Confidence in safety argument - An assessment framework based on belief function theory

Rui Wang

Université Fédérale

Toulouse Midi-Pyrénées

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ FÉDÉRALE TOULOUSE MIDI-PYRÉNÉES

Délivré par :

l'Institut National des Sciences Appliquées de Toulouse (INSA de Toulouse)

Présentée et soutenue le *02/May/2018* **par :**

Rui WANG

Confidence in safety argument - An assessment framework based on belief function theory

---

JURY

Thierry Denoeux	Professeur d'Université, UTC	Rapporteur
Mario Trapp	Maître de Conférence (eqiv.), TU Kaiserslautern	Rapporteur
Barbara Gallina	Maître de Conférence (eqiv.), Mälardalens Université	Examinateur
Didier Dubois	Directeur de Recherche, CNRS	Examinateur
Jérémie Guiochet	Maitre de Conférence, CNRS	Directeur de thèse
Gilles Motet	Professeur d'Université, INSA de Toulouse	Directeur de thèse
Walter Schön	Professeur d'Université, UTC	Invité

École doctorale et spécialité :

EDSYS : Informatique 4200018

Unité de Recherche :

Laboratoire d'analyse et d'architecture des systèmes

Directeur(s) de Thèse :

Gilles Motet et *Jérémie Guiochet*

Rapporteurs :

Thierry Denoeux et *Mario Trapp*

To my unforgettable adventure in France

Acknowledgments

The work presented in this thesis has been realised in the Laboratory for Analysis and Architecture of Systems (LAAS), a laboratory depending from the French National Center for Scientific Research (CNRS). I would like to thank, first of all, Madam Karama Kanoun and Mr. Mohamed Kaâniche, CNRS research directors, the heads of research department Crucial Computing (IC) and research group Dependable Computing and Fault Tolerance (TSF) respectively, for having welcomed me to carry out my work in this laboratory and the research team.

I would like furthermore express my sincere gratitude to Professor Gilles Motet and Dr. Jérémie Guiochet, my supervisors of the Ph.D. thesis, for their tremendous supports during three and half years. Without their guidance, encouragement, patience and precious time for frequent discussions, I could not have eventually accomplish this challenging task. I do feel lucky for having the opportunity to work with you both. I would like also to thank the rest of my thesis defence committee:

- Thierry Denoeux, Professor, UTC, reviewer
- Mario Trapp, Associate professor, TU Kaiserslautern, reviewer
- Barbara Gallina, Associate professor, Mälardalens University, Examiner
- Didier Dubois, Research director, CNRS, Examiner
- Walter Schön, Professor, UTC, Inviter

I really appreciate their acceptance for reviewing my work, and the insightful comments and interesting questions during my defence to improve the manuscript and widen my research from various perspectives. Special thanks are given to Prof. Schön for your first calculation on a scratch paper made on the plane back to UTC, which led to the start of an important part of my work.

I would like to thank Dr. Didier le Botlan for giving me the chance to be a teaching assistant for two semesters in INSA Toulouse, which let me have a glance of the pedagogical ideas of the French engineering school and acquire some necessary experience for my future career. More appreciations are for the safety engineers and researchers who helped me to complete the experimental study for the first validation of our research work.

I am also very grateful to the cool colleagues in TSF, especially the doctoral students, with whom I went through these unforgeable three years of the Ph.D. study. Specific appreciations are sent to my comrades: Carla, Thierry, Ulrich, William,

Alienor, Matthieu, Guillaume, Kalou, Benoît; and to my other dear friends: Ci, Haiyang, Donghai, Zhuo, Lili, Yu, Xi, Nan, Lunde, Min for their continuous supports and pleasant company. I would like to give my special thanks to Xiaoning, who made my last days of this adventure even harder, but will make my future easier and brighter. In the end, I would like to express my deepest gratefulness to my parents for their unconditional love throughout my study in France and my life in general.

Contents

Introduction

Dependability of software applications has always been a concern for the stakeholders. It is especially true for safety-critical systems, such as aeronautic systems, railway, automotive, nuclear, etc. A structured safety argument is a common method in practice to justify sufficient assurance of the system safety. Usually, most safety arguments nowadays are textual, whereas there is a growing trend for graphical representing methods to structure arguments. Establishing an evidence-based argument is even now required by some functional safety standards of various industrial sectors, such as avionic systems, automotive, railway, safety-related electronic systems, etc. A safety argument has a top statement to be justified (e.g., "{system X} is acceptably safe" or "the failure rate of {system X} is less than 10^{-9}"). Nevertheless, some issues arise when assessing the safety argument relating to piles of evidence documents, especially for computing systems. A regulation body has to decide on the acceptability of this statement, and this decision is based on the confidence in the argument. The available argument does not provide such confidence directly, and it heavily depends on subjective expertise. Thus, a framework is needed to make explicit and measure the confidence in safety argument with the challenges as follows:

