure of imagination" given that air-planes had been used before as weapon and the World Trade Center in New York had been targeted already in 1993 [314].

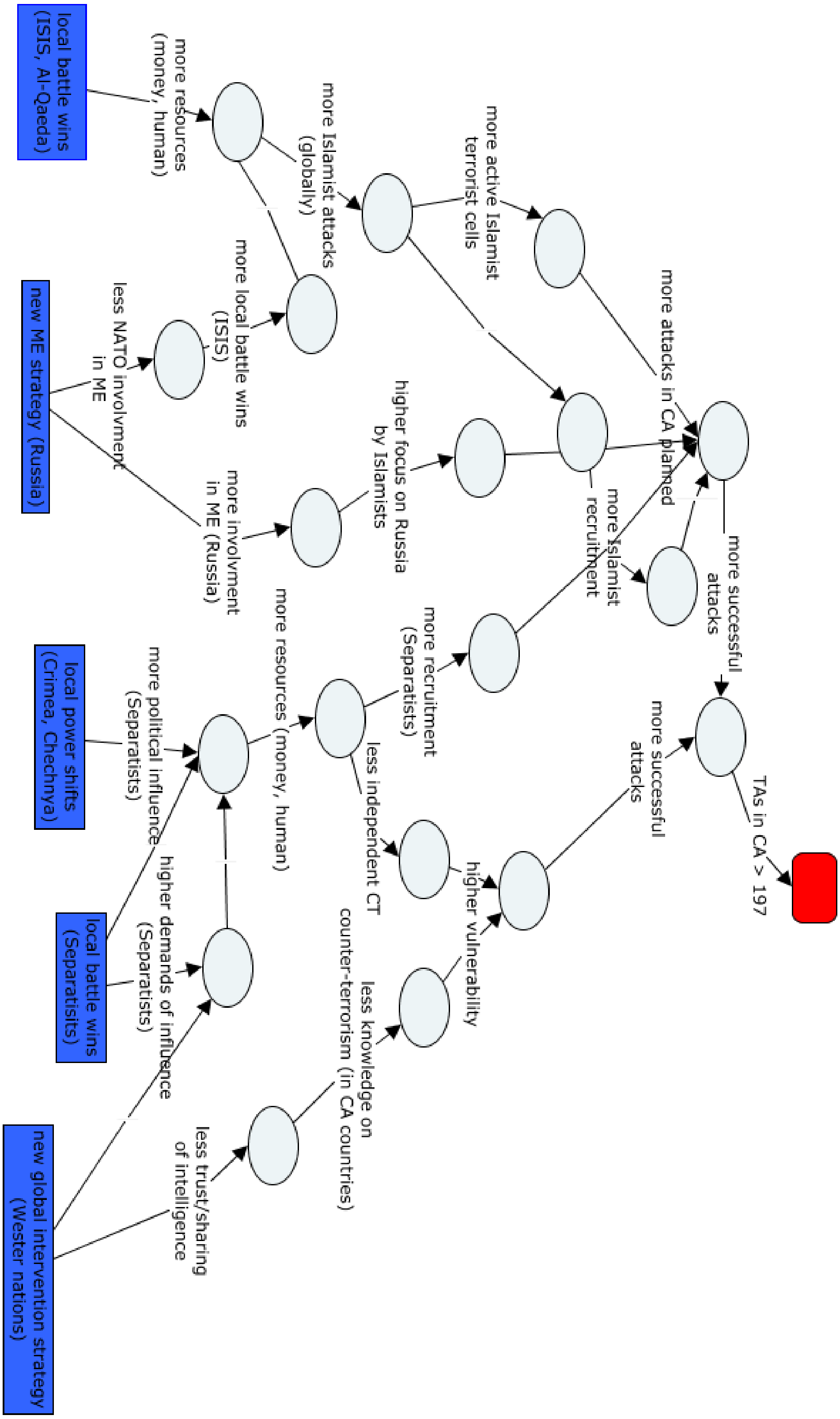


Figure 6.11: Unconditional scenarios for assessments.

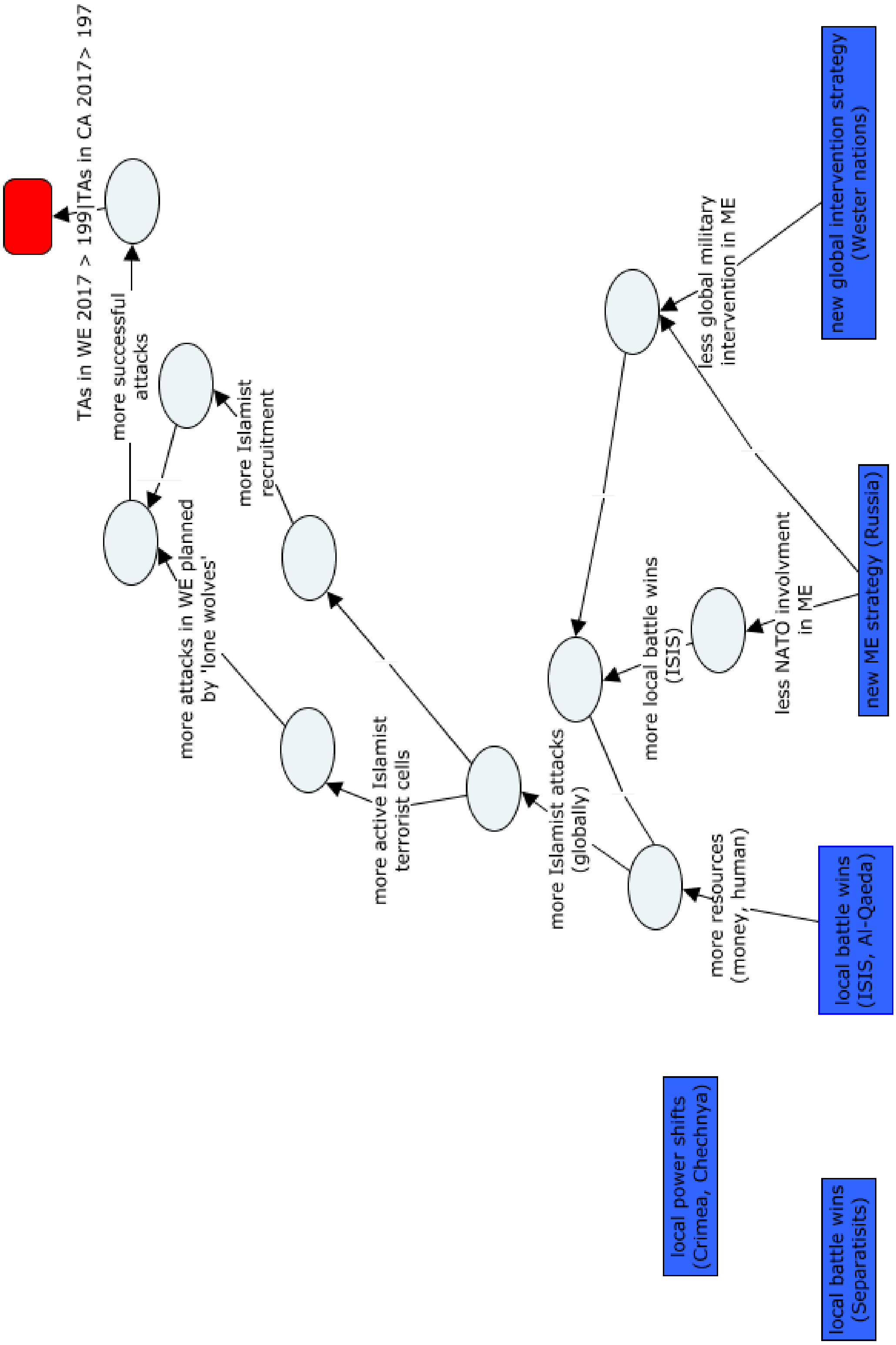


Figure 6.12: Conditional scenarios for assessments.

6.5 Discussion and chapter conclusions

When using expert judgement for assessing dependence, there is a trade-off between easing the assessment burden for experts and sufficiently capturing a real-world phenomenon of interest in our model [412]. Here we presented an elicitation method that aims to satisfy a decision-maker's desired level of detail for a model, whereas the procedure of eliciting dependence from experts provides an intuitive way of assessing some detailed dependence information (such as extreme parts of a joint distribution) while avoiding infeasible and inconsistent assessments. We argued that for the decision-maker a non-parametric setting of modelling multivariate uncertainties is applicable and therefore we addressed the potential assessment issues of under- and overspecification.
In addition to a discussion on various application areas in which our method might be of interest, we provided an illustrative example of terrorism risk. In future research, more applications are desirable to explore how the SRP method performs and to obtain insights around issues like alternative ways of framing the judgements, the implication of restricted feasibility ranges, or the elicitation of different forms (other than conditional probabilities). For example, an alternative to eliciting quantile-based assessments, we can elicit conditional expectations. This follows from the discussion of [412] on modelling and elicitation strategies that are determined by the choice of considering influencing factors of dependence relationships explicitly or implicitly. The latter is similar to PI methods which aim at satisfying reasonable conditions of a model output due to its easier understanding and quantification. This is of particular interest when we cannot observe (and hence elicit) our variables of interest directly. [35] show an elicitation procedure and minimum information modelling for expectations on the whole joint distribution. Hence, considering its elaboration based on our method could allow for a more detailed specification of multivariate uncertainty for non-observable model input parameters. In the actuarial context of section 6.4, we might ask experts to assess the conditional expectation for a risk measure, such as *probable maximum loss* (see [179]), which can be used (partly) as model output, whereas we assess dependence through PI on the function generating it.

Appendix

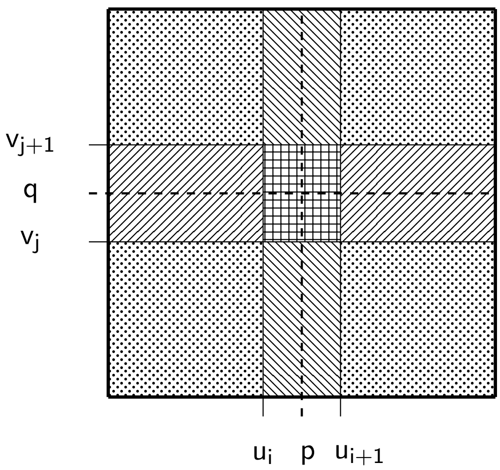
Proofs for section 6.2.1:

Proof for Proposition 1:

Suppose we are given values $u_0 = 0 < u_1 < \cdots < u_n < 1 = u_{n+1}$, and $v_0 = 0 < v_1 < \cdots < v_m < 1 = v_{m+1}$ (where $n, m > 0$), $0 < p, q < 1$, with p different to the u_i and q different to the v_j. Then a copula distribution on $QP(u, v)$ can be refined to a copula distribution on $QP(u, v; p, q)$.

Proof. In order to prove proposition 1, we divide the set $QP(u, v)$ into four subsets:

1. $A(p, q)$ has a single element which is the rectangle of $QP(u, v)$ containing the point (p, q).
2. $U(p, q)$ is the set of rectangles in $QP(u, v)$ that overlap the line $v = q$, except the one in $A(p, q)$.



=elements of $B(p, q)$
=elements of $V(p, q)$
=elements of $U(p, q)$
=elements of $A(p, q)$

Figure 6.13: Different set of rectangles in $QP(u, v)$.

3. $V(p, q)$ is the set of rectangles in $QP(u, v)$ that overlap the line $u = p$, except the one in $A(p, q)$.

4. $B(p, q)$ is the set of all rectangles in $QP(u, v)$ that are not in $A(p, q)$, $B(p, q)$, or $V(p, q)$.

Define also $A^*(p, q)$ to be the rectangles in $QP(u, v; p, q)$ which are sub-rectangles of $A(p, q)$, and define U^*, V^* and B^* similarly.
Note that $B^*(p, q) = B(p, q)$, that is, the rectangles in $B(u, v)$ do not get subdivided by the lines $u = p$, $v = q$. Rectangles in U^* are obtained by dividing rectangles in U by the line $v = q$, and rectangles in V^* are obtained by dividing rectangles in V by the line $u = p$.

We now define the refined copula distribution on $QP(u, v; p, q)$.
Let $\alpha = (p - u_i)/(u_{i+1} - u_i)$, and $\beta = (q - v_j)/(v_{j+1} - v_j)$. We specify how to define the refined copula distribution as follows:

1. For the rectangles in A^* , the lower left sub-rectangle is allocated $\alpha\beta$ of the mass of A, the lower right one gets proportion $(1 - \alpha)\beta$, the upper left one gets proportion $\alpha(1 - \beta)$, and the upper right one gets proportion $(1 - \alpha)(1 - \beta)$.

2. Each rectangle in U is subdivided into two sub-rectangles in U^* by the line $v = q$, and the lower sub-rectangle is allocated proportion β of its mass and the upper sub-rectangle is allocated proportion $(1 - \beta)$ of the mass.

3. Each rectangle in V is subdivided into two sub-rectangles in V^* by the line $u = p$, whereas the left sub-rectangle is allocated proportion α of its mass and the upper sub-rectangle is allocated proportion $(1 - \alpha)$ of its mass.

4. Any rectangle in $B^*(p, q) = B(p, q)$ is assigned the same probability as it was in in the copula distribution on $QP(u, v)$.

This allocation of probabilities to the rectangles of $QP(u, v; p, q)$ adds to 1, while it is straightforward to check that it is a copula distribution. $\square$

Regions of interest in illustrative case-study (section 6.4):

- Central Asia: Armenia, Azerbaijan, Belarus, Georgia, Kazakhstan, Kyrgyzstan, Russia, Tajikistan, Turkmenistan, Ukraine, Uzbekistan.
- Western Europe: Austria, Belgium, Denmark, Finland, France, Germany, Iceland, Ireland, Italy, Luxembourg, Malta, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom.

Chapter 7

Eliciting and modelling probabilistic dependence between multivariate uncertainties of bug-drug combinations

This chapter[1] addresses the future impact of antibacterial resistance through increasing resistance rates which are of concern for various stakeholders, ranging from clinicians, researchers and decision-makers in the pharmaceutical industry to healthcare policy-makers. Due to multidrug resistance, it is a complex challenge so that neglecting the dependence between uncertainties of future resistance rates might severely underestimate the systemic risk for certain bug-drug combinations. Therefore, in this chapter, we model the dependence between several important bug-drug combinations that are of interest for the UK. As a commonly encountered challenge in probabilistic dependence modelling is the lack of relevant historical data to quantify a model, we present a method for eliciting the dependence information from expert in a formal and structured manner. This allows for ensuring transparency and reproducibility of the results as well as mitigating common cognitive fallacy in dependence assessments. Such a methodological robustness is of particular importance when the elicitation results are used in prioritising future investments of antibiotics research and development.

7.1 Chapter introduction

Decision and risks analysis methods that support medical decision-making (MDM) under uncertainty have gained attention lately due to an increased emphasis on evidence-based medicine and patient-centered care [186, 129]. Nevertheless, uncertainty in healthcare decision problems has been common for much longer and more subtly, for instance across the components of (even regular) medical procedures, such as disease definition, diagnosis, procedure selection and outcome

[1] Based on: Werner, C., Bedford, T., Colson, A. and Morton, A. (to be submitted). Risk assessment of future antibiotic resistance - eliciting and modelling probabilistic dependence between multivariate uncertainties of bug-drug combinations, Working Paper

observation given that a clinician is typically confronted with several choices in each of these steps [128].

A recent topic, which frequently dominates the scientific media and that shows the implications of little guidance on how to deal with uncertainty in medical procedures, is the emergence of antibacterial resistance. The widespread prevalence of antibiotic resistance-elements in bacterial pathogens poses a great concern for a variety of stakeholders, including clinicians, researchers, policy-makers, the pharmaceutical industry as well as the public. It is often attributed to mis- and overprescription of antibiotics by clinicians together with patients' failure to follow the treatment course or even patients' expectation to receive antibiotics [92, 57]. A further commonly stated reason is the prevalence of antibiotics in meat products through its frequent use in livestock [401]. The complexity of antibacterial resistance poses a particular challenge in this regard. For various bug-drug combinations we cannot regard future resistance rates as independent from another given that this severely underestimates the risk of multidrug or even pan-resistance (the latter being the resistance to all available antibiotics). Hence, a better understanding of the relationships between bug-drug pairs' resistance rates is crucial for clinicians' guidance on drug prescription as well as drug discovery which, if successful, is the main antidote to antibacterial resistance. Investing into second line drugs to which a bug is not or only merely less resistant severely undermines the overall risk and an antibiotics portfolio diversification.

Indeed, complexity is regarded as one of the main sources of uncertainty in MDM [186]. We therefore focus on probabilistically modelling the dependence between future resistance rates with the aim of providing a framework to better understand and model the risk of future antibiotic resistance and establish the value of its future research. A common challenge when modelling dependence probabilistically is the lack of relevant historical data for quantifying a model. Whenever in addition simplifying assumptions, such as independence, are not justifiable, eliciting dependence information from experts is the most sensible approach. A formal and structured process to dependence elicitation supports transparency and reproducibility of the expert judgement study, mitigates experts' potential cognitive fallacies and ensures confidence in the model result. These are important desiderata of an elicitation process given prevalent misconceptions about the concept of probability in MDM despite the frequent presentation and communication of health risks in probabilistic form [138, 101].

Structured expert judgement methods have been used previously in several areas of MDM (even though only univariate uncertainties were considered), for instance in health technology assessment [373, 49, 178, 374, 48, 254], assessment of surgery effectiveness [76] or as well modelling the risk and efficacy of treatment types [381]. In the following, we present a structured process for eliciting dependence information from experts and we use the elicited assessment to quantify a (dependence) model. In addition to outlining the main elements of our elicitation process, we present an illustrative case-study in which it has been applied for quantifying the dependence between the uncertainties of certain bug-drug pairs' resistance rates. Such an application is of particular relevance for informing policy-making and guidance-setting in antibiotics research and prescribing. In the next section, we briefly provide an overview on multidrug resistance, the risk which poses the main motivation for the illustrative case-study. In section 7.3, we present our method for eliciting dependence from experts, structuring their knowledge about it and modelling it. The results of the elicitations together with the resulting models are given in section 7.4. In section 7.5, we briefly present how our method can inform policy-making and guidance provision before concluding the chapter.

7.2 Dependence between uncertainties of resistance rates through multidrug resistance

In this section, we briefly discuss the importance of considering multidrug resistance and the reasons for it. This underlines the question of why a robust risk assessment should account for dependence between resistance rates' uncertainties.

Currently, around $700,000$ deaths related to antibiotic resistance are recorded annually with a potential increase up to 10 million by 2050, a future scenario where all known antibiotics are ineffective [92]. While such an "antibiotic apocalypse" might be regarded as media hysteria or even scaremongering by various researchers in the field, the phenomenon of multidrug resistance together with its future impact on healthcare is evident (see e.g. [92, 224, 387]).

The first bacteria that became resistant to multiple drugs were enteric ones, such as Escherichia coli, Shigella and Salmonella. Their multidrug resistance has been observed first in the late 1950s to early 1960s. Nevertheless, at this time such discoveries did not lead to changes in health policy-making as they had been regarded as of little concern. This changed however a decade later when Haemophilus influenza and Neisseria gonorrhoeae, respiratory and genitourinary causing organisms, started to develop resistance to multiple drugs. From then onwards, multidrug resistance and even scenarios of pan-resistance have come to the fore of various stakeholders. For a more extensive historical overview on multidrug resistance, see [257].

Today, the topic of multidrug resistance is particularly relevant for global examples of hospital and community strains such as Klebsiella penuemoniae and Pseudomonas aeruginosa isolates, even though the case of MRSA (methicillin-resistant Staphylococcus aureus strains) most arguably attracted the most public attention [92]. Other occurrences of multidrug resistance pose a risk to certain regions more specifically. For instance, while we focus in this chapter explicitly on Escherichia coli and Klebsiella pneumonia resistance to Fluoroquinolones in the UK, it is estimated that in Southeast Asia and China about 60 to 70% of Escherichia coli are resistant to Fluoroquinolones in addition to five other drug families [257].

Bacteria can accrue multiple resistance determinants through the long-term use (> 10 days) of a single antibiotic as within this time-frame the antibiotic will choose not solely for bacteria that are resistant to itself but also other ones. Given such continued antimicrobial selection, resistance not only increases for the drug taken by a patient but also structurally unrelated drugs. As such, several observations on the linkage of different resistance genes have been made [257]. Other influencing factors on the dependence between resistance rates of bugs to various drugs are ecological ones as can be seen by the disparity between resistance rates in hospitals and local communities. Here, the density of the antibiotic usage impacts not only the individual users but also others sharing the same environment [408, 256]. A remark in this context make [98] who emphasize that high resistance rate correlations among humans and animals do not pose evidence for causation. This remark is of particular importance when analysing the resulting of our elicitation together with the experts' rationales.

7.3 Methods for dependence elicitation and modelling of multidrug resistance

After having presented briefly the implications of and reasons for multidrug resistance and hence having introduced the motivation for eliciting (and modelling) dependence between future antibacterial resistance rate uncertainties from experts, in this section we present the main components of our structured expert judgement process for achieving robust elicitation results about multidrug resistance. Some components of the overall process are only briefly discussed, such as training the experts and choosing the form of the elicited variables. The main foci are on eliciting the marginal distributions, structuring experts' knowledge about the dependence between resistant rates, eliciting and combining assessments as well as modelling the dependence. For an extensive, general discussion on the different components of dependence elicitation processes, we refer to [415].

7.3.1 Eliciting the marginal distributions of future antibacterial resistance rates

Before we can elicit dependence information from experts, we have to specify the marginal distributions of the variables of interest, i.e. the future resistance rates of the bug-drug combinations of interest. This is necessary as otherwise the experts condition their dependence assessments on different quantile values, e.g. on different median values, which would make a sensible aggregation of their assessments impossible [415]. While in some cases we can quantify the marginal distributions through relevant historical data, in other dependence elicitations (such as the one presented here) this information also has to be obtained from subject matter experts. In our case this is due to the predictive nature of the marginal uncertainties given that we are concerned with future resistance rates of bug-drug combinations, such as Klebsiella pneumonia and Escherichia coli isolates resistant to Carbapenems and third generation Cephalosporins, in the UK in the year 2021.

The elicitation method used for eliciting the marginal distributions is the Classical model [330, 79]. Briefly, in this method the experts are asked to provide their assessments over continuous quantities. However, rather than only providing a single point estimate, they give a number of quantiles in order to capture the quantities' uncertainty distributions. For the marginal antibacterial resistance rates, these were the 5^{th}, 25^{th}, 50^{th}, 75^{th} and 95^{th} quantiles. The 5^{th} quantile is the number for which the expert thinks that there is a 5% chance that the true value is below this value and a 95% chance that it is above. We can interpret the other assessed quantiles similarly whereas for the 50^{th} quantile assessment there is a 50% chance for the true value to be below it and a 50% to be above it and so on. Usually, we start the elicitation by asking first for the 5^{th} and 95^{th} quantiles before eliciting more central quantiles. This might mitigate the anchoring effect of subsequent assessments around some more central, previously assessed values, such as the median.

We ask the experts to assess two types of questions. The first are calibration (or seed) questions and the second are the actual target questions. The former assessments are used for combining the experts through linear pooling whereas each expert's weight is performance-based. Thus, the answer for each calibration question is known to the analyst or will be known within the time fame of the research, but it is not known to the experts at the moment of the elicitation. An expert is regarded as a good probability assessor if the provided assessments for the calibration questions capture the true values with the correct long run relative frequencies (statistical accuracy) while the given distributions are relatively narrow (informativeness). Statistical accuracy means that for a large number of questions half of the true values fall above the median quantiles and half fall below. 90% of the true values fall within the given 90% interval (from the 5^{th} to the 95^{th} quantile) and 50% of true values between the 25^{th} and 75^{th} quantiles. Informativeness on the other hand takes into account how concentrated the given distributions are with respect to a chosen background distribution. For determining the overall weighting scheme, statistical accuracy is more important than informativeness, i.e. non-informative but statistically accurate assessments are still useful as this result might entail an important message regarding the overall uncertainty in the field.

7.3.2 Eliciting and modelling the dependence between future antibacterial resistance rates

Various ways to model dependence probabilistically can be used, whereas each offers certain modelling conveniences and captures specific aspects of a dependence relationships. For example, graphical models, such as Bayesian (belief) nets [317], allow for a high involvement even of non-statistical experts in determining the (structural) dependence relationships. Other dependence models, such as regression models include (so called) a set of auxiliary variables, which are not directly part of the model variables but allow for an easier quantification of their uncertainty. [412, 415] discuss the choice of dependence models in certain modelling contexts and their influence on the expert judgement method in more detail.
In this chapter, we aim for deriving a parametric copula which satisfyingly captures the experts' assessments. This model choice is due a copula's convenience that the dependence relationship can be addressed separately from the marginal distributions and that it allows for modelling tail dependencies. Numerous commonly used parametric copulas can be grouped either into the Elliptical copula class or the Archimedean one. Copulas in the former group are radially symmetric, i.e. their upper and lower tail dependence is the same, whereas copulas in the latter group are asymmetric, meaning that they can capture upper or lower tail dependencies explicitly. In addition to modelling the overall dependence, tail dependence is an important feature of a joint distribution to capture as neglecting it can lead to little understanding of the overall risk and hence poor decision-making when events evoking extreme values appear. The case of multidrug antibacterial resistance (7.2) highlights the importance of identifying potential upper tail dependencies as the spread and increase of resistance affecting multiple drugs shows a possibly prevalent systemic risk. In other words, it might not be desirable to model that a high resistance rate increase for a bug to one drug is followed by a similarly high decrease in resistance for another drug, so that antibacterial resistance is balanced out, which would be given in a symmetric dependence.
For an introduction to copula theory and advanced discussions, we refer to [302, 208] and [125]. Briefly, recall that the random variable $U = F_X(X)$ of a continuous random variable X with distribution function $F_X(X)$ is uniformly distributed and the same is true for Y. In our context, the variables X and Y correspond to the rates of resistance for different bug-drug pairs, e.g. Klebsiella pneumonia isolates resistant to third generation Cephalosoporins (X) and Kleb-

siella pneumonia isolates to Carbapenems (Y) in the UK in the year 2021. With two random variables X and Y, the distribution of the vector $(F_X(X), F_Y(Y))$ is then supported on the unit square with uniform marginals. Such a bivariate distribution is called a copula, and its construction can be reversed, so that any set of univariate distribution functions with a copula represents a multivariate distribution. Formally, upper tail dependence (which is of interest here) is defined as:

$$\lambda_U(X,Y) = \lim_{u \to 1^-} P(Y > F_Y^{-1}(u) | X > F_X^{-1}(u)) \tag{7.3.1}$$

when a limit $\lambda_U \in [0,1]$ exists. Thus, $\lambda_U > 0$ indicates upper tail dependence whereas for $\lambda_U = 0$ the distribution tails are independent. Loosely formulated for us here, in the case of upper tail dependence it is more likely to observe high resistance rates for X given high resistance rates for Y. When mapping the experts' assessments onto parametric copulas with given rank correlations, we use the upper tail dependence coefficient $\lambda_U = (1 - 2u + C(u,u))/(1-u)$ to distinguish the various forms (see Figures 7.5 and 7.6 when discussing the results in 7.4). We refer to the tail dependence coefficient when discussing the results of the dependence elicitation.

Structuring experts' knowledge about the dependence relationships through conditional scenarios

Assessing dependence, e.g. in the form introduced in the next sub-section, can be cognitively challenging for experts or even counter-intuitive. As a result, experts might be prone to cognitive fallacies while, more generally, they can struggle to incorporate their knowledge on complex dependence relationships in a quantitative assessment. Therefore, in order to mitigate cognitive fallacies, enhance the understanding of the elicited form and allow for managing and sharing knowledge, we use a method for mapping conditional scenarios [414] prior to the quantitative assessments. Before presenting our method however, we briefly explain why dependence assessments can be cognitively demanding and more precisely which cognitive fallacies are common when assessing conditional probabilities (the choice of eliciting dependence in a conditional form is discussed in more detail below).

While it is common to conceptualise probabilistic dependence through conditionality, studies show that not only specific cognitive fallacies can easily occur but that understanding and interpreting conditional forms is (still) a challenge in today's statistics and probability education [114]. [63] remark in this regard that a main focus of education is on frequentist approaches to probability together with (idealised) random experiments. For understanding and conceptualising conditional probabilities, such a position is however not easily adopted for real world random phenomena. Nevertheless, the subjective view on probability, in which a conditional probability is more intuitive as one simply revises judgements given new information that has become available, is not commonly taught in curricula of numerous countries [53].

Some specific cognitive fallacies, which are of particular relevance for conditional probabilities, are confusion of the inverse [272, 127, 193], confusion of joint and conditional probabilities [134] and the causality heuristic [8]. An extensive introduction to these and other cognitive fallacies for assessing dependence can be found in [415]. For an overview on heuristics and biases in the area of MDM, see [46]. Briefly, the aforementioned fallacies can affect dependence judgements for multidrug resistance as follows.

The confusion of the inverse states that experts might confuse an elicited conditional probability $P(X|Y)$ with its inverse $P(Y|X)$. In our case, this can happen if experts might be unsure (or disagree) about the order in which the first and second line drugs are used, or if they find it easier to condition on the resistance rate of a second line being above a certain threshold.
The confusion of joint and conditional probabilities is often explained by the semantic misinterpretation of "and" which might be interpreted by some expert as an indicator of temporal order. As such, they assess a conditional probability instead of an elicited joint one. Similarly, we have anecdotal evidence [414] that this fallacy can also work in the opposite direction, thus experts assess a joint distribution instead of a conditional one. In the case of dependent resistance rates, (unwillingly) assessing the joint probability between two bug-drug combinations instead of a conditional probability can result in a severe underestimation of the multidrug resistance risk as an expert would typically assess a joint extreme event with a very small probability while in a conditional scenario systemic risks actually increase the extreme event probability.
Lastly, the causality heuristic refers to experts' mental (causal) models about the dependence relationships and states that a conditional probability is assessed higher if a causal explanation underlies the conditional dependence relationship, even though such as causal explanation has no role in probability theory. As shown, when discussing the results of the elicitation together with the experts' conditional scenarios underlying their assessments, various causal and non-causal explanations for their assessments are given whereas these should have equal influence on the assessment result rather than assessing higher probabilities when more causal explanations are given.
As aforementioned, a way to mitigate these cognitive fallacies and improve experts understanding of the conditional judgements is through structuring experts' knowledge prior to the quantitative assessments. For that, we use a conditional scenario mapping method introduced in [414]. Scenario are defined as "sequences that link triggering events to specified consequences or final states through intermediate conditions". Thus, the experts were first presented with the final state for the event that we condition onto later, i.e. a bug's resistance to third generation Cephalosporins being higher than either its 50th quantile or 95^{th} quantile. Then, they reasoned through backwards logic what the different reasons for this final state are, taking into account events from today to the year 2021. The events that are observable today are then classified as trigger events and "imported" into the probability space of the conditional distribution, i.e. a bug's resistance to Carbapenems being above either its 50^{th} or 95^{th} quantile. Taking these imported trigger events, the experts then reasoned in forwards logic how the conditional scenarios evolve up to the year 2021 while determining the conditional probability based on the number of relevant trigger events for both, the unconditional and conditional probability space. The conditional scenarios, which results are shown in Section 7.4, allow experts to reflect on their knowledge about the dependence relationships, clarify the inverses and any causal versus non-causal factors, and can be shared to challenge each expert's previous understanding of the assessments.

Eliciting dependence between the future resistance rates

The following elicitation process might be regarded as a pragmatic approach to modelling dependence as the final (copula) model is only based on a low number of assessments. Nevertheless, we argue that the following elicitation method and the resulting model offer a good balance between assessing detailed dependence information on the one hand, e.g. it captures the behaviour of the random variables in the extreme parts (tails) of their joint distribution, and on the other hand ensuring a low cognitive complexity for experts. In particular, eliciting too many variables easily leads to infeasible and incoherent assessments. The part of the elicitation process that focuses on eliciting the target variables is briefly described in the following steps:

- We elicit the conditional median in the form of $P(Y > 50^{th}\ quantile|X > 50^{th}\ quantile)$ for the variables of interest X and Y. For our exemplary bug-drug pair this can be framed as: *"For the year 2021, given that in the UK the rate of Escherichia coli isolates resistant to third generation Cephalosporins is higher than* 16.21% *[*50^{th} *quantile], what is the probability that the rate resistance of Escherichia coli isolates to Carbapenems is higher than* 1.996% *[*50^{th} *quantile]?"* (see Figure 7.1 on the left).
- As an intermediate step, we then vary the conditional variables, typically to explore the joint distribution tail, i.e. more extreme scenarios more explicitly. For that, we elicit the conditional probability of $P(Y > 50^{th}\ quantile|X > 95^{th}\ quantile)$. In our example this is asked for by: *"For the year 2021, given that in the UK the rate of Escherichia coli isolates resistant to third generation Cephalosporins is higher than* 38.59% *[*95^{th} *quantile], what is the probability that the rate resistance of Escherichia coli isolates to Carbapenems is higher than* 1.996% *[*50^{th} *quantile]?"*
- In a similar way, we now vary the other quantile, so that both refer to the joint distribution's tail, i.e. we elicit $P(Y > 95^{th}quantile|X > 95^{th}quantile)$. We frame the question as: *"For the year 2021, given that in the UK the rate of Escherichia coli isolates resistant to third generation Cephalosporins is higher than* 38.59% *[*95^{th} *quantile], what is the probability that the rate resistance of Escherichia coli isolates to Carbapenems is higher than* 17.24% *[*95^{th} *quantile]?"* (see Figure 7.1 on the right).
- With the above assessments, we can now compare each expert's judgements with different parametric copula forms. This is done by plotting the assessments against the converging conditional exceedance probabilities for selected parametric copulas simulated at the u^{th} quantile from 0.5 to 1 through the tail concentration function. Figures 7.5 and 7.6 show the comparison of parametric copulas at the 50^{th} and 95^{th} quantiles for various rank correlations. The copula choices and rank correlations can be varied for approximating the assessments better.
- With a first idea of which copula represents an expert's information reasonably well given a specific rank correlation, we can test the robustness of that choice, e.g. by "feeding back" the probabilities for non-elicited quantiles and check an expert's agreement for it.

All assessments are made in a conditional probability form as we regard it as more intuitive than eliciting other dependence parameters. For instance, joint probabilities are cognitively difficult to conceptualise and understand by experts [73] while correlation coefficient are only reliably assessed for dependence relationships in which experts get frequent feedback [432]. Further, we can use the scenario mapping method for conditional dependence relationships as shown in the previous sub-section.

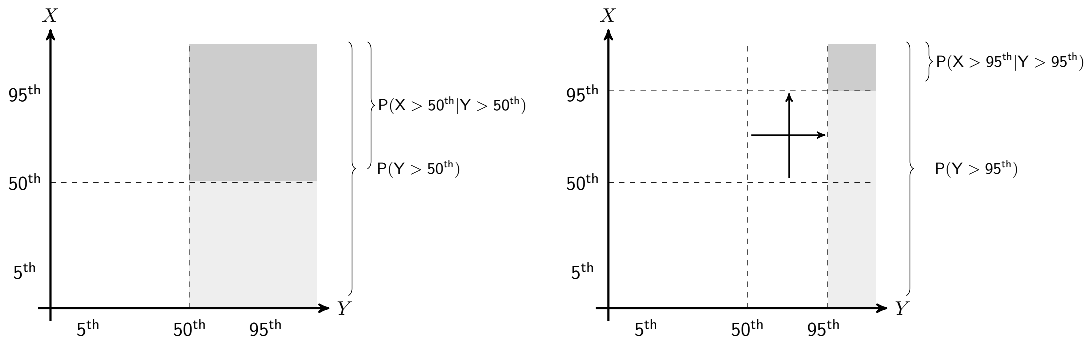


Figure 7.1: Schematic representation of elicitation sequence

Aggregating dependence assessments from various experts

In many expert judgement studies, it is of interest to elicit assessments from more than one expert in order to capture a broader range of knowledge and beliefs about the uncertainties of interest. Whenever we elicit more than one assessment, an important decision concerns aggregating several judgements in a sensible way. While some methods are based on behavioural aggregation approaches, thus these methods advocate that experts should achieve a consensus opinion for the variable of interest, other methods combine experts' judgements mathematically. Mathematical aggregation of expert judgements can avoid common shortcomings of behavioural aggregation methods, such as one expert dictating the final elicitation outcome (e.g. due to a strong personality and/or power relationships between the experts) or as well group-think, i.e. experts try to avoid discussion and conflict about elicited result [155]. Therefore, we only consider mathematical aggregation methods in the following. More specifically, we construct several combined assessments through different linear pooling methods. This allows to compare the different results and feed these back to the experts as the topic of combining dependence assessments is less well explored than it is the case for univariate uncertainty.

Equal Weighting Another way of combing expert judgements is simply through the weights determined by the earlier calibration questions for the marginal distributions. In other words, we assume that each expert's performance on the previous calibration questions together with the resulting weights not only reflects their ability to assess marginal probabilities, but to make assessments more generally, also about probabilistic dependence. While some research [291, 289] suggests that experts who perform well with the Classical model cannot be regarded as good dependence assessor, we remark that these are indicative results which is why we include a linear pool weighting scheme based on the previous calibration questions.

Performance-based weighting: Dependence-calibration score The last aggregation method requires the elicitation of dependence calibration variables in addition to the actual target variables. Experts' assessments are then combined based on a dependence calibration score which is introduced by [294] while an extensive introduction and discussion is given in [291] and [412, 415]. For that, the Hellinger distance measures the distance of experts' assessments and the actual rank correlation values. An information score as with Cooke's model is not derived.

In total we elicited nine calibration questions which are of two different types. This has the advantage of considering experts' performance from different perspectives, and (as in our case) it simply allows for including more calibration variables. The latter can be relevant as (in our experience) finding suitable calibration variables for dependence relationships might be challenging given that association measures, e.g. rank correlation coefficients, which then constitute the true values require more historical data to be relevant than this is the case for univariate uncertainties.
Our first type of calibration questions is: *"Given that in the UK in 2015, the rate of resistance for Escherichia coli isolates to third generation Cephalosporins was above its median of* 10.09% *(2010-2015), what is the probability that the rate of resistance for Escherichia coli isolates to Carbapenems was also above its median of* 0.433%*?"*
The rank correlation coefficients for the above framing are based on the dependence over time, i.e. the resistance rates per year for 2010 to 2015.
For the second type of calibration question, we ask: *"Given that the UK in 2015, was below the European (Italy, Spain, France, UK) median of* 25.43% *for the rate of resistance of Escherichia coli isolates to third generation Cephalosporins, what is the probability that it is also below the median (*1.23%*) for Escherichia coli isolates to Carbapenems?"*
The correlation coefficients for the second type of calibration questions considers the dependence over various countries but within the same year.
Note that we ask for conditional probabilities in the calibration questions even though the historical data are given as rank correlations. This is to keep it in the same form as the target variable elicitation for which we already justified the choice of conditional probabilities in Section 7.3.2. Nevertheless, we now need to transform the experts' assessments from conditional probabilities into (Spearman's) rank correlations. More technical details on this transformation together with an introduction on how to derive this transformation for other quantile values is given in [290].