- Confidence definition

  Clarifying the confidence concept is undoubted of great importance for this issue. In fact, the definition of confidence requires identifying the factors that influence the system assurance. It can be understood as discovering the uncertainties in the supporting evidence and structure of an argument. For instance, the uncertainty could be how much we trust in the supporting evidence, full confidence or still having some doubts. It could also be the degree of the contribution of a piece of evidence to the top statement. Moreover, these uncertainties are often subjective and hard to determine an exact probabilistic distribution. Thus, a suitable uncertainty theory is necessary for a formal definition of the confidence placed in an argument.

- Aggregation rules

  The aggregation rules are essential for propagating the confidence in an argument. It varies depending on the independent and mutual contributions of different supporting evidence, which relate to the argument types. Several

pieces of evidence, belonging to the same top statement, can be complementary or redundant. Hence, it should be integrated into the aggregation rules. Similarly, the choice of a mathematical method is crucial for merging the mentioned uncertainties of confidence measures.

- Expert judgement extraction

  The values of confidence measures may come from the stochastic uncertainty from available data in evidence or subjective judgements from experts. The probabilistic or frequentist issue has been well explored. But transforming the subjective opinions to quantitative measures is a challenge. Thus, the expert judgement extraction is another critical problem to be solved in the quantitative assessment approach.

- Parameter estimation

  The quantitative framework of confidence assessment is expected to produce a parametric argument model. How to determine the parameters are of great importance. The expert judgement discussed above are mainly regarding the evaluation of the uncertainties in the evidence; and the parameters to be estimated here are related to the uncertainties in the argument structure, such as the weights of the evidence, argument types, etc. A feasible method or process is in need to complete the argument model.

The objective of this thesis is to propose a quantitative framework to formalize and assess our confidence in the safety argument. This framework aims to address the limitations mentioned above. We are intent to develop it based on an uncertainty theory, Dempster-Shafer theory (D-S theory). More specifically, this framework focuses on dealing with the issue of the argumentation assessment in the following aspects: 1) Formal definition of confidence in safety arguments and related assessment parameters based on *belief function and mass function* of D-S theory; 2) Development of confidence aggregation rules for structured safety arguments with Dempster combination rule; The Goal Structuring Notation (GSN) is adopted to establish the structured arguments. 3) A proposition of a quantitative assessment framework of safety arguments, which integrates a feasible method for expert opinion extraction. The parameter estimation will be studied based on a case study of railway safety cases.

This thesis will be organised as follows:

In Chapter 1, we elaborate the theoretical background and literature review of the related work to the thesis. This chapter is composed of 3 corresponding parts:

(1) the introduction of safety argument and its development methodologies; (2) the uncertainties theories, including probabilistic and non-probabilistic ones; and (3) the confidence assessment for the safety argument via qualitative and quantitative approaches.

In Chapter 2, a confidence assessment method for structured arguments is proposed. We formally define and aggregate this confidence consistently using the Dempster-Shafer theory. The definitions of the confidence assessment parameters and aggregation rules are demonstrated for the single argument, double-node argument, and n-node argument, respectively. At the end of this chapter, the sensitivity analysis of the aggregation rules is carried out. Most of the work has been published in two conference papers [Wang et al., 2016a,b].

In Chapter 3, we propose a 4-step confidence assessment framework for the safety case of a critical system. This work is implemented based on the quantitative model of the confidence assessment for the safety argument proposed in Chapter 2. This systematic framework provides solutions to determine the parameters, i.e. *trustworthiness* and the *appropriateness* of premises in arguments. For the *trustworthiness*, we integrate a method for the judgement extraction into the mathematical model. It makes the model more practical for a real engineering application. For the *appropriateness*, we propose a method to reuse the framework itself to derive the corresponding parameters based on the collected expert judgements. Moreover, some considerations for the evaluation of the other elements (contexts, justifications, and assumptions) in a safety case are given. Most of this work is published in the Conference SafeComp [Wang et al., 2017a].

In Chapter 4, a case study on railway safety cases is carried out. The safety assurance rationale behind the EN5012x series standards is identified through the construction of structural safety cases based on the standards. Then, the evaluation of the confidence assessment parameters is realised by a survey towards safety experts. This parameter evaluation, in turn, validates the feasibility of the proposed confidence assessment framework. With these study results, an application guideline of this framework is provided based on the Wheel Slide Protection (WSP) system. This is an extension of the published work in a journal paper [Wang et al., 2017b].

In Appendix A, the complete questionnaire for safety argument assessment discussed in Chapter 4 is presented.

Chapter 1

Theoretical background

Contents

1.1 Introduction

This thesis focuses on the issue of assessing the confidence in the safety arguments. Especially, the quantified methods are considered. The related works and theoretical background include the following 3 aspects: (1) safety arguments; (2) uncertainty theories; and (3) assessment approaches for safety arguments.

Thus, in this chapter, theoretical background and literature review are composed of 3 corresponding parts: (1) the introduction of safety argument and its development methodologies (Section 1.2); (2) the uncertainties theories, including probabilistic and non-probabilistic theories (Section 1.3); and (3) the confidence

assessment for the safety argument via qualitative and quantitative approaches (Section 1.4).

1.2 Safety argumentation

Structured arguments play important role in communicating a system's attributes with various names: safety case [Kelly and Weaver, 2004; Bishop and Bloomfield, 1998], assurance case [Bloomfield et al., 2006], trust case [Cyra and Gorski, 2007], dependability case [Bloomfield et al., 2007], etc. This thesis focuses on the ones arguing the system safety, that is, the safety arguments. The definitions of safety arguments and the relating requirements of standards are introduced. Then, we present the common approaches for argument representation.

1.2.1 Definition of safety argument

The notion of safety case has already been adopted in various safety-critical sectors. It is generally considered as *"a documented body of evidence that provides a convincing and valid argument that a system is adequately safe for a given application in a given environment"* [Bishop and Bloomfield, 1998]. Other similar definition may be: *a safety case should communicate a clear, comprehensive and defensible argument that a system is acceptably safe to operate in a particular context* [Kelly, 1998].

These definitions reveal the common features of a safety case. It is firstly an **argument** reasonably formulated and documented. The argumentation is supposed to be based on solid and considerable **evidence**. Then, the safety case is used to convince the stakeholders of the **objective** to be achieved, the adequate safety. All the argumentation should be within a certain **context**, due to the absolute safety of a system can hardly be reached. Finally, the argument should be **well structured** (*convincing valid, clear, comprehensive and/or defensive*) to facilitate the safety assessment, certification and maintenance process.

The development of the safety case is a common practice in demonstrating the system safety. This is mainly due to the fact that, in several sectors, the safety-related regulations require developing a safety case. In addition, some regulation bodies explicitly require building safety assurance arguments.

1.2.2 Safety argumentation in standards

The standard ISO26262 [2011] aims to address the functional safety of electronic control systems for road vehicles. It is an application of the standard IEC61508 [2010] to automotive domain. The *Safety Integrity Levels (SILs)* originated from IEC61508 are four discrete levels (SIL1-4) associated with the requirement of necessary risk reduction for the studied system. The SIL4 is the highest integrity level. The notion of SIL is called *Automotive Safety Integrity Level (ASIL)* in ISO26262. The corresponding four levels range from ASIL A to ASIL D, where the ASIL D is dedicated to the highest integrity level. In order to assess the ASIL, a risk analysis of each hazardous event needs to be performed by estimating the *severity*, *probability of exposure* and *controllability*. The safety requirements differ according to the ASILs.

This standard defines a safety case as:

> *"An argument that the safety requirements for an item*[1] *are complete and satisfied by evidence compiled from work products of the safety activities during development".*

In the Part 2, the standard explicitly requires developing a safety case, which is expected to document the evidence of the achievement of a certain ASIL:

> *"6.4.6* ***Safety case***
>
> *6.4.6.1 This requirement shall be complied with for items that have at least one safety goal with an ASIL A, B, C, or D: a safety case shall be developed in accordance with the safety plan.*
>
> *6.4.6.2 The safety case should progressively compile the work products that are generated during the safety lifecycle."*

Additionally, after its success for the guideline MISRA C, the organisation MISRA in automotive sector plans to issue *Guidelines for Automotive Safety Case Arguments* [MISRA, 2017]. It aims to provide practical guidelines to develop and review a safety case for electrical and/or electronic (E/E) systems embedded in vehicles. Gallina et al. [2013] propose an novel approach to construct safety cases for product lines in alignment with ISO26262. It aims to ensure the systematic reuse of the development and certification artefacts of safety-critical systems.

[1] Item: system or array of systems to implement a function at the vehicle level, to which ISO26262 is applied.