7.4 Results of dependence elicitation for multidrug resistance rates

After having presented the main components of our structured dependence elicitation process, in this section we present the corresponding results. In accordance with the previously presented calibration and target questions, in the following we show the results for the dependence between Escherichia coli isolates' resistance to third generation Cephalosporins and its resistance to Carbapenems.
In total, six experts participated in the elicitation of the marginal distributions and four (of them) participated in the dependence elicitation. All experts have a broad subject matter expertise ranging from clinicians' day-to-day experience of prescribing antibiotics to research involvement in new antibiotics' development. We elicited the marginal distributions and dependence from each expert in separate expert judgement sessions. Figures 7.2 and 7.3 show the elicited marginal distributions for the resistance rate in 2021 of the bug-drug pairs of Escherichia coli isolates to Carbapenems and Escherichia coli isolates resistant to third generation Cephalosporins accordingly. In addition to each expert's individual assessment, we show the performance-based combination according to the Classical model (DM) and the equal weight combination (EW).

We observe that for resistance to Carbapenems, all experts apart from Expert

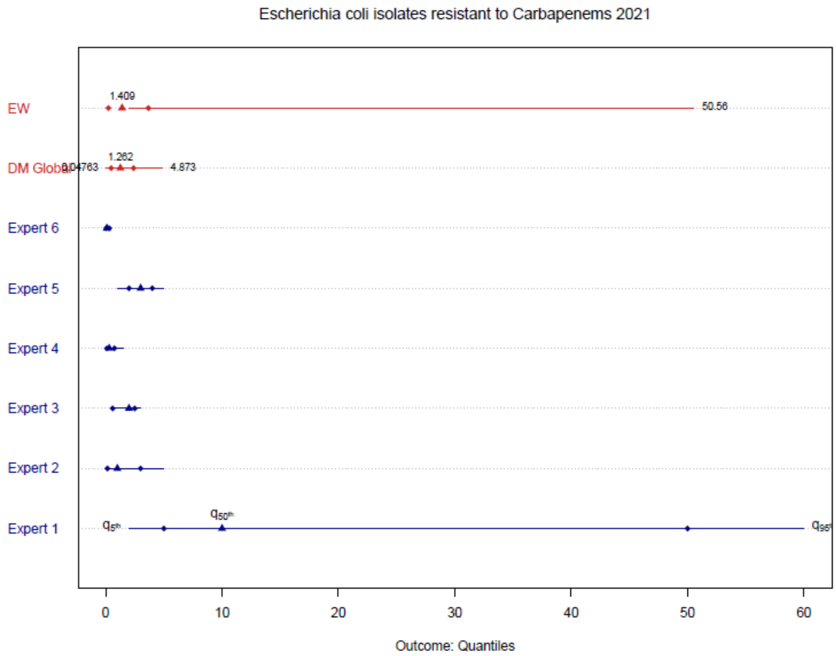


Figure 7.2: Elicited marginal distribution (5^{th}, 25^{th}, 50^{th}, 75^{th} and 95^{th} quantile) for the rate of resistance of Escherichia coli isolates resistant to Carbapenems in the UK in 2021.

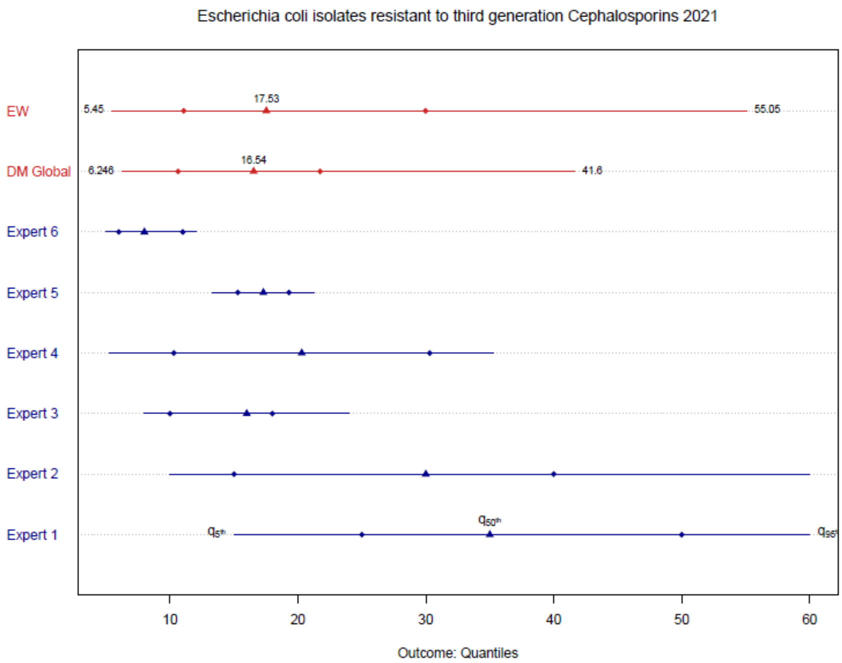


Figure 7.3: Elicited marginal distribution (5^{th}, 25^{th}, 50^{th}, 75^{th} and 95^{th} quantile) for the rate of resistance of Escherichia coli isolates resistant to third generation Cephalosporins in the UK in 2021.

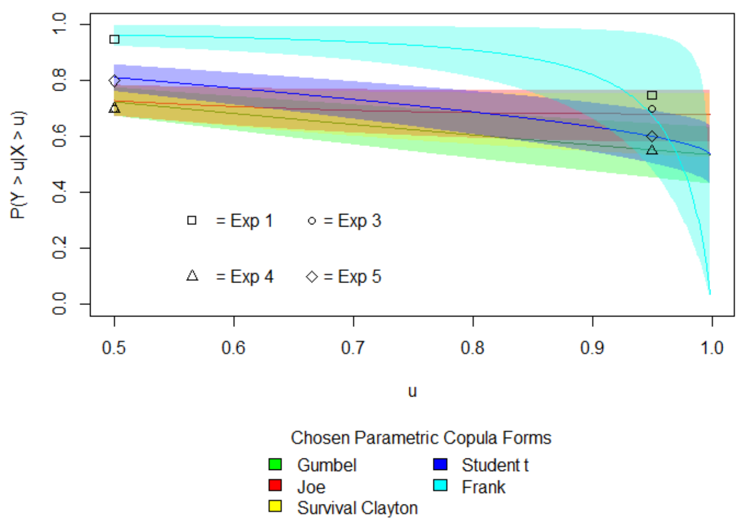


Figure 7.4: Elicited conditional exceedance probabilities for quantiles $u = 0.5$ and 0.95 with fitted copulas per expert.

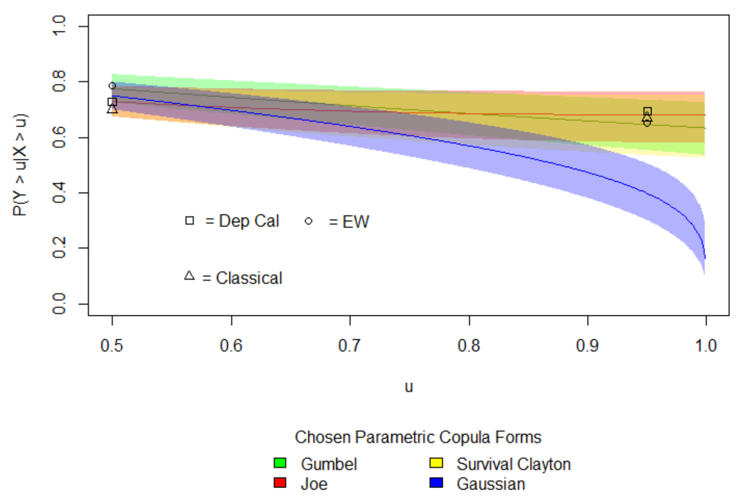


Figure 7.5: Elicited conditional exceedance probabilities for quantiles u=0.5 and 0.95 with fitted copulas per combined weighting.

1 assess narrow distributions with medians which are close to each other. This is reflected in both combinations through similar median values. Nevertheless, the DM is much more informative than the EW combination as Expert 1 receives less weight in the former. For the resistance to third generation Cephalosporins, the assessments are less in agreement and we therefore obtain a much wider DM distribution even though it is still more informative than the EW combination. Based on the elicitation results for the marginal distributions, we then elicited the dependence between both resistance rates according to the procedure in Section 7.3.2. Figure 7.4 shows each experts' assessment about the dependence between the future resistance rates of the bug-drug pairs together with some fitted parametric copula choices.

Similarly, Figure 7.5 presents the combined assessments (for each of the different aggregation methods) again together with the best-fitting copulas. We observe that the experts differ considerably in that two experts' best fitting copulas are symmetric (Expert 1's assessments result in a Frank copula with rank

Table 7.1: Individual and combined multidrug dependence assessment results.

Expert/Combination method	$> 0.5 \mid > 0.5$	$> 0.95 \mid > 0.95$
Classical model (marginal seeds)	0.7979	0.6711
Equal Weighting)	0.7875	0.65
Dependence Calibration	0.7295	0.6949
Expert 5)	0.8	0.6
Expert 4	0.7	0.55
Expert 3	0.7	0.7
Expert 1	0.95	0.75

correlation 0.9 and Expert 5's assessments correspond to a Student t copula with rank correlation 0.6 the shaded area adds ± 0.1) while the other two experts' assessments fit asymmetric copula with upper tail dependence (Expert 3's assessments correspond to Joe copula with rank correlation of 0.45 and Expert 4's assessments match a Gumbel copula with rank correlation of 0.45). In contrast to these individual differences, we see that the combined assessments for all aggregation methods result in copulas with upper tail dependence (Joe and Survival Clayton copulas with rank correlation 0.45 for the performance-based weighting scheme and a Gumbel copula with rank correlation of 0.55 for equal weighting). We observe that tail independent copulas, such as the Gaussian copula, do not fit well. Table 1 summarises the assessment results.

The results of the aggregated assessments are supported by the experts' combined rationales. Each expert's assessment has been supported by the conditional scenario mapping method introduced in section 7.3.2. We then combined all rationales and fed these back to the experts for sharing the overall knowledge and giving them a possibility to adjust their assessments. While no expert modified their assessments, all agreed on these meta conditional scenarios. Figure 7.6 and 7.7 show the unconditional and conditional scenarios for the previously discussed bug-drug combinations. We can see which trigger events for the unconditional probability space are impacting the conditional one together with the sub-set of events which are still relevant for the extreme (95^{th} quantile) scenarios (in the red shaded areas). For instance, we observe that a main scenario for upper tail dependence between Escherichia coli resistance to third generation Cephalosporins and Escherichia coli to Carbapenems is due to the potential emergence of a new strain. This leads to the target-oriented policy decisions across the National Health Service (NHS), which aim to reduce the prescriptions of Carbapenems and third generation Cephalosporins, not to be working anymore. Other scenarios supporting upper tail dependence consider trigger events that lead to a higher beta-lactam antibiotics' use or as well bottlenecks in the raw material supply chain. A scenario, against possible tail dependence, is that Carbapenem use might decrease until 2021 which then leads to an increase in the use of third generation Cephalosporins.

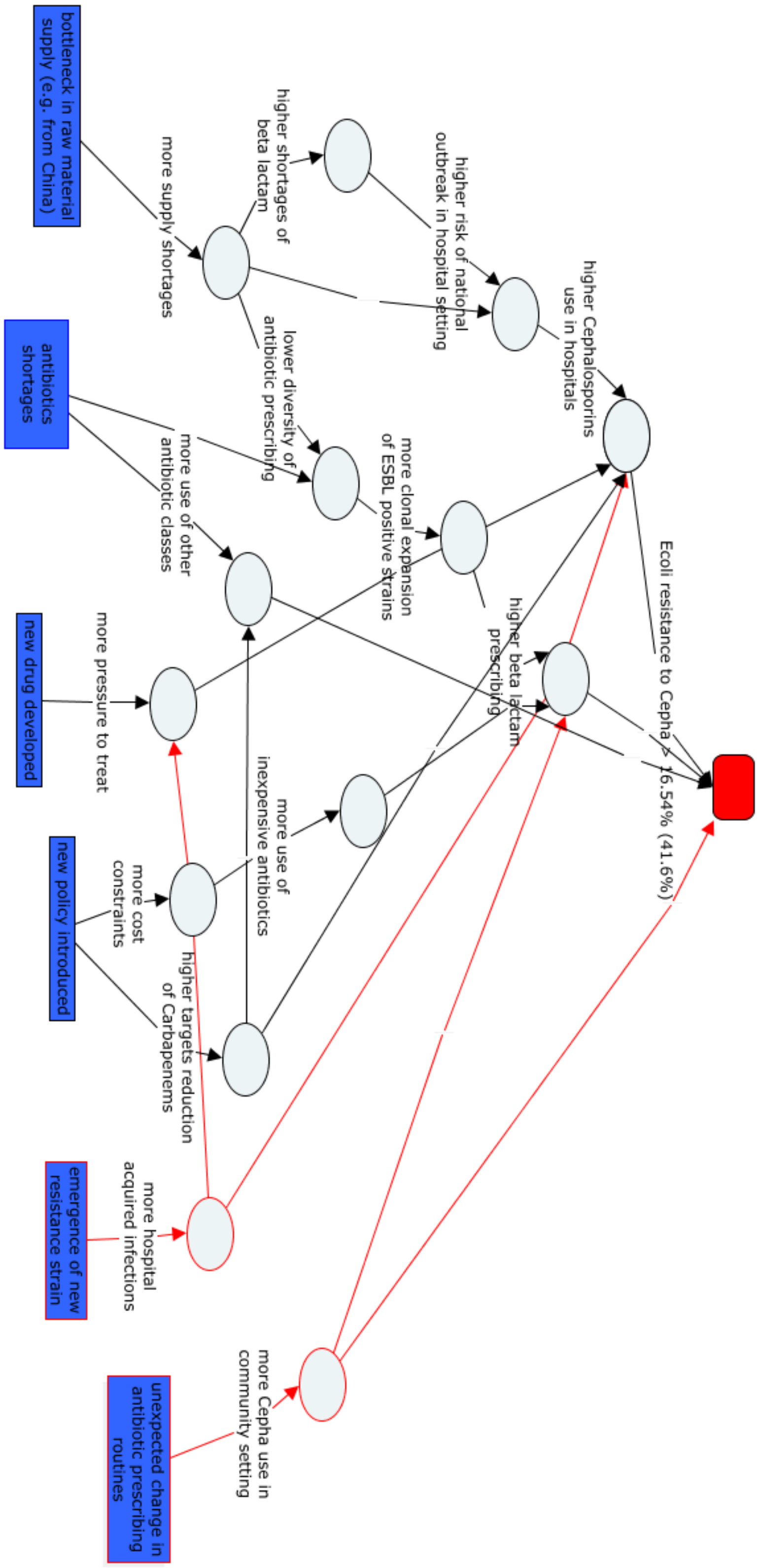


Figure 7.6: Unconditional meta scenario of all experts for Escherichia coli resistance to third generation Cephalosporins in the UK in 2021.

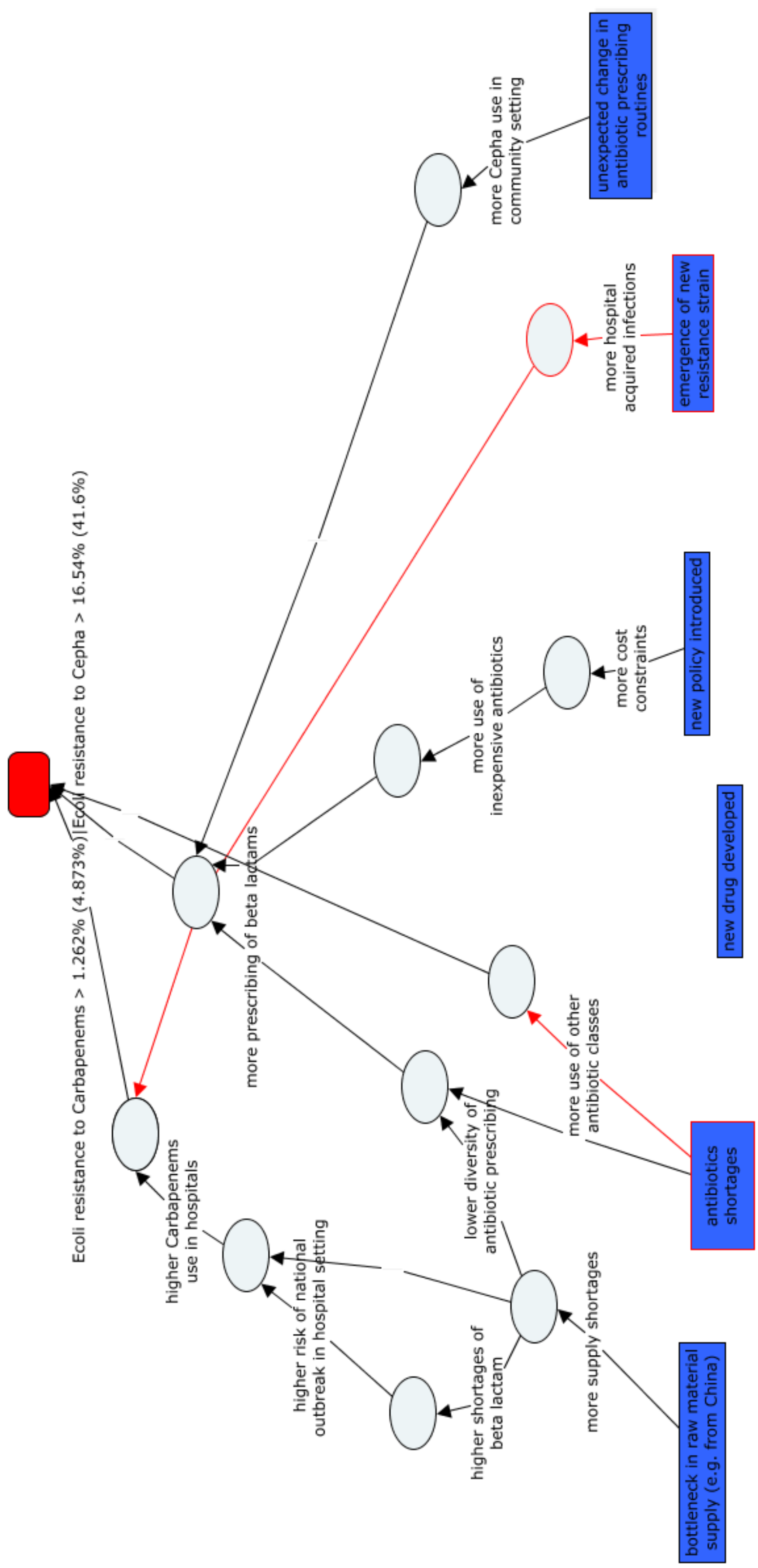


Figure 7.7: Conditional meta scenario of all experts for Escherichia coli resistance to Carbapenems in the UK in 2021.

7.5 Discussion on informing the value of future antibiotics' R&D and chapter conclusions

In this chapter we have addressed the challenges of using quantitative risk assessment techniques to model future antibacterial resistance to multiple drugs. Given the lack of relevant historical data for quantifying a (dependence) model in this context, we have introduced a structured expert judgement process for eliciting dependence between future resistance rates uncertainties. Proposing prescriptive decision aiding methods is in line with the trend that decision models are gaining importance and acceptance as formal methods to inform health policy making [209].

A main result of our case-study is the identification of upper tail dependence, indicating a potential future systemic risk which negatively affects the usefulness of some common first and second line drugs against their corresponding bugs. A main motivation for our case-study (as mentioned in the chapter introduction) is to offer a method that informs decision-makers who manage a portfolio of antibiotics and make decisions about adding new ones to it through R&D investments. In the following, we briefly discuss how our findings and structured dependence elicitation more generally can be used in medical decision-making informing policy-making.

Referring to the term "concern-driven risk management", [97] criticises the World Health Organization [418] for basing their recommendations and guidance to identifying critically important antibiotics and thus prioritising R&D activities on qualitative levels of concern. This means that risk-based decisions are taken if a regulatory agency "is sufficiently concerned about risks from current human behaviours" (such antibiotic use) rather than considering quantitative modelling methods. However, past experience shows that such concern-driven recommendations can have devastating consequences, such as happened in Denmark and the European Union where animal illnesses and mortality surged after banning five animal antibiotics. While a quantitative risk model might lack relevant data for informing such decisions, our method provides a way of addressing such modelling challenges and provide evidence-based decisions.

Another example in this regard, in which omitting the possible effects of dependence between resistance rate uncertainties were of particular importance, is the guidance given by the U.S. Food and Drug Administration's Center for Veterinary Medicine [305]. [93] remarks that due to not considering any possible correlations, no sensible risk management recommendations can be given. Again, using a structured expert judgement process for specifying uncertainties and their dependence improves risk management decisions and thus impacts positively the current value of information for quantifying and comparing the economic consequences of different actions. As such, using formal uncertainty assessment methods when determining medical innovations and future research needs is crucial [349].

Part IV

General Conclusions and Bibliography

Chapter 8

General Conclusions

Within this thesis, I have made various contributions to the research on eliciting dependence from experts which is an important area of the decision and risk analysis literature. This concluding chapter summarises my research results by reflecting on the research objectives of this thesis together with the connection between the presented contributions/chapters, by re-addressing the research questions and by discussing the limitations of the findings presented here while proposing suggestions for further research.

8.1 Reflections on the research objectives and the connection between the research contributions

This thesis highlights the importance of using structured processes to elicit dependence information for probabilistic modelling from experts when relevant historical data are lacking. Research in this area can be approached in different ways. In this section, I therefore explain and reflect on how the different contributions (and hence chapters) presented in this thesis relate to and follow from each other. This helps clarifying how the research design adopted here helps achieving the research objectives.

In the introduction to this thesis, Figure 1.1 provides an overview on how the chapters, which are based on published (or to be published) contributions, are placed in a conceptual framework of prescriptive, supportive methods for decision and risk models. Chapter 2 is not shown as it outlines preliminaries and is of introductory nature rather than a research contribution. This framework is based on the discussion of section 1.4 about the proposed research methodology for prescriptive methods and models in decision and risk analysis. Similar frameworks can be found, for example in [75].

On the left hand side of Figure 1.1, the main elements of a decision/risk analysis process are shown at a (fairly) general level. As decision and risk analysts, we first need to understand the problem context. At such a general level, this involves observing the problem together with identifying as well as defining it. These steps are part of the element shown as a cloud shape in which usually the decision-maker is still involved. In this thesis, the (broad) problem context that we observe concerns dependence relationships between uncertainties. These occur in various application areas and can have unexpected consequences for decision outcomes if neglected by decision-makers. Therefore, together with the decision-maker we should first determine whether dependencies can be neglected or not, and if not, decide which attributes of dependence relationships are of interest for the decision. This determines how much detailed information is necessary for making a sensible decision and hence identifies and defines the problem to be modelled.

In the next step, we make choices on appropriate models that we consider suitable for modelling the decision/risk analysis problem at hand. These choices are often based on the needs and preferences of the decision-maker. For the topic of this thesis, this implies that after identifying and defining the dependencies between uncertainties, which might have an impact on the decision and are of interest for a decision-maker, we choose models that capture the dependence relationships' attributes of interest. In this regard, various dependence models exist whereas different model types address different aspects of dependence relationships. A model aspect, which is of relevance for several dependence relationships discussed, assessed and modelled in the reviews and case-studies of this thesis, and which is of importance for sensible dependence assessment in many application areas, is that of tail dependence. Often, we need to consider potential tail dependencies to better understand extreme risks. For these, we might observe that (in a bivariate relationship) conditional on being above a critical threshold for one particular variable, the other variable is also above a critical threshold. Further, we might also want to capture whether tail dependencies are either asymmetric or symmetric. For example, when considering monetary losses as risk factors, in the latter situation losses are balanced out by gains, whereas this is not the case for the former, i.e. with asymmetric tails [137].

At this point of the decision/risk analysis framework, we might enter a loop in the process of choosing various models and selecting the one(s) which fit(s) best in the problem context. However, a decision-maker might desire a different model with different or less assumptions and/or different capabilities (in terms of the produced output) and therefore, we need to alter the current model or select a different one. This part of model altering and selection might then be repeated several times. This is also known as the *Management Science process* [419]. For dependence models, it can be common to propose parametric, low-dimensional models first. They are easier to understand, quantify and use. However, due to their strong assumptions and potential restrictions, for instance in case of modelling tail dependence, the decision-maker asks then for a more flexible dependence model. For example, a model that captures tail dependencies and implies less parametric assumptions. In this way, we can enter a loop of refining the model choices with the decision-maker.

The above two steps are similarly applicable in other decision and risk problem contexts (than ones in which dependence modelling is directly of interest). For example, when applying forecasting/time-series methods (through historical data), we might start by proposing models that can be quantified with little data and are, at the same time, easily understandable. In this modelling area, some of the easiest models are naïve ones, which simply take the last observed value as forecast, or as well forecast through simple averaging of some observed past values [56]. Decision-makers however might require more sophisticated models as solutions to their problems.

Similarly in simulation-based decision and risk analyses, such as when applying discrete event simulation, we might choose simple models first. Then, in a loop-like process, we develop (in agreement with a decision-maker) more sophisticated simulations. These either incorporate more detail or use a multi-methodological approach to represent sub-systems (which provide the input variables of certain model parts) more accurately. Common examples are simulations which are

primarily process-oriented but apply an agent-based simulation for certain parts of the model as these might require a different level of detail (see e.g.[233]).
Once we have chosen one or more appropriate model(s) that can provide the information and output required by the decision-maker, we need to identify the potential modelling challenges which are prevalent for the particular model(s).
Often, and as it is the case in this thesis, the modelling challenges in a decision and risk analysis context are related to the prescriptive purpose of a model. As aforementioned, prescriptive decision theory aims at offering guidance or procedures which allow for making optimal choices of decisions. This can be based on the difference between normative and descriptive decision theory, whereas prescriptive models differ to ones that seek to explain or describe a phenomenon of interest (see section 1.4 for a more extended introduction to the models considered in this thesis).
The prescriptive methods are then our specific tools and techniques which can be used to offer such guidance. Typically, we use an optimisation technique to identify the optimal solution most efficiently. However, we can also run scenarios (e.g. through Monte Carlo simulations altering model parameters) or use *trial-and-error* as ways to gain a better insights on our best choice in terms of which decision to make. Each technique has its own advantages and drawbacks and the choice is likely to be based again on the decision-maker's preferences. An optimisation technique can be most efficient (if a solution can be obtained in that way) but may constitute a *black box* to the decision-maker if only optimal solutions are presented without much more information on how it was obtained. Running scenarios or using trial-and-error on the other hand might be more time-consuming, but can offer a better familiarisation with the behaviour of a model and the impact of changes made to the model parameters.
In that way, the above techniques can be applied for identifying an optimal solution if the underlying model is a dependence model, either as a decision or risk model itself or if it is "plugged into" (and complements) another model. Similarly, we require some technique, such as the above ones, for simulation and forecasting models when they are part of a prescriptive analysis.
A modelling challenge to use optimisation or scenario-based techniques can be a lack of relevant historical data on the uncertainties involved in the decision problem and on the preferences for choosing certain decisions. Therefore, structured elicitation techniques for guiding uncertainty and preference assessments are key for making optimal decisions and according to [37], "the art and science of elicitation of values (about consequences) and judgments (about uncertainties) lies at the heart of prescriptive endeavours."
This statement is particularly fitting for dependence between uncertainties as it not only adds complexity to an assessment, but also because relevant historical data on dependencies are often scarce, even if relevant data on marginal distributions are available.
The above discussion on the broad decision and risk analysis context with a focus on its prescriptive purpose already shows how dependence models and hence the assessment of dependencies are part of a more general decision and risk analysis literature. Given that this thesis has contributed to this area (with a prescriptive focus), we can now outline where the contributions lie within such a general framework, how the research has been influenced by the steps of the framework and lastly how the contributions are related along this framework. This is done by re-considering the research design, methodology and objectives formulated in the introduction of this thesis (see sections 1.3 and 1.4).

Recall that our research design and methodology to achieve our objectives is based on two pillars of knowledge production, that of (1) research authors structuring knowledge and that of (2) users applying knowledge when solving problems (in the real world). We already clarified earlier why the latter is particularly relevant when contributing to prescriptive decision and risk analysis research while the former is a common prequisite for most research areas.
Nevertheless, in the research on expert judgement methods for dependence, this structured knowledge has not been given, apart from the few, less comprehensive overviews mentioned in chapter 3. Therefore, I decided to provide reviews on dependence elicitation as ways to offer such structured knowledge for researchers interested in dependence assessments first. This is in line with the first objective that is to propose a taxonomy for the current research on expert judgement methods for dependence assessment and set its future agenda.
In line with the framework of Figure 1.1, the first literature review (that of chapter 3) is motivated by the various dependence models and how the modelling choice (which occurs after having identified and defined the decision and risk problem) impacts the elicitation method. The elicited form is in this regard particularly influenced by the model choice and at the same time is a key decision for the elicitation. As such, chapter 3 addresses the upper part of the framework to bridge the gap between real-world problems, models and modelling challenges, which in this case is that of lack of relevant data and the question in which form to elicit it (considered as one of the most basic decisions for an elicitation).
The second literature review, chapter 4 in this thesis, takes a different starting point and addresses the lower part of the process, i.e. focussing on the prescriptive purpose of elicitation processes. As aforementioned, structured, prescriptive processes for supporting experts in making dependence assessments are crucial as they impact the optimal decision directly. Therefore, I provide a comprehensive review on how well the different elements of such processes have been explored in the research community.
After obtaining some structure for the knowledge in the research area, I then address the second research objective. I do so by proposing novel prescriptive methods, which address some of the identified research gaps, and test these in real-world decision and risk problems for ensuring their practical and pragmatic value. Hence, on the right of Figure 1.1, we can see that chapters 5 to 7 are the chapters that contribute to the second research objective. As can be seen, the research gaps addressed by the proposed novel methods are primarily identified from the literature on elicitation processes. However, I should note that when applied in case-studies and with specific models, the methods are also influenced by findings on elicitation in specific modelling contexts.
Following this reflection on the research contributions as part of the broader decision and risk analysis literature and while doing so, having clarified the chosen order of presenting these contributions in this thesis, I will now re-address and discuss the research questions proposed in the introduction.

8.2 Re-addressing the research questions

Research Question 1: *Which dependence models are most prevalent in the decision and risk analysis research and for which of these has expert judgement been used to address the lack of relevant historical data? What are the foci of these models and how can they be considered in assessment methods?*

Chapter 3 of this thesis, which is based on [412], offers a systematic review of the literature on dependence models for which expert judgement has been used for assessment purposes. In that, we include proposed structured expert judgement methods for specific models as well as non-structured elicitations of dependencies as it is for example often the case when conditional probabilities are elicited for BNs. The latter are considered in particular when surveying the case-study literature in which dependence has been elicited for models in real-world applications.

Not only, but also because of the recorded non-structured elicitations, BNs are the most widely used dependence models in the modelling literature that includes expert judgement by accounting for around 64% of models used. In addition to a general increased interest in BNs (which developed over approximately the last three decades since [317]) in the decision and risk analysis literature, this preference in model choice might also be attributed to the fact that with such graphical models experts are often "called in" for deriving the structure of the model and it might be then *natural* to do additional quantitative elicitations on the corresponding conditional probability table afterwards if data are lacking.

This outlines already that one main modelling aspect, which is of interest in the literature, concerns graphical dependence models. For these, experts can be used for the elicitation of structural information which in turn can clarify experts' quantitative assessments. We have therefore explored the idea of hybrid models in particular for tail dependence further in chapter 5 and [414].

Another, modelling aspect that has been highlighted in this review is that of higher dimensions and extreme joint distributions, as ones with tail dependence. In the modelling literature we can see a fair interest into modelling more complex dependence relationships together with a higher interest in models that capture dependence in the extreme parts, i.e. joint tails, e.g. through copulas. Given that a lack of relevant historical data is already now identified as a modelling challenge as shown by the few rather ad hoc methods mentioned in the review, we identified the modelling aspect of capturing tail dependence as one of increased current and future interest. In order to address this, we also proposed the SRP method in chapter 6 and [413].

While copulas and BNs are mentioned as the main models of interest for future research, we should however still consider other research in this area, such as that of assessing parametric joint distributions and regression models as in certain application areas, such as ecological and environmental risk assessment, they are of high interest and an ongoing research topic.

Research Question 2: *Which* key *elements of processes for eliciting dependence from experts have partly or completely been neglected in past and current research?*

Both literature reviews (chapters 3 and 4) present some of the main choices that need to be made when eliciting dependence from experts. The first focuses on the choices that result from the model choice, in particular whether we can elicit the natural input parameter of a model or if it needs to be transformed. While some findings exist on the ease with which experts can assess a particular association measure, more research in that is necessary.

Chapter 4 on the other hand addresses the whole elicitation process. The most crucial research gaps with regards to the elements of such processes (in the sense that they have been the most neglected but are necessary for any elicitation) are the choice of the elicited form, training experts, structuring experts' knowledge and beliefs and aggregating expert judgements. We re-address these elements below when discussing the future research of dependence elicitation.

Research Question 3: *What is the status-quo of related research topics of interest for cross-fertilisation from the elicitation of experts' (univariate) probabilities (e.g. behavioural research on cognitive fallacies) and to what extent have these taken into account in current dependence elicitation methods and processes?*
In chapter 4, we have presented some findings on a main cross-fertilisation topic, that of behavioural research on cognitive fallacies. While some cognitive fallacies for dependence assessment (in particular for conditional dependence, e.g. through conditional probabilities) have been identified and studied, they might almost seen as a by-product of the much more active psychological research on probability assessment in general (usually for univariate quantities). As such, it is important to note that in future more research in this area is desirable, in particular to guide the chosen elicited form and to provide experts with training. For training experts, we mentioned some potential studies for cross-fertilisation in chapter 4 as we might obtain some insight from the problems that student face when learning the concepts of various dependence parameters.
Another cross-fertilisation topic which has been identified as lacking, but is not addressed in depth in this thesis, is the question of what makes a good dependence assessor. In the introduction we have briefly discussed who an expert is, however in this regard this question needs to be extended to whether the same properties are true for an expert on dependence. Research on this could then further inform the area of choosing experts and sensible aggregation methods - for instance, when wanting to use calibration questions.

Research Question 4: *How can a method support experts in structuring their knowledge and beliefs about dependence relationships in order to mitigate common cognitive fallacies and enhance experts' confidence in their dependence assessments?*
In order to address this research question, we have proposed a conditional scenario mapping method in chapter 5 and applied it in a case-study on risk assessment in the UK Higher Education sector. While developing this method we have taken the presented desiderata of our method as guidance whereas these served to identify the shortcomings of other methods. Future methods for this purpose might base their work on these desiderata or even extend this list.
Through the feedback from the experts, who take part in the case-study, we can identify how well we achieve supporting experts in mitigating cognitive fallacies. For instance, anecdotal evidence in the study presented in chapter 5 indicates that our method can avoid the confusion of joint and conditional probabilities whereas at the same time we enhance experts' confidence in later quantitative assessments. Especially through decomposing (conditional) dependence relationships through backwards and forwards logic as well as structuring experts' reasoning via suggested event types and an appropriate graphical representation we aim at mitigating cognitive fallacies and make assessments more intuitive.

Research Question 5: *How can we support experts in making feasible dependence assessments while not restricting the level of detail and flexibility of a dependence model as desired by a decision-maker?*

In chapter 6 we propose the SRP method which guides experts in making feasible assessments on detailed aspects of a joint distribution. This might be important when a detailed specification is desired by a decision-maker, such as when tail dependencies are of interest. The method provides an algorithm, which gives upper and lower bounds for assessments, to ensure feasibility. We identified this as a sensible approach as it ensures that experts will stay within feasible ranges and it should be preferred to altering experts' assessments after they have been given. Of course, it is possible that experts are not willing to stay within those ranges which highlights the importance of other elements in the expert judgement process, such as structuring knowledge prior to a quantitative assessment.

Research Question 6: *How can we ease experts' cognitive burden when making detailed dependence assessment, in particular when assessing tail dependence?*
As mentioned in Research Question 1, we identified an increased interest in modelling tail dependencies in the overall research literature on dependence models. As such, models become more flexible and sophisticated to capture extreme joint distributions which are often more useful for a decision-maker than ones with low-dimensional assumptions. As with other dependence models, and in fact uncertainty models generally, a modelling challenge is that of lack of relevant historical data. In fact, the tails of joint distributions lack data by definition through having low probability.
Therefore, we have had a particular focus on how our methods of chapter 5, 6 and 7 address tail dependencies in addition to how compatible they are with models of interest, such as copulas. Especially chapter 7 shows an elicitation method which can be applied if we simply want to identify parametric copulas for checking if tail dependence is of relevance and if yes, whether it is symmetric or asymmetric.

8.3 Limitations of the presented research

The main limitation with the research presented in this thesis is that for the original research only a very limited number of case-studies exists in which their robustness has been tested. It is therefore key to continue the evaluation of these methods through further case-studies, including different experts and problem areas. Some possible extension and alteration, which might be of interest, have been mentioned when concluding the chapters in the original research part.
A further limitation of the proposed original methods is the lack of supporting material. While protocols and also to some extent online/software tools exist, these can be improved for future studies, also in order to maintain a high quality in applying these methods.
Lastly, for the literature reviews the common limitations apply, such as only peer-reviewed studies and reports that are available in the selected online libraries have been considered whereas some (also) lack having been applied in a variety of settings. As such, we have not included potential non-published approaches that might be found in industry applications.

8.4 Future research

Nowadays, the term “complexity” is a buzzword and as such it is not surprising that the interest in dependence modelling applications for complex systems is increasing. In this regard, BNs have already seen a sharp increase in use (approximately) over the last three decades and this interest is likely to continue. However, also other models, such as copulas, are seeing more frequent use more recently. Therefore, research on dependence elicitation becomes more pressing as a lack of historical data will always be of concern - similar to the univariate case. In fact, the case of having sufficient relevant historical data to quantify a model satisfactorily might be even less common as such data on dependence relationships are often unavailable even if marginal distributions are given. In this thesis we have identified some of the main areas to be addressed in dependence elicitation and we have proposed novel methods for some main elicitation questions. In future more research efforts in this direction are necessary. Building onto the answer to our first research question, the main future research might be focused on the following.