In railway domain, the functional safety of electronic systems for signalling are ensured by the EN5012X series standards: EN50126 [1999], EN50128 [2011], and EN50129 [2003]. They are also based on the IEC61508 [2010] standard. The safety case is considered as:

> *"The documented demonstration that the product complies with the specified safety requirements".*

Particularly, EN50129 introduces a high-level structure for any safety case of the railway signalling system. It clarifies the necessary evidence that justifies the rigorous development processes and safety life-cycle activities, which ensures the adequate confidence in the system safety. The structure of the safety case is mainly based on the acceptance conditions: 1) evidence of quality management, 2) evidence of safety management, 3) evidence of functional and technical safety.

In software engineering domain, the standard ISO/IEC15026-1 [2013] of *systems and software assurance* requires explicitly the development of an *assurance case*. Such cases are used to justify the system attributes, for example safety, reliability, maintainability, human factors, operability and security. Thus, the assurance case is often named as safety case or reliability and maintainability case. The definition of assurance case proposed in this standards is:

> *"Representation of a claim or claims, and the support for these claims".*
>
> *NOTE: An assurance case is reasoned, auditable artefact created to support the contention its claim or claims are satisfied. It contains the following and their relationships:*
>
> - *one or more claims about properties;*
> - *arguments that logically link the evidence and any assumptions to the claim(s);*
> - *a body of evidence and possibly assumptions supporting these arguments for the claim(s).*

Furthermore, Part 2 of this standard [ISO/IEC15026-2, 2011] provides the basic structure and contents of an assurance case in order to improve the communication among shareholders.

For aeronautic systems, the primary requirement of certification for the airbone software or software embedded in the CNS/ATM (Communication, Navigation, Surveillance and Air Traffic Management) systems is to be compliant with the

standard DO-178C/ED-12C [2011] or DO-278A/ED-109A [2011] published by the issuing bodies RTCA/EUROCAE. In both standards, the assurance case is recommended to be used while adopting an alternative method. It aims to present the bridge between evidence and *"the claims of compliance with the system safety objectives"*. For the safety justification of the complete systems, the assurance case is not explicitly required. However, the European Organisation for the Safety of Air Navigation (Eurocontrol) issues the Safety Regulatory Requirements BE [2001] for risk assessment and mitigation in ATM systems. This requirement ESARR4 with the target users of all ATM services providers clearly indicates the use of:

> *Correct and complete arguments to demonstrate that the constituent part under consideration, as well as the overall ATM System are, and will remain, tolerably safe* [2] *including, as appropriate, specifications of any predictive, monitoring or survey techniques being used.*

A *Safety Case Development Manual* [BE, 2006] issued by Eurocontrol is a guidance for the construction of Safety Cases. In this manual, the Goal Structuring Notation is recommended as a graphical representation method of safety argument, which will be introduced in next sub-section.

In the defence sector, the Defence Standards of UK successively clarify the definitions of safety case:

> *For software: "The software safety case shall present a well-organised and reasoned justification based on objective evidence, that the software does or will satisfy the safety aspects of the Statement of Technical Requirements Specification" [DEF STAN 00-55] [MoD, 1997].*

> *For system "The safety case is a structured argument, supported by a body of evidence that provides a compelling, comprehensible and valid case that a system is safe for a given application in a given environment" [DEF STAN 00-56] [MoD, 1996].*

Referencing to these two standards, the regulation of software on British Military Aircraft JSP 318B [MoD, 1999] (Regulation of the Airworthiness of Ministry of Defence Aircraft) clarifies that:

> *"Safety analysis is carried out to support the Safety Case. The safety analysis that is to be undertaken is detailed in DEF STAN 00-56 and, for software, in DEF STAN 00-55."*

[2] *"I.e., meeting allocated safety objectives and requirements" [BE, 2001].*

The JSP 318B lists the necessary requirements of a safety case. It should:

- *Define the configuration to which it applies.*
- *Describe the safety requirements, targets and attributes.*
- *Provide a justification for the airworthiness of the design; this means addressing both new equipment and systems, and the effect of changes to existing equipment and systems.*
- *Detail the evidence for airworthiness, including as appropriate the results of analyses, tests and trials carried out by the Designer, DERA4 Boscombe Down and other independent organisations, safety questionnaires for Service Engineered Modifications (SEMs), etc.*
- *Identify the limitations and procedures necessary to achieve the required level of safety for the subject configuration.*

We may summarize that in different safety critical domains, the safety argumentation is explicitly mentioned as an essential documentation to record the justified confidence in the system safety. This confidence comes from the arguments and supporting evidence in the safety case obtained from the rigorous development process. Standards are developed in the purpose for the industrial applications and productions. They are often relatively practical. Thus, an interesting difference of the safety case definitions between safety argumentation community and standard is the use of "acceptably/tolerably safe" as the objective, where in standards "compliance with safety requirements" is considered. Both are actual safety objectives. The "acceptably safe" is supported by the "compliance with safety requirements" (as clarified by BE [2001]), but at a higher level of abstraction. In other words, the "compliance with safety requirements" can be interpreted as an instantiation of "acceptably safe" objective. In this thesis, the modelling approach of a safety case is explored according to the safety assurance rationale or guidelines of safety case development in the mentioned standards.

1.2.3 Structuring approaches for safety argumentation

Despite various definitions of safety cases, there is a consensus that the argumentation of a safety case is used to link the safety objectives and safety evidence. These arguments should be well and reasonable structured. Many approaches are proposed for organising arguments. Some of them are general argumentation methods,

but not limited to safety justification. In this section, the following presentation approaches are discussed:

- Textual argumentation
- Govier's argument supporting notation
- Toulmin's argument model
- Goal Structuring Notation

1.2.3.1 Textual argumentation

Safety cases are commonly documented with plain text. Safety engineers use the natural language to describe how the safety targets are fulfilled. The plain text can be very flexible in expressing the safety arguments. However, the quality of such arguments strongly depends on the argument organisation. Kelly [Kelly, 1998] points out that the proficiency degree of using the written language impacts the expression of arguments. The unclear language semantics may bring ambiguity in the argumentation. Moreover, the cross-reference in the text would undermine the flow of the main argument.

EN50129 provides a guidance to develop a railway safety case. A part of safety argument template is extracted from this standard. This part of the safety case shows the suggested structure for safety justification with respect to system functional and technical safety. It is enumerated as the Section 2 of the safety case in the template. The contents of this section is shown as follows:

Section 2 Assurance of correct functional operation

- *2.1 System architecture description*
- *2.2 Definition of interfaces*
- *2.3 Fulfilment of System Requirements Specification*
- *2.4 Fulfilment of Safety Requirements Specification*
- *2.5 Assurance of correct hardware functionality*
- *2.6 Assurance of correct software functionality*

The rationale of the argument organization is that the system safety in terms of functional and technical safety can be ensured by achieving four targets listed in subsections 2.3-2.6. The first two subsections are the contextual information of the system. They are used to help understanding of the safety justification. However,

the great length of *system architecture description* and *interface definition* influences the main argument stream.

Let us take another specific example of textual argument. As a real safety case always has confidential issue, we present hereafter a fragment of argument in the railway domain. It is written based on the standard EN50129 and more precisely on its Subsection 2.4: Fulfilment of Safety Requirements Specification.

> *"This section aims to demonstrate how specified safety functional requirements of {System X} are fulfilled by the design. The {System X} Safety Requirements Specification [Doc1] obtained during the hazard analysis (see Hazard Analysis Documents [Doc2]) are introduced into the system design (see {System X} Functional Requirement Specification [Doc3] and {System X} Architecture Description [Doc4]). They are traced throughout the system development lifecycle (see {System X} V&V report [Doc5]). Safety team is responsible to ensure the completeness of the identified safety requirements and the fulfilment of each safety requirement by the design. The apportioned safety requirements should finally be analysed and validated at system level, that is, all hazards have been closed or reasonably explained (see Hazard Log [Doc6])".*

The strategies to ensure the fulfilment of safety requirements are proposed in this paragraph. Several documents are referred as safety evidence. However, it is not easy to have a clear image of the relationship among this premises in terms of safety assurance.

Furthermore, these limitations of textual arguments may cause the lack of understanding among the co-authors of the safety case, which leads to more difficulties for the maintenance and reuse of the safety case.

1.2.3.2 Govier's argument supporting notation

Govier [1991] discusses the arguments from the angle of philosophy. She emphasises the importance of arguments structure, especially for the complex arguments of which the conclusion (top claim) and premises are not fairly obvious. A clear argument structure presents the way how premises contribute to the conclusion. It helps to understand the reasoning of an argument, which significantly impacts the evaluation of an argument.