With regards to an appropriate and robust choice of an elicited form, in particular the literature review of chapter 3 and [412] conclude that it is on the one hand crucial to ensure intuitiveness or at least experts' familiarity of the elicited form while on the other hand eliciting direct model inputs whenever sensible. The direct elicitation of the natural model input is important for avoiding the introduction of additional assumptions when transferring the elicited form into the model input. Further, it is important to limit the possibility of making infeasible judgements. This is why in many studies in which experts are used to quantify dependence a focus is laid on direct or "natural" inputs. However, it is not stated nor discussed whether experts are familiar with this choice nor whether they considered the elicitation part as intuitive. For enhancing future dependence elicitation methods, more feedback on which forms experts are familiar with is crucial. This can come case-studies of dependence elicitations (even if this is not the main focus of the paper) together with more insight from research, such as [73]. Though, as we already aforementioned regarding the latter, case-studies of actual decision and risk analysis context provide more accurate feedback than stylised elicitation contexts and we need to be more critical with such stylised contexts (in particular if we elicit assessments from students rather than actual experts).

In particular, when making a sequence of assessments on a multivariate distribution, such as done in chapter 6 and [413], it is "easy" to make infeasible judgements and additional guidance on possible upper and lower feasibility bounds needs to be given. This becomes however more complicated when eliciting within different forms. With the advent of novel modelling approaches and tail dependence being en vogue, new approaches to eliciting dependence for copulas might emerge and for these the choice of the elicited form and the question on how to approach sequential assessments become even more pressing than in the multitude of studies in which simply conditional probabilities as direct inputs for BNs are elicited - even though it remains questionable in how many of these studies the experts had a good idea about the concept of conditional probabilities.

In fact, for BNs of higher dimensions, i.e. whenever we include growing conditioning sets, we need to question experts' understanding of the elicited form and similar questions to those above arise. For instance, we can use algorithms which provide experts with feasibility bounds, such as proposed in our elicitation method, for which we explicitly focused on eliciting conditional probabilities but avoid problems of such growing conditioning sets.

Lastly in the context of choosing the elicited form, more research in the role of feedback is important as some studies seem to suggest that directly eliciting correlation coefficients can be a preferred method when the experts obtain frequent feedback in this form. In situations of no feedback, such a choice can however be too complicated to conceptualise for an expert.

Another, area of necessary future research is training experts. In chapter 4 and [415], we outlined some review findings on training experts before eliciting dependence from them. Yet, this was only possible by including common findings from the literature on how students learn the relevant concepts in their mathematical education. While we believe that possibilities for cross-fertilisation between this area and training experts exists, it is clear that the work in this part for elicitation processes, i.e. how to train experts for dependence assessments, is lacking and need to be explored for improving future dependence elicitation protocols.

In addition to suitable training for experts, structuring their knowledge on dependence relationships is another main area for future research. As a main motivation for our conditional scenario mapping approach [414] served the findings of chapter 4 and [415], i.e. that until then no approach to structure experts judgements had existed. Of course, several dependence models, such as BNs, entail a prior, qualitative structuring of dependence relationships, however no methods existed to extend such efforts to other dependence models and in particular tail dependencies. Additionally to ease experts' cognitive burden, in particular for justifying the experts' assessments and transparency reasons, the detailed recording of their rationales is also crucial and should be further explored.

Finally, we have mentioned mathematical aggregation of dependence assessments in chapters 3 and 4 and the case-study of chapters 5 and 7. We used and compared some available methods but aggregating assessment has not been the focus of these studies. And while of course, behavioural aggregation methods and equal-weighting methods are also applicable for dependence assessments, it is clear that this area needs more research in future.

The above discussion on some main possible future research areas for eliciting dependence from experts shows that still much work is needed, nevertheless the advances along elicitation processes presented in this thesis show that this can be fruitful and rewarding.

Bibliography

[1] A E Abbas. Entropy methods for joint distributions in decision analysis. *IEEE Transactions on Engineering Management*, 53(1):146–159, 2006. 116

[2] A E Abbas, D V Budescu, and Y Gu. Assessing joint distributions with isoprobability contours. *Management Science*, 56(6):997–1011, 2010. 58

[3] K Abou-Moustafa, F De La Torre, and F Ferrie. Designing a metric for the difference between gaussian densities. *Brain, Body and Machine*, pages 57–70, 2010. 91

[4] F Ackermann and C Eden. Using causal mapping: individual and group; traditional and new. In *Systems Modelling: Theory and Practice*, pages 127–145. Wiley, UK, 2004. 103, 104

[5] F Ackermann, C Eden, T Williams, and S Howick. Systemic risk assessment: a case study. *Journal of the Operational Research Society*, 58(1):39–51, 2007. 104

[6] F Ackermann, S Howick, J Quigley, L Walls, and T Houghton. Systemic risk elicitation: Using causal maps to engage stakeholders and build a comprehensive view of risks. *European Journal of Operational Research*, 238(1):290–299, 2014. 104

[7] J Aczél and C Wagner. A characterization of weighted arithmetic means. *SIAM Journal on Algebraic Discrete Methods*, 1(3):259–260, 1980. 90

[8] I Ajzen. Intuitive theories of events and the effects of base-rate information on prediction. *Journal of Personality and Social Psychology*, 35(5):303, 1977. 71, 75, 144

[9] S A Al-Awadhi and P H Garthwaite. An elicitation method for multivariate normal distributions. *Communications in Statistics-Theory and Methods*, 27(5):1123–1142, 1998. 44

[10] S A Al-Awadhi and P H Garthwaite. Prior distribution assessment for a multivariate normal distribution: an experimental study. *Journal of Applied Statistics*, 28(1):5–23, 2001. 44

[11] S A Al-Awadhi and P H Garthwaite. Quantifying expert opinion for modelling fauna habitat distributions. *Computational Statistics*, 21(1):121–140, 2006. 46, 48

[12] B J M Ale, L J Bellamy, R M Cooke, L H J Goossens, A R Hale, A L C Roelen, and E Smith. Towards a causal model for air transport safety-an ongoing research project. *Safety Science*, 44(8):657–673, 2006. 103, 105

[13] L G Allan. A note on measurement of contingency between two binary variables in judgment tasks. *Bulletin of the Psychonomic Society*, 15(3):147–149, 1980. 34, 71

[14] P G Altbach and J Knight. The internationalization of higher education: Motivations and realities. *Journal of Studies in International Education*, 11(3-4):290–305, 2007. 108

[15] M Amer, T U Daim, and A Jetter. A review of scenario planning. *Futures*, 46:23–40, 2013. 101

[16] J L Anderson. Horizon mission methodology - a tool for the study of technology innovation and new paradigms. Technical report, NASA Working Paper AIAA PAPER 93-1134, 1993. 103, 104

[17] P Arbenz and D Canestraro. Estimating copulas for insurance from scarce observations, expert opinion and prior information: a bayesian approach. *ASTIN Bulletin: The Journal of the IAA*, 42(1):271–290, 2012. 42, 44

[18] T Augustin and M E G V Cattaneo. Foundations of probability. In *International Encyclopedia of Statistical Science*, pages 542–544. Springer, 2011. 18, 20

[19] T Aven. On how to define, understand and describe risk. *Reliability Engineering & System Safety*, 95(6):623–631, 2010. 30

[20] T Aven. *Misconceptions of risk*. John Wiley & Sons, UK, 2011. 29

[21] T Aven. The risk concept-historical and recent development trends. *Reliability Engineering & System Safety*, 99:33–44, 2012. 29, 127

[22] T Aven and S Guikema. On the concept and definition of terrorism risk. *Risk Analysis*, 35(12):2162–2171, 2015. 127

[23] T Aven and V Kristensen. Perspectives on risk: review and discussion of the basis for establishing a unified and holistic approach. *Reliability Engineering & System Safety*, 90(1):1–14, 2005. 29

[24] T Aven and B S Krohn. A new perspective on how to understand, assess and manage risk and the unforeseen. *Reliability Engineering & System Safety*, 121:1–10, 2014. 127

[25] T Aven and O Renn. On risk defined as an event where the outcome is uncertain. *Journal of Risk Research*, 12(1):1–11, 2009. 30

[26] N Balakrishnan and V B Nevzorov. *A primer on statistical distributions*. John Wiley & Sons, UK, 2004. 44

[27] M Bar-Hillel. The base-rate fallacy in probability judgments. *Acta Psychologica*, 44(3):211–233, 1980. 74

[28] C Batanero and C Díaz. Training school teachers to teach probability: reflections and challenges. *Chilean Journal of Statistics*, 3(1):3–13, 2012. 86

[29] C Bechlivanidis and D A Lagnado. Does the why tell us the when? *Psychological Science*, 24(8):1563–1572, 2013. 75, 113

[30] T Bedford. *Interactive expert assignment of minimally-informative copulae.* University of Strathclyde, 2002. 43

[31] T Bedford and R M Cooke. Vines: A new graphical model for dependent random variables. *Annals of Statistics*, pages 1031–1068, 2002. 116

[32] T Bedford, A Daneshkhah, and K J Wilson. Approximate uncertainty modeling in risk analysis with vine copulas. *Risk Analysis*, 36:792–815, 2016. 43, 116, 126

[33] T Bedford, R Denning, M Revie, and L Walls. Applying bayes linear methods to support reliability procurement decisions. In *Reliability and Maintainability Symposium*, pages 341–346. IEEE, 2008. 47

[34] T Bedford and K J Wilson. On the construction of minimum information bivariate copula families. *Annals of the Institute of Statistical Mathematics*, 66(4):703–723, 2014. 126

[35] T Bedford, K J Wilson, and A Daneshkhah. Assessing parameter uncertainty on coupled models using minimum information methods. *Reliability Engineering & System Safety*, 125:3–12, 2014. 116, 117, 126, 136

[36] E J Bedrick, R Christensen, and W Johnson. A new perspective on priors for generalized linear models. *Journal of the American Statistical Association*, 91(436):1450–1460, 1996. 47

[37] Bell, D E and Raiffa, H and Tversky, A. *Decision making: Descriptive, normative, and prescriptive interactions.* Cambridge University Press, USA, 1988. 7, 8, 9, 159

[38] M Ben-Daya, D Ait-Kadi, S O Duffuaa, J Knezevic, and A Raouf. *Handbook of maintenance management and engineering*, volume 7. Springer, 2009. 103

[39] P L Bernstein. *Against the gods: The remarkable story of risk.* Wiley, USA, 1996. 17, 28

[40] B Bes, S Sloman, C G Lucas, and É Raufaste. Non-bayesian inference: Causal structure trumps correlation. *Cognitive Science*, 36(7):1178–1203, 2012. 75, 79

[41] R Beyth Marom. Perception of correlation re-examined. *Memory and Cognition*, 10(6):511–519, 1982. 34

[42] J E Bickel and J E Smith. Optimal sequential exploration: A binary learning model. *Decision Analysis*, 3(1):16–32, 2006. 116, 126

[43] V M Bier. Choosing what to protect. *Risk Analysis*, 27(3):607–620, 2007. 128

[44] P Bishop, A Hines, and T Collins. The current state of scenario development: an overview of techniques. *Foresight*, 9(1):5–25, 2007. 103, 104

[45] N Blomqvist. On a measure of dependence between two random variables. *The Annals of Mathematical Statistics*, pages 593–600, 1950. 51

[46] J S Blumenthal-Barby and H Krieger. Cognitive biases and heuristics in medical decision making: a critical review using a systematic search strategy. *Medical Decision Making*, 35(4):539–557, 2015. 144

[47] K Bocker, A Crimmi, and H Fink. Bayesian risk aggregation: Correlation uncertainty and expert judgement. In *Rethinking Risk Measurement and Reporting*. Risk Books, UK, 2010. 44

[48] L Bojke, K Claxton, Y Bravo-Vergel, M Sculpher, S Palmer, and K Abrams. Eliciting distributions to populate decision analytic models. *Value in Health*, 13(5):557–564, 2010. 140

[49] L Bojke, B Grigore, D Jankovic, J Peters, M Soares, and K Stein. Informing reimbursement decisions using cost-effectiveness modelling: A guide to the process of generating elicited priors to capture model uncertainties. *Pharmaco Economics*, pages 1–11, 2017. 140

[50] F Bolger. The selection of experts for (probabilistic) expert knowledge elicitation. In L C Dias, A Morton, and J Quigley, editors, *Elicitation: The science and art of structuring judgement*, volume 261, chapter 16, pages 393–443. Springer International Series in Operations Research and Management Science, USA, 2018. 3, 4

[51] F Bolger and G Wright. Assessing the quality of expert judgment: Issues and analysis. *Decision Support Systems*, 11(1):1–24, 1994. 58, 84

[52] E Borgonovo. *Sensitivity Analysis*. Springer, USA, 2017. 7

[53] M Borovcnik and R Kapadia. From puzzles and paradoxes to concepts in probability. In *Probabilistic Thinking*, pages 35–73. Springer, Switzerland, 2014. 72, 144

[54] R Bradfield, G Wright, G Burt, G Cairns, and K Van Der Heijden. The origins and evolution of scenario techniques in long range business planning. *Futures*, 37(8):795–812, 2005. 98, 101

[55] R Bradley, F Dietrich, and C List. Aggregating causal judgments. *Philosophy of Science*, 81(4):491–515, 2014. 88, 91

[56] Peter J Brockwell and Richard A Davis. *Introduction to time series and forecasting*. springer, Switzerland, 2016. 158

[57] D A Broniatowski, E Y Klein, and V F Reyna. Germs are germs, and why not take a risk? patients expectations for prescribing antibiotics in an inner-city emergency department. *Medical Decision Making*, 35(1):60–67, 2015. 140

[58] G G Brown and L A Cox Jr. How probabilistic risk assessment can mislead terrorism risk analysts. *Risk Analysis*, 31(2):196–204, 2011. 128

[59] G J Browne, S P Curley, and P G Benson. Evoking information in probability assessment: Knowledge maps and reasoning-based directed questions. *Management Science*, 43(1):1–14, 1997. 87

[60] C Bunea and T Bedford. The effect of model uncertainty on maintenance optimization. *IEEE Transactions on Reliability*, 51(4):486–493, 2002. 44

[61] D W Bunn and A A Salo. Forecasting with scenarios. *European Journal of Operational Research*, 68(3):291–303, 1993. 101

[62] S Campbell. Determining overall risk. *Journal of Risk Research*, 8(7-8):569–581, 2005. 29

[63] P Carranza and A Kuzniak. Duality of probability and statistics teaching in french education. *Joint ICMI/IASE study: teaching statistics in school mathematics*, 2008. 72, 144

[64] K Chaloner, T Church, T A Louis, and J P Matts. Graphical elicitation of a prior distribution for a clinical trial. *The Statistician*, pages 341–353, 1993. 44

[65] K Chaloner and G T Duncan. Some properties of the dirichlet-multinomial distribution and its use in prior elicitation. *Communications in Statistics-Theory and Methods*, 16(2):511–523, 1987. 44, 45

[66] L J Chapman and J P Chapman. Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology*, 74(3):271–280, 1969. 35, 71, 79

[67] P Checkland. *Systems thinking, systems practice*. Wiley, UK, 1990. 98

[68] Peter Checkland and John Poulter. Soft systems methodology. In M Reynolds and S Holwell, editors, *Systems approaches to managing change: A practical guide*, pages 191–242. Springer, UK, 2010. 103

[69] K Chelst and S E Bodily. Structured risk management: filling a gap in decision analysis education. *Journal of the Operational Research Society*, pages 1420–1432, 2000. 101

[70] S Low Choy, R O'Leary, and K Mengersen. Elicitation by design in ecology: using expert opinion to inform priors for bayesian statistical models. *Ecology*, 90(1):265–277, 2009. 47, 48

[71] C W Churchman. Operations research as a profession. *Management Science*, 17(2):3–37, 1970. 104

[72] A Clauset, M Young, and K S Gleditsch. On the frequency of severe terrorist events. *Journal of Conflict Resolution*, 51(1):58–87, 2007. 127

[73] R T Clemen, G W Fischer, and R L Winkler. Assessing dependence: Some experimental results. *Management Science*, 46(8):1100–1115, 2000. 50, 53, 54, 56, 57, 58, 82, 84, 146, 164

[74] R T Clemen and T Reilly. Correlations and copulas for decision and risk analysis. *Management Science*, 45(2):208–224, 1999. 5, 41, 49, 53, 54, 56, 57, 84

[75] R T Clemen and T Reilly. *Making hard decisions with DecisionTools*. Cengage Learning, USA, 2013. 67, 80, 157

[76] A Colson, S Adhikari, A Sleemi, and R Laxminarayan. Quantifying uncertainty in intervention effectiveness with structured expert judgement: an application to obstetric fistula. *BMJ Open*, 5(6):7–23, 2015. 140

[77] United States Nuclear Regulatory Commission. Reactor safety study: An assessment of accident risks in us commercial nuclear power plants-executive summary-nureg 75-014. Technical report, United States Nuclear Regulatory Commission, 1975. 67

[78] United States Nuclear Regulatory Commission. Reactor risk reference document-nureg-1150. Technical report, United States Nuclear Regulatory Commission, 1987. 67

[79] R M Cooke. *Experts in uncertainty: opinion and subjective probability in science*. Oxford University Press, USA, 1991. 4, 33, 36, 48, 81, 91, 109, 130, 142

[80] R M Cooke. Parameter fitting for uncertain models: modelling uncertainty in small models. *Reliability Engineering & System Safety*, 44(1):89–102, 1994. 38, 48, 116

[81] R M Cooke. The anatomy of the squizzel: the role of operational definitions in representing uncertainty. *Reliability Engineering & System Safety*, 85(1):313–319, 2004. 14, 24, 25

[82] R M Cooke. Uncertainty analysis comes to integrated assessment models for climate change and conversely. *Climatic Change*, 117(3):467–479, 2013. 34, 48, 67, 68

[83] R M Cooke. Validating expert judgment with the classical model. In *Experts and Consensus in Social Science*, pages 191–212. Springer, 2014. 36

[84] R M Cooke and T Bedford. *Probabilistic risk analysis: Foundations and methods*. Cambridge University Press, UK, 2001. 14, 15, 16, 17, 18, 20, 21, 25, 26, 28, 38, 61, 67, 74, 87, 97, 98, 101, 103, 104

[85] R M Cooke and L H J Goossens. Expert judgement elicitation for risk assessments of critical infrastructures. *Journal of Risk Research*, 7(6):643–656, 2004. 3, 4, 88, 99, 113

[86] R M Cooke and L H J Goossens. Tu delft expert judgment data base. *Reliability Engineering & System Safety*, 93(5):657–674, 2008. 91

[87] R M Cooke and L J H Goossens. Procedures guide for structured expert judgment. *Project Report to the European Commission*, 18820, 1999. 67, 99, 113

[88] R M Cooke and G Kelly. Climate change uncertainty quantification: Lessons learned from the joint eu-usnrc project on uncertainty analysis of probabilistic accident consequence codes. Technical report, RFF Discussion Paper, 2010. 51

[89] R M Cooke and B Kraan. Dealing with dependencies in uncertainty analysis. In *Probabilistic Safety Assessment and Management*, pages 625–630. Springer, 1996. 34, 51