Considering the cooperative contribution of premises, a graphical notation is used to show different argument structures (see Figure 1.1). The premises and

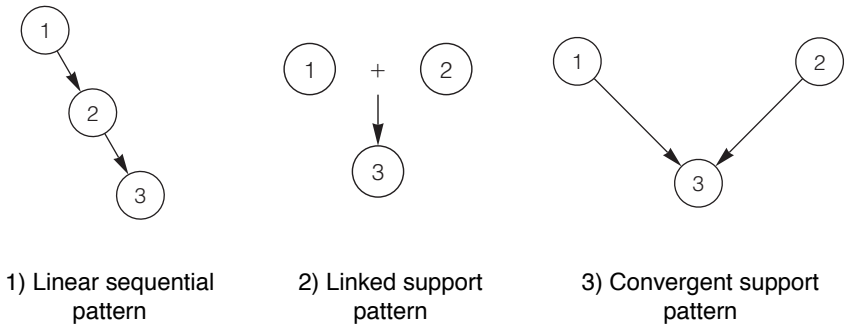


Figure 1.1: Govier's three patterns of argument structures [Govier, 1991]

conclusions are not distinguished in this notation. In Figure 1.1, ① and ② are the premises; and ③ represents the conclusion. Three basic argument patterns are proposed for the arguments with two or more premises:

1) *Linear sequential patten*

   Premises support the conclusion subsequently. In 1) of Figure 1.1, ② is deduced from ①; and ③ is deduced from ②. For this pattern type, the intermediate premise(s) can be regarded as sub-argument(s). All premises are necessary to obtain the conclusion.

2) *Linked support pattern*

   Premises shall be linked to support the conclusion. No conclusion can be deduced without any one of the premises. In 2) of Figure 1.1, both premises ① and ② are needed for conclusion ③. The falseness of either premise leads to the rejection of the conclusion based on this argument.

3) *Convergent support pattern*

   In contrast with the *linked support* argument, each premise of a *convergent support* argument contributes to the conclusion. In 3) of Figure 1.1, either premises ① or ② can hold the conclusion ③. If one of the premises is false, the other one is able to support ③. The truth of both premises increases the confidence in the conclusion due to that *"more dimensions the topic are considered"*.

A real argument may be composed of the combinations of these three patterns of support. For example, in Figure 1.2, the presented argument has 5 premises (①-⑤) and a main conclusion (⑥). This argument includes a *linear support* (① → ④), a *linked support* (② OR ③ → ⑤) and a *convergent support* (④ AND ⑤ → ⑥).

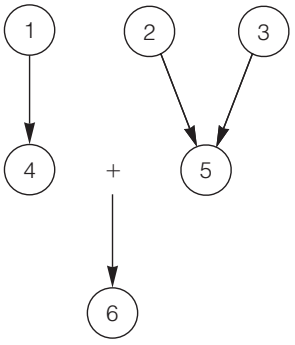


Figure 1.2: Combination of argument patterns of support [Govier, 1991]

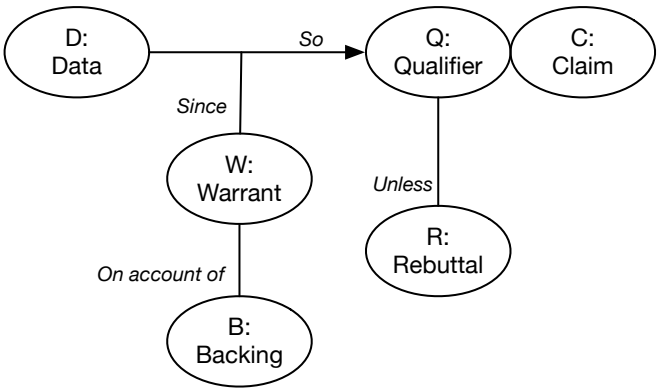


Figure 1.3: Toulmin argument model

1.2.3.3 Toulmin argument model

The philosopher Stephen Toulmin [1969] proposes a way to make the distinction of different elements in a realistic argument. This work highlights the essential compositions of a textual argument and the function of each argument element. The layout of an argument is abstracted as shown in Figure 1.3.

The six distinctive elements presented in Figure 1.3 are explained by Toulmin and interpreted by the author in terms of safety arguments as follows:

- *Claim (C): the claim or conclusion whose merits we are seeking to establish*

  A *claim* is a statement being argued, also known as a target, an objective or a goal in safety arguments.

- *Data (D): the facts we appeal to as a foundation for the claim*

  The *data* is considered as the facts or evidence used to prove the claim.

- *Warrants (W): the general, hypothetical statements that serve as bridges between the claim and the data*

The *warrants* are actually the intermediate logical steps (not always explicit) to establish the claim from the starting point of the available evidence.

- *Qualifiers (Q): the conditions, exceptions, or qualifications that limit the degree of force which our data confer on our claim in virtue of our warrant.*

  These *qualifiers* are the contexts or assumptions in which the inference from evidence to claim is bounded.

- *Rebuttals (R): conditions of exception*

  The *rebuttals* are the counter-arguments or statements indicating circumstances when the general argument does not hold true.

- *Backing (B): the statement based on which the warrants themselves would possess the authority or currency*

  The *backing* is the statements that serve to support the warrants (i.e., arguments that don't necessarily prove the main point being argued, but which do prove the warrants are true). It can be considered as justifications or contexts in safety arguments.

The Toulmin's notation of argument provides a typical pattern for an argument. According to this pattern, the statements in the argument are classified into six elements according to their various impacts on the argument reasoning. This distinction makes it possible to explicitly assess an argument, because both the strengths and limits are clarified for an argument. It reveals how an argument can get closer to the truth. The qualifiers and rebuttals help to present a comprehensive argument rather than a strong assertion.

1.2.3.4 Goal Structuring Notation

Based on the existing structuring approaches of argumentation, Kelly [1998] proposes a notation for the safety argument, called Goal Structure Notation (GSN). It helps to make the presentation of an argument more readable and adaptable. This notation is goal-based. It aims to break down the top goal into sub-goals until there are available evidence supporting the sub-goals. GSN allows the representation of the evidence, objective, argument, context, etc. The main elements of GSN are presented in Figure 1.4. An example of GSN is given in Figure 1.5. This safety argument fragment is derived from the Hazard Avoidance Pattern [Kelly and McDermid, 1997]. The explanation of the elements are listed below according to the work [Kelly, 1998]:

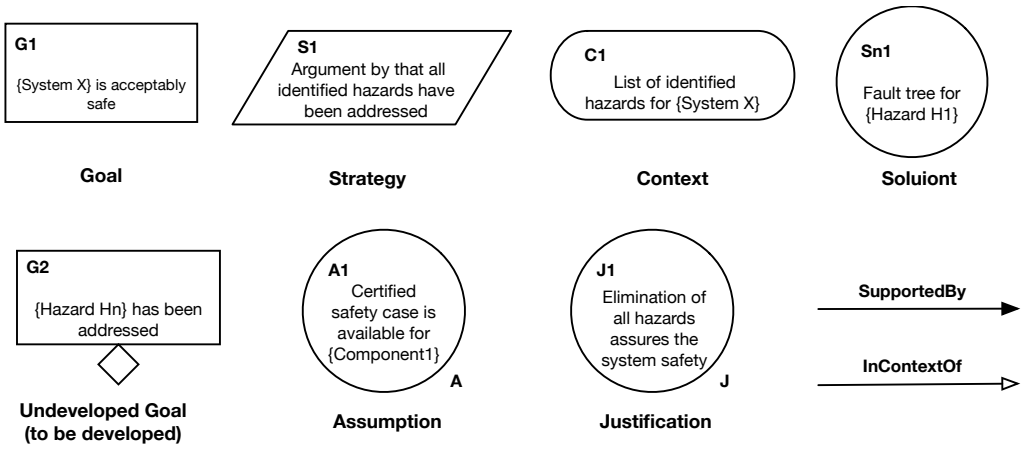


Figure 1.4: Main elements of the Goal Structure Notation

- Goal: the goal is the claim about the system in the aspects of the system design, implementation, operation or maintenance. For instance, a goal can be "G1: {System X} is acceptably safe". When a GSN safety argument is developed, sub-goals are introduced. The arrow denoted *SupportedBy* [GSN Standard, 2011] (see Figure 1.4) connects the parent and child goals (originally denoted *SolvedBy* in [Kelly, 1998]). In Figure 1.5, several sub-goal examples are given: "G2-Gn: Hazard H_i has been addressed".

- Solution: the available source of information to directly support a goal. Solutions may include all forms of evidence. For example, solution can be tests results, verification reports, fault tree (see Figure 1.4), etc.

- Strategy: the description of how to realise a goal decomposition. It always appears between parent and child goals. For instance, in Figure 1.5, strategy S1 shows how the goal "G1: {System X} is acceptably safe" is inferred from sub-goals.

- Context: a reference to contextual information, or a statement of contextual information. It can be related to a goal, a strategy or a solution. These elements are linked to the context object with the arrow denoted *InContextOf* shown in Figure 1.4. For example, a context can be "C1: List of identified hazards for {System X}" is the context in which G1 can be inferred from G2-Gn.