[90] F J Costello. How probability theory explains the conjunction fallacy. *Journal of Behavioral Decision Making*, 22(3):213–234, 2009. 78

[91] A Cottrell. Keynes's theory of probability and its relevance to his economics: three theses. *Economics & Philosophy*, 9(1):25–51, 1993. 22

[92] J A G Cox and T Worthington. The antibiotic apocalypse-scaremongering or scientific reporting? *Trends in Microbiology*, 25(3):167–169, 2017. 140, 141

[93] L A Cox Jr. *Quantitative health risk analysis methods: modeling the human health impacts of antibiotics used in food animals*, volume 82. Springer Science & Business Media, USA, 2006. 154

[94] L A Cox Jr. Some limitations of 'risk= threat× vulnerability× consequence' for risk analysis of terrorist attacks. *Risk Analysis*, 28(6):1749–1761, 2008. 127

[95] L A Cox Jr. Game theory and risk analysis. *Risk Analysis*, 29(8):1062–1068, 2009. 128

[96] L A Cox Jr. Improving risk-based decision making for terrorism applications. *Risk Analysis*, 29(3):336–341, 2009. 128

[97] L A Cox Jr. *Risk analysis of complex and uncertain systems*, volume 129. Springer Science & Business Media, USA, 2009. 154

[98] L A Cox Jr and R S Singer. Confusion over antibiotic resistance: ecological correlation is not evidence of causation. *Foodborne Pathogens and Disease*, 9(8):776–776, 2012. 141

[99] J Crook and F Moreira. Checking for asymmetric default dependence in a credit card portfolio: A copula approach. *Journal of Empirical Finance*, 18(4):728–742, 2011. 98

[100] M Crotty. *The foundations of social research: Meaning and perspective in the research process.* Sage, USA, 1998. 6

[101] C L Cuite, N D Weinstein, K Emmons, and G Colditz. A test of numeric formats for communicating risk probabilities. *Medical Decision Making*, 28(3):377–384, 2008. 140

[102] A A R Cunha and D C Morais. Analysing the use of cognitive maps in an experiment on a group decision process. *Journal of the Operational Research Society*, 67(12):1459–1468, 2016. 104

[103] A Curry and W Schultz. Roads less travelled: different methods, different futures, 2009. 103

[104] A Daneshkhah and J Oakley. Eliciting multivariate probability distributions. In K Bocker, editor, *Rethinking Risk Measurement and Reporting.* Risk Books, UK, 2010. 34, 53

[105] A Darwiche. *Modeling and reasoning with Bayesian networks.* Cambridge University Press, USA, 2009. 39

[106] R M Dawes. *Rational choice in an uncertain world: the psychology of judgement and decision making.* Sage Publications, USA, 1988. 71, 74, 113

[107] B De Finetti. *Theory of probability: a critical introductory treatment*, volume 6. John Wiley & Sons, UK, 2017. 24, 26

[108] J De Mey. The aftermath of september 11: The impact on and systemic risk to the insurance industry. *The Geneva Papers on Risk and Insurance Issues and Practice*, 28(1):65–70, 2003. 127

[109] K De Witte and L López-Torres. Efficiency in education: a review of literature and a way forward. *Journal of the Operational Research Society*, 68(4):339–363, 2017. 108

[110] M H DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121, 1974. 90

[111] M H DeGroot and S E Fienberg. The comparison and evaluation of forecasters. *The Statistician*, pages 12–22, 1983. 36

[112] D J Delgado-Hernández, O Morales-Nápoles, D De-León-Escobedo, and J C Arteaga-Arcos. A continuous bayesian network for earth dams risk assessment: An application. *Structure and Infrastructure Engineering*, 10(2):225–238, 2014. 56

[113] D Denyer, D Tranfield, and J E Van Aken. Developing design propositions through research synthesis. *Organization Studies*, 29(3):393–413, 2008. 8

[114] C Díaz, C Batanero, and J M Contreras. Teaching independence and conditional probability. *Boletín de Estadística e Investigación Operativa*, 26(2):149–162, 2010. 72, 86, 144

[115] J M Dickey, D V Lindley, and S J Press. Bayesian estimation of the dispersion matrix of a multivariate normal distribution. *Communications in Statistics-Theory and Methods*, 14(5):1019–1034, 1985. 44

[116] F Dietrich and C List. Probabilistic opinion pooling. In A Hajek and C Hitchcock, editors, *The Oxford Handbook of Probability and Philosophy*, pages 179–207. Oxford University Press, UK, 2017. 90

[117] F J Diez. Parameter adjustment in bayes networks. the generalized noisy or-gate. In *Proceedings of the Ninth international conference on uncertainty in artificial intelligence*, pages 99–105. Morgan Kaufmann Publishers, 1993. 40, 41

[118] Dresch, A and Pacheco Lacerda, D and Valle Antunes, J A Jr. *Design Science Research: A method for science and technology advancement*. Springer, Switzerland, 2015. 9

[119] M J Druzdel and L C Van Der Gaag. Building probabilistic networks: Where do the numbers come from? *IEEE Transactions on Knowledge and Data Engineering*, 12(4):481–486, 2000. 39

[120] M J Druzdzel and L C Van Der Gaag. Elicitation of probabilities for belief networks: Combining qualitative and quantitative information. In *Proceedings of the Eleventh conference on uncertainty in artificial intelligence*, pages 141–148. Morgan Kaufmann Publishers, 1995. 40

[121] D Dubois, H Prade, and R Sabbadin. Decision-theoretic foundations of qualitative possibility theory. *European Journal of Operational Research*, 128(3):459–478, 2001. 13

[122] W M DuCharme. Response bias explanation of conservative human inference. *Journal of Experimental Psychology*, 85(1):66–74, 1970. 71, 77

[123] G Ducot and G J Lubben. A typology for scenarios. *Futures*, 12(1):51–57, 1980. 104

[124] J Dunjó, V Fthenakis, J A Vílchez, and J Arnaldos. Hazard and operability (hazop) analysis. a literature review. *Journal of Hazardous Materials*, 173(1):19–32, 2010. 103

[125] F Durante and C Sempi. *Principles of copula theory.* CRC Press, USA, 2015. 41, 119, 143

[126] D Duxbury and B Summers. Financial risk perception: Are individuals variance averse or loss averse. *Economics Letters*, 84(1):21–28, 2004. 28

[127] D M Eddy. Probabilistic reasoning in clinical medicine: Problems and opportunities. In D Kahneman, P Slovic, and A Tversky, editors, *Judgment under uncertainty: Heuristics and biases*, pages 249–267. Cambridge University Press, USA, 1982. 71, 74, 144

[128] D M Eddy. Variations in physician practice: the role of uncertainty. *Health Affairs*, 3(2):74–89, 1984. 140

[129] D M Eddy. Evidence-based medicine: a unified approach. *Health Affairs*, 24(1):9–17, 2005. 139

[130] C Eden and F Ackermann. The role of gdss in scenario development and strategy making. In *String Processing and Information Retrieval Symposium and International Workshop on Groupware*, pages 234–242. IEEE, 1999. 101

[131] A B Eder, K Fiedler, and S Hamm-Eder. Illusory correlations revisited: The role of pseudocontingencies and working-memory capacity. *The Quarterly Journal of Experimental Psychology*, 64(3):517–532, 2011. 71, 79

[132] W Edwards. Optimal strategies for seeking information: Models for statistics, choice reaction times, and human information processing. *Journal of Mathematical Psychology*, 2(2):312–329, 1965. 71, 77

[133] A Eichler and M Vogel. Three approaches for modelling situations with randomness. In E J Chernoff and B Sriraman, editors, *Probabilistic thinking*, pages 75–99. Springer, Switzerland, 2014. 86

[134] H J Einhorn and R M Hogarth. Judging probable cause. *Psychological Bulletin*, 99(1):3–19, 1986. 71, 72, 73, 75, 76, 77, 107, 113, 144

[135] F G Elfadaly and P H Garthwaite. Eliciting dirichlet and connor-mosimann prior distributions for multinomial models. *Test*, 22(4):628–646, 2013. 44

[136] M Eling and D Toplek. Modeling and management of nonlinear dependencies–copulas in dynamic financial analysis. *Journal of Risk and Insurance*, 76(3):651–681, 2009. 126

[137] P Embrechts, A McNeil, and D Straumann. Correlation and dependence in risk management: properties and pitfalls. *Risk Management: Value at Risk and Beyond*, pages 176–223, 2002. 126, 158

[138] R F Eyler, S Cordes, B R Szymanski, and L Fraenkel. Utilization of continuous 'spinners' to communicate risk. *Medical Decision Making*, pages 725–729, 2017. 140

[139] B C Ezell, S P Bennett, D Von Winterfeldt, J Sokolowski, and A J Collins. Probabilistic risk analysis and terrorism risk. *Risk Analysis*, 30(4):575–589, 2010. 128, 133

[140] P L Fackler. Modeling interdependence: an approach to simulation and elicitation. *American Journal of Agricultural Economics*, 73(4):1091–1097, 1991. 51

[141] R Falk. Conditional probabilities: insights and difficulties. In D Tall, editor, *Proceedings of the Second International Conference on Teaching Statistics*, pages 292–297, 1983. 75, 76

[142] M Farrow. Practical building of subjective covariance structures for large complicated systems. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 52(4):553–573, 2003. 46

[143] M Farrow, M Goldstein, and T Spiropoulos. Developing a bayes linear decision support system for a brewery. In S French and J Q Smith, editors, *The Practice of Bayesian Analysis*, pages 71–106. UK, 1997. 47

[144] N Fenton and M Neil. *Risk assessment and decision analysis with Bayesian networks*. CRC Press, USA, 2012. 88

[145] N E Fenton, M Neil, and J G Caballero. Using ranked nodes to model qualitative judgments in bayesian networks. *IEEE Transactions on Knowledge and Data Engineering*, 19(10):1420–1432, 2007. 41, 84

[146] K Fiedler, B Brinkmann, T Betsch, and B Wild. A sampling approach to biases in conditional probability judgments: beyond base rate neglect and statistical format. *Journal of Experimental Psychology: General*, 129(3):399, 2000. 74

[147] P C Fishburn. *Utility for decision making*. Wiley, USA, 1970. 25, 26, 27

[148] S Fleetwood. Ontology in organization and management studies: A critical realist perspective. *Organization*, 12(2):197–222, 2005. 6

[149] M J Flores, A E Nicholson, A Brunskill, K B Korb, and S Mascaro. Incorporating expert knowledge when learning bayesian network structure: a medical case study. *Artificial Intelligence in Medicine*, 53(3):181–204, 2011. 88

[150] EFSA=European Food and Safety Authority. Guidance on expert knowledge elicitation in food and feed safety risk assessment. Technical report, EFSA Journal 12(6), 2014. 33, 67

[151] J Fountain and P Gunby. Ambiguity, the certainty illusion, and the natural frequency approach to reasoning with inverse probabilities. *New Zealand Economic Papers*, 45(1-2):195–207, 2011. 71

[152] L A Franco. Forms of conversation and problem structuring methods: a conceptual development. *Journal of the Operational Research Society*, 57(7):813–821, 2006. 98

[153] S French. *Decision theory: an introduction to the mathematics of rationality*. Halsted Press, UK, 1986. 21

[154] S French. Uncertainty and imprecision: Modelling and analysis. *Journal of the Operational Research Society*, 46(1):70–79, 1995. 105

[155] S French. Aggregating expert judgement. *Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales. Serie A Matematicas*, 105(1):181–206, 2011. 3, 33, 39, 68, 90, 147

[156] S French. Cynefin: uncertainty, small worlds and scenarios. *Journal of the Operational Research Society*, 66(10):1635–1645, 2015. 98

[157] S French and D Rios Insua. *Statistical decision theory*. Wiley, UK, 2000. 101

[158] I Gal. Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review*, 70(1):1–25, 2002. 86

[159] M C Galavotti. The modern epistemic interpretations of probability: Logicism and subjectivism. In *Handbook of the History of Logic*, volume 10, pages 153–203. Elsevier, 2011. 20, 21, 22, 23

[160] B J Garrick, J E Hall, M Kilger, J C McDonald, T O'Toole, P S Probst, E R Parker, R Rosenthal, A W Trivelpiece, and L A Van Arsdale. Confronting the risks of terrorism: making the right decisions. *Reliability Engineering & System Safety*, 86(2):129–176, 2004. 128

[161] J B Garrick. Perspectives on the use of risk assessment to address terrorism. *Risk Analysis*, 22(3):421–423, 2002. 128

[162] P H Garthwaite and S A Al-Awadhi. Non-conjugate prior distribution assessment for multivariate normal sampling. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(1):95–110, 2001. 44

[163] P H Garthwaite, S A Al-Awadhi, F G Elfadaly, and D J Jenkinson. Prior distribution elicitation for generalized linear and piecewise-linear models. *Journal of Applied Statistics*, 40(1):59–75, 2013. 47

[164] P H Garthwaite and J M Dickey. An elicitation method for multiple linear regression models. *Journal of Behavioral Decision Making*, 4(1):17–31, 1991. 47

[165] P H Garthwaite, J B Kadane, and A O'Hagan. Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association*, 100(470):680–701, 2005. 4, 34, 53, 54, 84

[166] S I Gass. Decision-aiding models: validation, assessment, and related issues for policy analysis. *Operations Research*, 31(4):603–631, 1983. 7, 114

[167] Saul I Gass and Lambert S Joel. Concepts of model confidence. *Computers & Operations Research*, 8(4):341–346, 1981. 114

[168] I Gavanski and C Hui. Natural sample spaces and uncertain belief. *Journal of Personality and Social Psychology*, 63(5):766–780, 1992. 75

[169] C Genest and A C Favre. Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, 12(4):347–368, 2007. 44

[170] C Genest, M Gendron, and M Bourdeau-Brien. The advent of copulas in finance. *The European Journal of Finance*, 15(7-8):609–618, 2009. 44

[171] T Gilovich, D Griffin, and D Kahneman. *Heuristics and biases: The psychology of intuitive judgment.* Cambridge University Press, USA, 2002. 70

[172] D V Gokhale and S J Press. Assessment of a prior distribution for the correlation coefficient in a bivariate normal distribution. *Journal of the Royal Statistical Society. Series A (General)*, pages 237–249, 1982. 34, 54, 56, 84

[173] M Goldstein and D Wooff. *Bayes linear statistics, theory and methods*, volume 716. John Wiley & Sons, UK, 2007. 46

[174] J P Gosling, A Hart, H Owen, M Davies, J Li, and C MacKay. A bayes linear approach to weight-of-evidence risk assessment for skin allergy. *Bayesian Analysis*, 8(1):169–186, 2013. 47

[175] John Paul Gosling. Shelf: the sheffield elicitation framework. In L C Dias, A Morton, and J Quigley, editors, *Elicitation: The science and art of structuring judgement*, volume 261, chapter 4, pages 61–93. Springer International Series in Operations Research and Management Science, New York, 2018. 129

[176] M Granger and M Henrion. *Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis.* Cambridge University Press, USA, 1990. 67, 84

[177] G Gredebäck, A Winman, and P Juslin. Rational assessments of covariation and causality. In *Proceedings of the 22nd annual conference of the Cognitive Science Society*, pages 190–195. Erlbaum Hillsdale, 2000. 34

[178] B Grigore, J Peters, C Hyde, and K Stein. A comparison of two methods for expert elicitation in health technology assessments. *BMC Medical Research Methodology*, 16(1):85, 2016. 140

[179] P Grossi and H Kunreuther. *Catastrophe modeling: a new approach to managing risk.* Springer Science & Business Media, New York, 2005. 127, 136

[180] I Hacking. *The Emergence of Probability: A Philosophical Study of Early Ideas about Probability, Induction and Statistical Inference: A Philosophical Study of Early.Induction and Statistical Inference.* Cambridge University Press, USA, 1975. 15

[181] T Hailperin. Probability logic. *Notre Dame Journal of Formal Logic*, 25(3):198–212, 1984. 18

[182] Y Y Haimes. On the complex definition of risk: A systems-based approach. *Risk Analysis*, 29(12):1647–1654, 2009. 101, 107, 128

[183] A Hájek. What conditional probability could not be. *Synthese*, 137(3):273–323, 2011. 18, 25

[184] A Hájek. Interpretations of probability. In *The Stanford Encyclopedia of Philosophy*. Citeseer, 2012. 21

[185] D L Hamilton. *Cognitive processes in stereotyping and intergroup behavior*. Psychology Press, USA, 2015. 79

[186] P K J Han, W M P Klein, and N K Arora. Varieties of uncertainty in health care: a conceptual taxonomy. *Medical Decision Making*, 31(6):828–838, 2011. 139, 140

[187] A M Hanea, M F McBride, M A Burgman, B C Wintle, F Fidler, L Flander, C R Twardy, B Manning, and S Mascaro. I nvestigate d iscuss e stimate a ggregate for structured expert judgement. *International Journal of Forecasting*, 33(1):267–279, 2017. 88, 108

[188] A M Hanea, O Morales-Napoles, and D Ababei. Non-parametric bayesian networks: Improving theory and reviewing applications. *Reliability Engineering & System Safety*, 144:265–284, 2015. 40, 82

[189] Anca Maria Hanea, Mark Burgman, and Victoria Hemming. Idea for uncertainty quantification. In L C Dias, A Morton, and J Quigley, editors, *Elicitation: The science and art of structuring judgement*, volume 261, chapter 5, pages 95–117. Springer International Series in Operations Research and Management Science, New York, 2018. 129

[190] D Hanea and B Ale. Risk of human fatality in building fires: A decision tool using bayesian networks. *Fire Safety Journal*, 44(5):704–710, 2009. 88

[191] M Hänninen, O A Banda Valdez, and P Kujala. Bayesian network model of maritime safety management. *Expert Systems with Applications*, 41(17):7837–7846, 2014. 41

[192] R Hastie. Causal thinking in judgments. pages 590–628. Wiley, USA, 2016. 34, 72, 73, 76, 114

[193] R Hastie and R M Dawes. *Rational choice in an uncertain world: the psychology of judgement and decision making.* Sage Publications, USA, 2001. 71, 74, 113, 144

[194] HEFCE. Risk management: A guide to good practice for higher education institutions. Technical report, Higher Education Funding Council for England, 2001. 108

[195] HEFCE. Risk management in higher education: A guide to good practice. Technical report, Higher Education Funding Council for England, 2005. 108

[196] J C Helton. Treatment of uncertainty in performance assessments for complex systems. *Risk Analysis*, 14(4):483–511, 1994. 15

[197] M Henrion. Some practical issues in constructing belief networks. In *Uncertainty in Artificial Intelligence*, pages 132–139. 1989. 39, 40, 88

[198] U Hoffrage, S Lindsey, R Hertwig, and G Gigerenzer. Communicating statistical information. *Science*, 290(5500):2261–2262, 2000. 55

[199] S C Hora. Eliciting probabilities from experts. In *Advances in Decision Analysis - From Foundations to Applications*, pages 129–163. 2007. 85, 87

[200] S C Hora and E Kardeş. Calibration, sharpness and the weighting of experts in a linear opinion pool. *Annals of Operations Research*, 229(1):429–450, 2015. 90

[201] R A Howard. Knowledge maps. *Management Science*, 35(8):903–922, 1989. 87, 97, 103

[202] R A Howard and J E Matheson. Influence diagrams. *Decision Analysis*, 2(3):127–143, 2005. 87, 97, 103

[203] Hume, D (edited by Beauchamp, T L). *An Enquiry Concerning Human Understanding*. Oxford University Press, USA, 1748/2000. 72, 73

[204] A James, S L Choy, and K Mengersen. Elicitator: An expert elicitation tool for regression in ecology. *Environmental Modelling & Software*, 25(1):129–145, 2010. 47

[205] E T Jaynes. Prior probabilities. *IEEE Transactions on systems science and cybernetics*, 4(3):227–241, 1968. 20

[206] E T Jaynes. *Probability theory: The logic of science*. Cambridge University Press, USA, 2003. 116

[207] D Jenkinson. The elicitation of probabilities: A review of the statistical literature. Technical report, BEEP Working Paper, University of Sheffield, 2005. 33

[208] H Joe. *Dependence modeling with copulas*. CRC Press, USA, 2014. 41, 42, 98, 99, 119, 143

[209] A John-Baptiste, M M Schapira, C Cravens, J D Chambers, P J Neumann, J Siegel, and W Lawrence. The role of decision models in health care policy: A case study. *Medical Decision Making*, 36(5):666–679, 2016. 154

[210] J Johnes. Operational research in education. *European Journal of Operational Research*, 243(3):683–696, 2015. 108

[211] W E Johnson. Probability: The deductive and inductive problems. *Mind*, 41(164):409–423, 1932. 22

[212] V R R Jose, R F Nau, and R L Winkler. Scoring rules, generalized entropy, and utility maximization. *Operations Research*, 56(5):1146–1157, 2008. 36

[213] V R R Jose, R F Nau, and R L Winkler. Sensitivity to distance and baseline distributions in forecast evaluation. *Management Science*, 55(4):582–590, 2009. 36

[214] J Kadane and L J Wolfson. Experiences in elicitation. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(1):3–19, 1998. 35, 49, 55, 56, 81, 84

[215] J B Kadane, J M Dickey, R L Winkler, W S Smith, and S C Peters. Interactive elicitation of opinion for a normal linear model. *Journal of the American Statistical Association*, 75(372):845–854, 1980. 34, 44, 47, 48

[216] Kahn, H and Wiener, A J. *The Year 2000: A Framework for Speculation on the Next Thirty-three Years.* MacMillan, USA, 1967. 101

[217] D Kahneman and S Frederick. Representativeness revisited: Attribute substitution in intuitive judgment. *Heuristics and biases: The psychology of intuitive judgment*, 49:49–81, 2002. 74, 113

[218] D Kahneman and A Tversky. Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 3(3):430–454, 1972. 74

[219] D Kahneman and A Tversky. On the psychology of prediction. *Psychological Review*, 80(4):237–251, 1973. 71, 74, 76

[220] D Kahneman and A Tversky. The simulation heuristic. Technical report, Stanford University, 1981. 114

[221] S F Kao and E A Wasserman. Assessment of an information integration account of contingency judgment with examination of subjective cell importance and method of information presentation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(6):1363–1386, 1993. 71, 78

[222] S Kaplan. The words of risk analysis. *Risk Analysis*, 17(4):407–417, 1997. 101

[223] S Kaplan and B J Garrick. On the quantitative definition of risk. *Risk Analysis*, 1(1):11–27, 1981. 28, 29, 101, 128

[224] G Karam, J Chastre, M H Wilcox, and J L Vincent. Antibiotic strategies in the era of multidrug resistance. *Critical Care*, 20(1):136–145, 2016. 141

[225] D L Keefer and S E Bodily. Three-point approximations for continuous random variables. *Management Science*, 29(5):595–609, 1983. 46

[226] G L Keeney and D Von Winterfeldt. Identifying and structuring the objectives of terrorists. *Risk Analysis*, 30(12):1803–1816, 2010. 128

[227] R L Keeney and D von Winterfeldt. Eliciting probabilities from experts in complex technical problems. *IEEE Transactions on Engineering Management*, 38(3):191–201, 1991. 34, 67

[228] G Keren and K H Teigen. Yet another look at heuristics and biases approach. In *Blackwell Handbook of Judgment and Decision Making*, pages 89–109. USA, 2006. 76

[229] J M Keynes. *A treatise on probability.* MacMillan and Co, UK, 1921. 21

[230] C Kirchsteiger. Preface: International workshop on promotion of technical harmonisation on risk-based decision-making, 2002. 30

[231] J P C Kleijnen. Verification and validation of simulation models. *European Journal of Operational Research*, 82(1):145–162, 1995. 114

[232] D N Kleinmuntz, M G Fennema, and M E Peecher. Conditioned assessment of subjective probabilities: Identifying the benefits of decomposition. *Organizational Behavior and Human Decision Processes*, 66(1):1–15, 1996. 87

[233] V A Knight, J E Williams, and I Reynolds. Modelling patient choice in healthcare systems: development and application of a discrete event simulation with agent-based decision making. *Journal of Simulation*, 6(2):92–102, 2012. 159

[234] J J Koehler. The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences*, 19(1):1–17, 1996. 74

[235] A N Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung, Ergebnisse der Mathematik*. Springer, Zurich, 1933. 18, 101

[236] S Kotz and J R Van Dorp. Generalized diagonal band copulas with two-sided generating densities. *Decision Analysis*, 7(2):196–214, 2010. 43

[237] C Kousky and Rr M Cooke. The unholy trinity: fat tails, tail dependence, and micro-correlations. *Resources for the Future Discussion Paper*, 9(36):1–36, 2009. 126

[238] B Kraan. *Probabilistic inversion in uncertainty analysis: and related topics*. PhD thesis, Delft University of Technology, Netherlands, 2002. 49, 53, 81

[239] B Kraan and T Bedford. Probabilistic inversion of expert judgments in the quantification of model uncertainty. *Management Science*, 51(6):995–1006, 2005. 38, 48, 116

[240] B Kraan and R M Cooke. Post-processing techniques for the joint cec/usnrc uncertainty analysis of accident consequence codes. *Journal of Statistical Computation and Simulation*, 57(1-4):243–259, 1997. 48

[241] W H Kruskal. Ordinal measures of association. *Journal of the American Statistical Association*, 53(284):814–861, 1958. 34, 51, 53, 55, 58, 78, 81

[242] S Kullback and R A Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951. 125

[243] Z Kunda and R E Nisbett. The psychometrics of everyday life. *Cognitive Psychology*, 18(2):195–224, 1986. 34, 54

[244] D Kurowicka and R M Cooke. *Uncertainty analysis with high dimensional dependence modelling*. John Wiley & Sons, UK, 2006. 33, 34, 38, 48, 50

[245] M Kynn. The 'heuristics and biases' bias in expert elicitation. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171(1):239–264, 2008. 72, 77

[246] Lad, F. *Operational Subjective Statistical Methods: a mathematical, philosophical, and historical introduction*. Wiley-Interscience, USA, 1996. 5

[247] J Ladyman, J Lambert, and K Wiesner. What is a complex system? *European Journal for Philosophy of Science*, 3(1):33–67, 2013. 104

[248] D A Lagnado and Sloman S A. Inside and outside probability judgments. pages 157–176. Blackwell Publishing, USA, 2004. 14

[249] D A Lagnado and D R Shanks. Probability judgment in hierarchical learning: A conflict between predictiveness and coherence. *Cognition*, 83(1):81–112, 2002. 78

[250] D A Lagnado and S A Sloman. Inside and outside probability judgment. In *Blackwell Handbook of Judgment and Decision Making*, pages 157–176. 2006. 77

[251] D C Lane. What we talk about when we talk about systems thinking. *Journal of the Operational Research Society*, 67(3):527–528, 2016. 98

[252] David C Lane. Diagramming conventions in system dynamics. *Journal of the Operational Research Society*, 51(2):241–245, 2000. 108

[253] K B Laskey. Model uncertainty: Theory and practical implications. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 26(3):340–348, 1996. 16

[254] J Leal, S Wordsworth, R Legood, and E Blair. Eliciting expert opinion for economic models: an applied example. *Value in Health*, 10(3):195–203, 2007. 140

[255] N G Leveson. Applying systems thinking to analyze and learn from events. *Safety Science*, 49(1):55–64, 2011. 98

[256] S B Levy. The challenge of antibiotic resistance. *Scientific American*, 278(3):32–39, 1998. 141

[257] S B Levy and B Marshall. Antibacterial resistance worldwide: causes, challenges and responses. *Nature Medicine*, 10:122–129, 2004. 141

[258] D V Lindley. The philosophy of statistics. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 49(3):293–337, 2000. 24

[259] J D C Little. Models and managers: The concept of a decision calculus. *Management Science*, 16(8):466–485, 1970. 9

[260] T C Liu and Y C Lin. The application of simulation-assisted learning statistics (sals) for correcting misconceptions and improving understanding of correlation. *Journal of Computer Assisted Learning*, 26(2):143–158, 2010. 86

[261] D Lowe, L Martingale, and M Yearworth. Guiding interventions in a multi-organisational context: combining the viable system model and hierarchical process modelling for use as a problem structuring method. *Journal of the Operational Research Society*, 67(12):1481–1495, 2016. 104

[262] P M Lurie and M S Goldberg. An approximate method for sampling correlated random variables from partially-specified distributions. *Management Science*, 44(2):203–218, 1998. 49

[263] D MacKenzie and T Spears. The formula that killed wall street: The gaussian copula and modelling practices in investment banking. *Social Studies of Science*, 44(3):393–417, 2014. 3

[264] D R Mandel and D R Lehman. Integration of contingency information in judgments of cause, covariation, and probability. *Journal of Experimental Psychology: General*, 127(3):269–285, 1998. 79

[265] D J Mayston. Measuring and managing educational performance. *Journal of the Operational Research Society*, 54(7):679–691, 2003. 108

[266] K J McConway. Marginalization and linear opinion pools. *Journal of the American Statistical Association*, 76(374):410–414, 1981. 90

[267] S B McGrayne. *The theory that would not die: how Bayes' rule cracked the enigma code, hunted down Russian submarines, & emerged triumphant from two centuries of controversy.* Yale University Press, USA, 2011. 75, 77

[268] C R M McKenzie and L A Mikkelsen. A bayesian view of covariation assessment. *Cognitive Psychology*, 54(1):33–61, 2007. 79

[269] A J McNeil, R Frey, and P Embrechts. *Quantitative risk management: Concepts, techniques and tools.* Princeton University Press, USA, 2015. 3

[270] B Meder and G Gigerenzer. Statistical thinking: No one left behind. In *Probabilistic Thinking*, pages 127–148. 2014. 71

[271] D L Medin, J D Coley, G Storms, and B L Hayes. A relevance theory of induction. *Psychonomic Bulletin & Review*, 10(3):517–532, 2003. 75

[272] P E Meehl and A Rosen. Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin*, 52(3):194–216, 1955. 71, 74, 144

[273] A M H Meeuwissen and T Bedford. Minimally informative distributions with given rank correlation for use in uncertainty analysis. *Journal of Statistical Computation and Simulation*, 57(1-4):143–174, 1997. 43, 116, 126

[274] M W Merkhofer. Quantifying judgmental uncertainty: Methodology, experiences, and insights. *IEEE Transactions on Systems, Man, and Cybernetics*, 17(5):741–752, 1987. 67

[275] J Meyer, M Taieb, and I Flascher. Correlation estimates as perceptual judgments. *Journal of Experimental Psychology: Applied*, 3(1):3, 1997. 56

[276] E Michel-Kerjan and B Pedell. How does the corporate world cope with mega-terrorism? puzzling evidence from terrorism insurance markets. *Journal of Applied Corporate Finance*, 18(4):61–75, 2006. 127

[277] G Midgley. The sacred and profane in critical systems thinking. *Systems Practice*, 5(1):5–16, 1992. 105

[278] J Mingers and J Rosenhead. Problem structuring methods in action. *European Journal of Operational Research*, 152(3):530–554, 2004. 98

[279] C J Mitchell, J De Houwer, and P F Lovibond. The propositional nature of human associative learning. *Behavioral and Brain Sciences*, 32(2):183–198, 2009. 34, 52, 78

[280] L Mkrtchyan, L Podofillini, and V N Dang. Bayesian belief networks for human reliability analysis: A review of applications and gaps. *Reliability Engineering & System Safety*, 139:1–16, 2015. 41, 84

[281] F A Moala and A OHagan. Elicitation of multivariate prior distributions: A nonparametric bayesian approach. *Journal of Statistical Planning and Inference*, 140(7):1635–1655, 2010. 44, 45, 53

[282] N Möller, S O Hansson, and M Peterson. Safety is more than the antonym of risk. *Journal of Applied Philosophy*, 23(4):419–432, 2006. 29

[283] G Montibeller and V Belton. Causal maps and the evaluation of decision options-a review. *Journal of the Operational Research Society*, 57(7):779–791, 2006. 108

[284] G Montibeller and D Von Winterfeldt. Cognitive and motivational biases in decision and risk analysis. *Risk Analysis*, 35(7):1230–1251, 2015. 70, 71, 72, 80

[285] L V Montiel and J E Bickel. A simulation-based approach to decision making with partial information. *Decision Analysis*, 9(4):329–347, 2012. 116

[286] L V Montiel and J E Bickel. Approximating joint probability distributions given partial information. *Decision Analysis*, 10(1):26–41, 2013. 116

[287] O Morales-Nápoles. *Bayesian belief nets and vines in aviation safety and other applications*. PhD thesis, Delft University of Technology, Netherlands, 2010. 52, 56, 81, 84, 92

[288] O Morales-Nápoles, D J Delgado-Hernández, D De-León-Escobedo, and J C Arteaga-Arcos. A continuous bayesian network for earth dams' risk assessment: methodology and quantification. *Structure and Infrastructure Engineering*, 10(5):589–603, 2014. 56

[289] O Morales-Nápoles, A M Hanea, and D T H Worm. Experimental results about the assessments of conditional rank correlations by experts: Example with air pollution estimates. In ESREL, editor, *Proceedings of the 22nd European Safety and Reliability Conference Safety, Reliability and Risk Analysis*. Taylor & Francis, 2013. 51, 52, 84, 91, 147

[290] O Morales-Nápoles, D Kurowicka, and A Roelen. Eliciting conditional and unconditional rank correlations from conditional probabilities. *Reliability Engineering & System Safety*, 93(5):699–710, 2008. 40, 49, 50, 51, 54, 148

[291] O Morales-Nápoles, D Paprotny, D Worm, L Abspoel-Bukman, and W Courage. Characterization of precipitation through copulas and expert judgement for risk assessment of infrastructure. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, 3(4):3–18, 2017. 126, 147

[292] O Morales-Nápoles, D Paprotny, D T H Worm, L M Abspoel, and W Courage. Characterization of precipitation through copulas and expert judgement for risk assessment of infrastructure. In ESREL, editor, *Proceedings of the 25th European Safety and Reliability Conference Safety, Reliability and Risk Analysis*. Taylor & Francis, 2015. 89

[293] O Morales-Nápoles, D Worm, A M Hanea, and I Kalkman. Calibration and combination of expert's dependence estimates. *Reliability Engineering & Systems Safety*, under review. 92

[294] O Morales-Nápoles and D T H Worm. Hypothesis testing of multidimensional probability distributions. Technical report, WP4 GAMES2R TNO Report No. 0100003764, 2013. 91, 92, 147

[295] O Morales-Nápoles, D T H Worm, L M Abspoel, E Huibregtse, and W Courage. Trends and uncertainties regarding rain intensity in the netherlands. Technical report, TNO 2015 R10009, 2015. 84, 89

[296] J Morecroft. System dynamics. In *Systems approaches to managing change: a practical guide*, pages 25–85. Springer, UK, 2010. 103, 104

[297] M G Morgan, M Henrion, and M Small. *Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis.* Cambridge University Press, USA, 1992. 56

[298] H Moskowitz and D Bunn. Decision and risk analysis. *European Journal of Operational Research*, 28(3):247–260, 1987. 34

[299] H Moskowitz and R K Sarin. Improving the consistency of conditional probability assessments for forecasting and decision making. *Management Science*, 29(6):735–749, 1983. 51

[300] A Mosleh and G Apostolakis. The assessment of probability distributions from expert opinions with an application to seismic fragility curves. *Risk Analysis*, 6(4):447–461, 1986. 90

[301] S Nadkarni and P P Shenoy. A causal mapping approach to constructing bayesian networks. *Decision Support Systems*, 38(2):259–281, 2004. 39

[302] R B Nelsen. *An introduction to copulas.* Springer Science & Business Media, 2007. 119, 143

[303] L Norrington, J Quigley, A Russell, and R Van der Meer. Modelling the reliability of search and rescue operations with bayesian belief networks. *Reliability Engineering & System Safety*, 93(7):940–949, 2008. 88

[304] D W North. *The Invariance Approach to the Probabilistic Encoding of Information.* PhD thesis, Stanford University, Stanford, USA, 1970. 126

[305] U.S. Dept of Health, Food Human Services, and Center for Veterinary Medicine Drug Administration. Guidance for industry 152: Evaluating the safety of antimicrobial new animal drugs with regard to their microbiological effects on bacteria of human health concern. Technical report, FDA-CVM, 2003. 154

[306] A O'Hagan, C E Buck, A Daneshkhah, J R Eiser, P H Garthwaite, D J Jenkinson, J E Oakley, and T Rakow. *Uncertain judgements: eliciting experts' probabilities.* John Wiley & Sons, UK, 2006. 4, 5, 15, 16, 17, 33, 34, 35, 44, 45, 50, 53, 71, 74, 76, 80, 86

[307] A O'Hagan, E B Glennie, and R E Beardsall. Subjective modelling and bayes linear estimation in the uk water industry. *Applied Statistics*, pages 563–577, 1992. 47

[308] A O'Hagan and J E Oakley. Probability is perfect, but we can't elicit it perfectly. *Reliability Engineering & System Safety*, 85(1-3):239–248, 2004. 15, 17

[309] R A O'Leary, S Low Choy, J V Murray, M Kynn, R Denham, T G Martin, and K Mengersen. Comparison of three expert elicitation methods for logistic regression on predicting the presence of the threatened brush-tailed rock-wallaby petrogale penicillata. *Environmetrics*, 20(4):379–398, 2009. 47, 48

[310] F Ouchi. *A literature review on the use of expert opinion in probabilistic risk analysis.* World Bank, USA, 2004. 33

[311] D Over. Rationality and the normative/descriptive distinction. In *Blackwell Handbook of Judgment and Decision Making*, pages 3–18. USA, 2004. 70

[312] M Papathomas and A OHagan. Updating beliefs for binary variables. *Journal of Statistical Planning and Inference*, 135(2):324–338, 2005. 59

[313] G Parmigiani and L Inoue. *Decision theory: principles and approaches*, volume 812. John Wiley & Sons, UK, 2009. 24, 25

[314] E Paté-Cornell. On 'black swans' and 'perfect storms': risk analysis and management when statistics are not enough. *Risk Analysis*, 32(11):1823–1833, 2012. 104, 114, 133

[315] E Paté-Cornell and S Guikema. Probabilistic modeling of terrorist threats: A systems analysis approach to setting priorities among countermeasures. *Military Operations Research*, 7(4):5–23, 2002. 128

[316] S A Patterson and G E Apostolakis. Identification of critical locations across multiple infrastructures for terrorist actions. *Reliability Engineering & System Safety*, 92(9):1183–1203, 2007. 133

[317] J Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference.* Morgan Kaufmann, USA, 1988. 40, 82, 84, 88, 97, 116, 143, 161

[318] J Pearl. Causality: models, reasoning, and inference. *IIE Transactions*, 34(6):583–589, 2000. 39

[319] J Pearl. *Causality: models, reasoning, and inference.* Cambridge University Press, USA, 2009. 73, 82

[320] E S Pearson and J W Tukey. Approximate means and standard deviations based on distances between percentage points of frequency curves. *Biometrika*, 52(3/4):533–546, 1965. 46

[321] D F Percy. Bayesian enhanced strategic decision making for reliability. *European Journal of Operational Research*, 139(1):133–145, 2002. 45

[322] D F Percy. Subjective priors for maintenance models. *Journal of Quality in Maintenance Engineering*, 10(3):221–227, 2004. 45

[323] L D Phillips. A theory of requisite decision models. *Acta Psychologica*, 56(1):29–48, 1984. 105

[324] Pidd, M. *Systems Modelling - Theory and Practice*. John Wiley and Sons, UK, 2004. 98

[325] A Pollatsek, A D Well, C Konold, P Hardiman, and G Cobb. Understanding conditional probabilities. *Organizational Behavior and Human Decision Processes*, 40(2):255–269, 1987. 75

[326] C A Pollino, O Woodberry, A Nicholson, K Korb, and B T Hart. Parameterisation and evaluation of a bayesian network for use in an ecological risk assessment. *Environmental Modelling & Software*, 22(8):1140–1152, 2007. 88

[327] G Pólya. Heuristic reasoning and the theory of probability. *The American Mathematical Monthly*, 48(7):450–465, 1941. 70

[328] J Poplawska, A Labib, and D M Reed. From vicious to virtuous circles: problem structuring for quantified decision making in operationalization of corporate social responsibility. *Journal of the Operational Research Society*, 68(3):291–307, 2017. 104

[329] K R Popper. The propensity interpretation of probability. *The British Journal for the Philosophy of Science*, 10(37):25–42, 1959. 18, 24

[330] J Quigley, A Colson, W Aspinall, and R Cooke. Elicitation in the classical model. In L C Dias, A Morton, and J Quigley, editors, *Elicitation: The science and art of structuring judgement*, volume 261, chapter 2, pages 15–36. Springer International Series in Operations Research and Management Science, USA, 2018. 4, 129, 142

[331] E G Quijano, D Rios Insua, and J Cano. Critical networked infrastructure protection from adversaries. *Reliability Engineering & System Safety*, 2016. 128

[332] H Raiffa. Decision analysis: a personal account of how it got started and evolved. *Operations Research*, 50(1):179–185, 2002. 101

[333] F P Ramsey. Mr keynes on probability, 1922. 22

[334] F P Ramsey. Truth and probability. *The foundations of mathematics and other logical essays*, pages 156–198, 1926. 76

[335] F P Ramsey and E J Lowe. *Notes on philosophy, probability and mathematics*. UK, 1990. 21

[336] L Regis. A bayesian copula model for stochastic claims reserving. In *Actuarial and financial mathematics conference*, page 113, 2011. 44

[337] T Reilly. Sensitivity analysis for dependent variables. *Decision Sciences*, 31(3):551–572, 2000. 56

[338] O Renn, W J Burns, J X Kasperson, R E Kasperson, and P Slovic. The social amplification of risk: Theoretical foundations and empirical applications. *Journal of Social Issues*, 48(4):137–160, 1992. 98, 107

[339] S Renooij. Probability elicitation for belief networks: issues to consider. *The Knowledge Engineering Review*, 16(3):255–269, 2001. 39

[340] M Revie. *Evaluation of Bayes linear modelling to support reliability assessment during procurement*. PhD thesis, University of Strathclyde, UK, 2008. 46

[341] M Revie, T Bedford, and L Walls. Evaluation of elicitation methods to quantify bayes linear models. *Proceedings of the Institution of Mechanical Engineers, Part O - Journal of Risk and Reliability*, 224(4):322–332, 2010. 46, 56, 58, 84

[342] M Revie, T Bedford, and L Walls. Supporting reliability decisions during defense procurement using a bayes linear methodology. *IEEE Transactions on Engineering Management*, 58(4):662–673, 2011. 46, 47

[343] J Rios and D Rios Insua. Adversarial risk analysis for counterterrorism modeling. *Risk Analysis*, 32(5):894–915, 2012. 128

[344] D Rios Insua, J Rios, and D Banks. Adversarial risk analysis. *Journal of the American Statistical Association*, 104(486):841–854, 2009. 128

[345] A G L Romme. Making a difference: Organization as design. *Organization science*, 14(5):558–573, 2003. 8

[346] E A Rosa. The logical structure of the social amplification of risk framework (sarf): Metatheoretical foundations and policy implications, 2003. 30

[347] J Rosenhead. Past, present and future of problem structuring methods. *Journal of the Operational Research Society*, 57(7):759–765, 2006. 98

[348] H Rosoff and D Von Winterfeldt. A risk and economic analysis of dirty bomb attacks on the ports of los angeles and long beach. *Risk Analysis*, 27(3):533–546, 2007. 128

[349] C Rothery, K Claxton, S Palmer, D Epstein, R Tarricone, and M Sculpher. Characterising uncertainty in the assessment of medical devices and determining future research needs. *Health Economics*, 26(1):109–123, 2017. 154

[350] B M Rottman and R Hastie. Reasoning about causal relationships: Inferences on causal networks. *Psychological Bulletin*, 140(1):109–139, 2014. 73, 123

[351] G Rowe and G Wright. Expert opinions in forecasting: the role of the delphi technique. In *Principles of forecasting*, pages 125–144. Springer, 2001. 90, 108

[352] B Russell. On the notion of cause. *Proceedings of the Aristotelian Society*, 13:1–26, 1912. 72

[353] T P Ryan. *Modern regression methods*, volume 655. John Wiley & Sons, 2008. 47

[354] F Salmon. The formula that killed wall street. *Wired*, 3:16–20, 2009. 63

[355] W C Salmon. *The foundations of scientific inference*. University of Pittsburgh Press, USA, 1966. 19

[356] L J Savage. Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, 66(336):783–801, 1971. 23

[357] Savage, L J. *The foundations of statistics*. John Wiley and Sons, USA, 1954. 7, 13, 26, 98

[358] R D Shachter. Probabilistic inference and influence diagrams. *Operations Research*, 36(4):589–604, 1988. 87, 97

[359] D R Shanks. Judging covariation and causation. pages 220–239. Blackwell Publishing, USA, 2004. 34, 54

[360] C E Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948. 126

[361] Z Shen, M Odening, and O Okhrin. Can expert knowledge compensate for data scarcity in crop insurance pricing? *European Review of Agricultural Economics*, 43(2):237–269, 2015. 44

[362] J Shin, H Son, and G Heo. Development of a cyber security risk model using bayesian networks. *Reliability Engineering & System Safety*, 134:208–217, 2015. 133

[363] J Siebert, D Von Winterfeldt, and R S John. Identifying and structuring the objectives of the islamic state of iraq and the levant (isil) and its followers. *Decision Analysis*, 13(1):26–50, 2015. 128

[364] H A Simon. Information-processing theory of human problem solving. *Handbook of Learning and Cognitive Processes*, 5:271–295, 1978. 8

[365] H A Simon. Rational decision making in business organizations. *The American Economic Review*, 69(4):493–513, 1979. 7

[366] H A Simon. *The sciences of the artificial*. MIT Press, USA, 1996. 9

[367] Simon, H A. *Models of man; social and rational*. Wiley, UK, 1957. 7, 70

[368] M Sklar. Fonctions de repartition an dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231, 1959. 41

[369] S Sloman, A K Barbey, and J M Hotaling. A causal model theory of the meaning of cause, enable, and prevent. *Cognitive Science*, 33(1):21–50, 2009. 107

[370] S A Sloman and D Lagnado. The problem of induction. *The Cambridge handbook of thinking and reasoning*, pages 95–116, 2005. 75

[371] J Smedslund. The concept of correlation in adults. *Scandinavian Journal of Psychology*, 4(1):165–173, 1963. 34, 71, 78

[372] J E Smith and D Von Winterfeldt. Anniversary article: decision analysis in management science. *Management Science*, 50(5):561–574, 2004. 34

[373] M Soares and L Bojke. Expert elicitation to inform health technology assessment. In L C Dias, A Morton, and J Quigley, editors, *Elicitation: The science and art of structuring judgement*, volume 261, chapter 18, pages 479–494. Springer International Series in Operations Research and Management Science, USA, 2018. 140

[374] M O Soares, L Bojke, J Dumville, C Iglesias, N Cullum, and K Claxton. Methods to elicit experts beliefs over uncertain quantities: application to a cost effectiveness transition model of negative pressure wound therapy for severe pressure ulceration. *Statistics in Medicine*, 30(19):2363–2380, 2011. 140

[375] C S Spetzler and C A S Stael von Holstein. Exceptional paper-probability encoding in decision analysis. *Management Science*, 22(3):340–358, 1975. 35, 67, 80, 88, 122

[376] P Spirtes, C N Glymour, and R Scheines. *Causation, prediction, and search*. MIT Press, USA, 2000. 73

[377] C A S Stael von Holstein and J E Matheson. A manual for encoding probability distributions. Technical report, Decisions and Designs Inc., 1978. 67

[378] K E Stanovich and R F West. Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, 23(5):645–665, 2000. 7

[379] Stanovich, K E and West, R F. Advancing the rationality debate. *Behavioral and Brain Sciences*, 23(05):701–717, 2000. 7, 70

[380] START. Global terrorism database, 2016. data retrieved from: `https://www.start.umd.edu/gtd`. 130

[381] M D Stevenson, J Oakley, M Lloyd Jones, A Brennan, J E Compston, E V McCloskey, and P L Selby. The cost-effectiveness of an rct to establish whether 5 or 10 years of bisphosphonate treatment is the better duration for women with a prior fracture. *Medical Decision Making*, 29(6):678–689, 2009. 140

[382] T J Stewart, S French, and J Rios. Integrating multicriteria decision analysis and scenario planning-review and extension. *Omega*, 41(4):679–688, 2013. 101

[383] M Stone. The opinion pool. *The Annals of Mathematical Statistics*, 32(4):1339–1342, 1961. 90

[384] P Suppes. *A probabilistic theory of causality*. North-Holland Publishing Company, The Netherlands, 1970. 72, 73, 107

[385] P Suppes. Ramsey's psychological theory of belief. In *Cambridge and Vienna: Frank P. Ramsey and the Vienna Circle*, pages 35–53. Springer, 2006. 22, 23

[386] A D Swain and H E Guttmann. Handbook of human-reliability analysis with emphasis on nuclear power plant applications-final report. Technical report, US Nuclear Regulatory Commission NUREG/CR-1278, 1983. 57, 84

[387] J Tanwar, S Das, Z Fatima, and S Hameed. Multidrug resistance: an emerging crisis. *Interdisciplinary Perspectives on Infectious Diseases*, pages 1–8, 2014. 141

[388] K Tentori, V Crupi, and S Russo. On the determinants of the conjunction fallacy: probability versus inductive confirmation. *Journal of Experimental Psychology: General*, 142(1):235–255, 2013. 78

[389] M Thiollent. Uses of knowledge: some methodological alternatives. *Speciale Uitgave van Systemica Tijdsschrift van de Systeemgroep Nederland*, pages 115–124, 1985. 8

[390] A Tversky and D Kahneman. Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5(2):207–232, 1973. 79

[391] A Tversky and D Kahneman. Causal schemas in judgments under uncertainty. *Progress in social psychology*, 1:49–72, 1980. 71, 76, 77

[392] W Ulrich. Beyond methodology choice: critical systems thinking as critically systemic discourse. *Journal of the Operational Research Society*, 54(4):325–342, 2003. 105

[393] J Utts. What educated citizens should know about statistics and probability. *The American Statistician*, 57(2):74–79, 2003. 74

[394] J E Van Aken and G Romme. Reinventing the future: adding design science to the repertoire of organization and management studies. *Organization Management Journal*, 6(1):5–12, 2009. 8

[395] L C van der Gaag, S Renooij, C L M Witteman, B M P Aleman, and B G Taal. How to elicit many probabilities. In *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, pages 647–654. Morgan Kaufmann Publishers, 1999. 41

[396] K Van der Heijden. *Scenarios: the art of strategic conversation*. John Wiley & Sons, UK, 2011. 101, 103, 104

[397] J R Van Dorp. Statistical dependence through common risk factors: With applications in uncertainty analysis. *European Journal of Operational Research*, 161(1):240–255, 2005. 44, 57

[398] R J Vanderbei. *Linear programming*. Springer, 2015. 118

[399] C A Varum and C Melo. Directions in scenario planning literature-a review of the past decades. *Futures*, 42(4):355–369, 2010. 101

[400] J Venn. *The logic of chance*. MacMillan and Co, UK, 1876. 20

[401] A R Vieira, P Collignon, F M Aarestrup, S A McEwen, R S Hendriksen, T Hald, and H C Wegener. Association between antimicrobial resistance in escherichia coli isolates from food animals and blood stream isolates from humans in europe: an ecological study. *Foodborne Pathogens and Disease*, 8(12):1295–1301, 2011. 140

[402] G Villejoubert and D R Mandel. The inverse fallacy: An account of deviations from bayess theorem and the additivity principle. *Memory & cognition*, 30(2):171–178, 2002. 74

[403] R Von Mises. *Probability, statistics, and truth.* Courier Corporation, UK, 1957. 21

[404] J Von Neumann and O Morgenstern. *Theory of games and economic behavior (commemorative edition).* Princeton University Press, USA, 2007. 7

[405] Von Neumann, J and Morgenstern, O. *Theory of Games and Economic Behavior.* Princeton University Press, 1947. 7

[406] D Von Winterfeldt and T M O'Sullivan. Should we protect commercial airplanes against surface-to-air missile attacks by terrorists? *Decision Analysis*, 3(2):63–75, 2006. 128

[407] L Walls and J Quigley. Building prior distributions to support bayesian reliability growth modelling using expert judgement. *Reliability Engineering & System Safety*, 74(2):117–128, 2001. 67, 88

[408] R A Weinstein. Controlling antimicrobial resistance in hospitals: infection control and use of antibiotics. *Emerging Infectious Diseases*, 7(2):188–192, 2001. 141

[409] S Weisberg. *Applied linear regression*, volume 528. John Wiley & Sons, USA, 2005. 47

[410] V Werneck. Sobre o processo de construção do conhecimento: o papel do ensino e da pesquisa. *Ensaio: avaliação e políticas públicas em educação*, 14(51), 2006. 8

[411] C Werner. Report and R-tool manual: Conditional scenario mapping and dependence elicitation in higher education risk assessment. Technical report, University of Strathclyde, 2017. 108, 109, 110

[412] C Werner, T Bedford, R M Cooke, A M Hanea, and O Morales-Nápoles. Expert judgement for dependence in probabilistic modelling: a systematic literature review and future research directions. *European Journal of Operational Research*, 258(3):801–819, 2017. 81, 82, 85, 87, 90, 99, 117, 119, 136, 143, 147, 161, 164

[413] C Werner, T Bedford, and J Quigley. Sequential refi

ned partitioning for probabilistic dependence assessment. *Risk Analysis*, 2018. 161, 164

[414] C Werner, T Bedford, and J Quigley. Mapping conditional scenarios for knowledge structuring in (tail) dependence elicitation. *Journal of the Operational Research Society*, under review. 123, 131, 133, 144, 145, 161, 165

[415] C Werner, A M Hanea, and O Morales-Nápoles. Eliciting multivariate uncertainty from experts: Considerations and approaches along the expert judgement process. In L C Dias, A Morton, and J Quigley, editors, *Elicitation: The science and art of structuring judgement*, volume 261, chapter 8,

pages 171–210. Springer International Series in Operations Research and Management Science, New York, 2018. 113, 118, 123, 130, 133, 142, 143, 144, 147, 165

[416] D White. Application of systems thinking to risk management: a review of the literature. *Management Decision*, 33(10):35–45, 1995. 104

[417] P Whittle. *Probability via expectation.* Springer Science & Business Media, Germany, 1992. 19

[418] Australia WHO Working Group Consultation Canberra. Critically important antibacterial agents for human medicine for risk management strategies of non-human use. Technical report, WHO, 2005. 154

[419] T Williams. *Management science in practice.* Wiley, UK, 2008. 9, 158

[420] T M Williams, F Ackermann, C Eden, and S Howick. Project risk: systemicity, cause mapping and a scenario approach. *Managing risks in projects*, pages 343–352, 1997. 104

[421] H H Willis. Guiding resource allocations based on terrorism risk. *Risk analysis*, 27(3):597–606, 2007. 29, 128

[422] K J Wilson. An investigation of dependence in expert judgement studies with multiple experts. *International Journal of Forecasting*, 33(1):325–336, 2017. 90

[423] R L Winkler. Uncertainty in probabilistic risk assessment. *Reliability Engineering & System Safety*, 54(2-3):127–132, 1996. 14, 15, 16, 17, 20

[424] R L Winkler and R T Clemen. Multiple experts vs. multiple methods: Combining correlation assessments. *Decision Analysis*, 1(3):167–176, 2004. 91

[425] B C Wintle, F Fidler, P A Vesk, and Joslin Moore. Improving visual estimation through active feedback. *Methods in Ecology and Evolution*, 4(1):53–62, 2013. 108

[426] B W Wisse, S P van Gosliga, N P van Elst, and A I Barros. Relieving the elicitation burden of bayesian belief networks. In *BMA*, 2008. 40

[427] G Woo. Quantitative terrorism risk assessment. *The Journal of Risk Finance*, 4(1):7–14, 2002. 127, 129

[428] G Woo. *Calculating catastrophe.* World Scientific, 2011. 15, 17, 20, 25, 128, 129

[429] G Woo. Understanding the principles of terrorism risk moeling from the attack in westminster. , Risk Management Solutions Discussion Paper, 2017. 129

[430] G Wright and P Goodwin. Decision making and planning under low levels of predictability: Enhancing the scenario method. *International Journal of Forecasting*, 25(4):813–825, 2009. 71, 103, 104, 114

[431] J S Wu, G E Apostolakis, and D Okrent. Uncertainties in system analysis - probabilistic versus nonprobabilistic theories. *Reliability Engineering & System Safety*, 30(1-3):163–181, 1990. 15, 16

[432] K Wunderlich, M Symmonds, P Bossaerts, and R J Dolan. Hedging your bets by learning reward correlations in the human brain. *Neuron*, 71(6):1141–1152, 2011. 58, 84, 146

[433] I Yaniv and S Choshen-Hillel. Exploiting the wisdom of others to make better decisions: Suspending judgment reduces egocentrism and increases accuracy. *Journal of Behavioral Decision Making*, 25(5):427–434, 2012. 108

[434] J F Yates and E R Stone. *The risk construct*. John Wiley & Sons, USA, 1992. 27

[435] R K Yin. *Case study research and applications: Design and methods*. Sage Publications, USA, 2009. 9

[436] A Zagorecki and M J Druzdzel. An empirical study of probability elicitation under noisy-or assumption. In *Flairs Conference*, pages 880–886, 2004. 40

[437] R E Zapata-Vázquez, A O'Hagan, and L Soares Bastos. Eliciting expert judgements about a set of proportions. *Journal of Applied Statistics*, 41(9):1919–1933, 2014. 44, 45

[438] K Zwirglmaier and D Straub. Approaches to bayesian network structure elicitation. In ESREL, editor, *Risk, Reliability and Safety: Innovating Theory and Practice. Proceedings of the 26th European Safety and Reliability Conference*. Taylor & Francis. 87, 88