- Justification: statement or description that provides the rationale behind the adoption of some strategy or the presentation of some goal. For instance, a

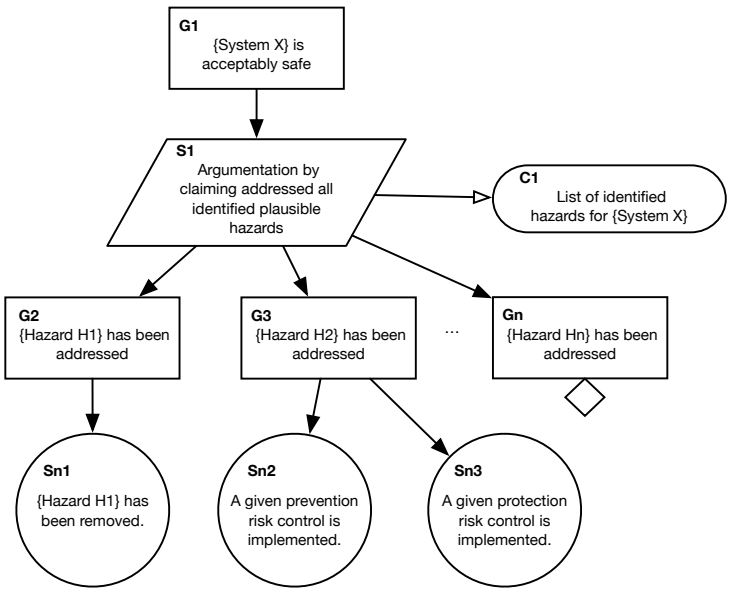


Figure 1.5: GSN example adapted from Hazard Avoidance Pattern [Kelly and McDermid, 1997]

justification can be *J1: Elimination of all hazards assures the system safety* (see Figure 1.4).

- Assumption: additional information linked to a goal or strategy, which is sometimes necessary for the goal decomposition. For instance, an assumption can be *A1: Certified safety case is available for* {*Component1*} (see Figure 1.4).

- Undeveloped goal - The undeveloped goal is the claim to be developed further, which is represented with $\diamond$ (see Figure 1.4).

1.2.3.5 Structured Assurance Case Metamodel

The Structured Assurance Case Metamodel (SACM) 2.0 is proposed by the Object Management Group [OMG, 2018]. This metamodel combines their efforts on the Structured Assurance Evidence Metamodel (SAEM) and Argumentation Metamodel (ARG). It mainly aims: on one hand, to facilitate establishing the structured assurance cases required by ISO/IEC15026-2 [2011]; on the other hand, to standardize the graphical notations increasingly used for structured safety cases, such as GSN introduced in previous subsection. This OMG technical report [OMG, 2018] presents the mappings between SACM and GSN

The OMG working group dedicated to SACM provides the normative specifications of metamodels for corresponding argument elements. The essential elements for an argument, in general, are premises, inference (argument structure) and conclusions. Hence, SACM includes three parts of the normative specifications with detailed descriptions, such as: superclass, attributes, associations, semantics, constraints, etc. These specification includes:

- Part 1. Common elements
  - SACM Base classes
  - SACM Packages
  - SACM Terminology
- Part 2. SACM Argumentation metamodel
- Part 3. SACM Artifact Metamodel

1.2.3.6 Discussion

After introducing the different argument structuring approaches, it is worth analysing the relationship among them. The textual argument is of the most common use for safety cases. However, it suffers from the variation of quality and high cost for maintenance. Toulmin's argument pattern makes explicit the six possible elements within a textual or a realistic argument. It helps readers to understand the reasoning process and to assess the degree of confidence in the final claim. Govier explores the argumentation in the sense of relationship among premises, rather than the function of each premise. We may consider that her support patterns are the extensions for Toulmin's *warrant*. It facilitates the analysis of the confidence propagation from premises to the conclusion (claim). This consideration is necessary for evaluating the complex argument. Kelly's GSN clarifies the argument elements based on Toulmin's notation; and this goal-based notation is also suitable for large scale arguments, especially for safety argument.

1.3 Uncertainty theories for confidence assessment

In this section, we present the possible sources of uncertainty in general. The limits of conventional probabilistic theory is discussed. New requirements of uncertainty measures and propagation are proposed. Based on these requirements, the emerging

uncertainty theories are summarized and compared to select one convenient and powerful mathematical tool to support the argumentation assessment process.

which may stem from complex development processes involving large amount of human decision-makings or merely the misunderstandings among software engineers, shareholders and end users. More assurance techniques are necessary to be added into development process. In turn,

1.3.1 Uncertainty concepts

Uncertainty is a general description of a state of knowledge impeding assessing truth values of propositions. In this subsection, we are going to identify the different categories of uncertainty sources and discuss the emerging uncertainty theories regarding to the probability theory.

1.3.1.1 Sources of uncertainty

The classification of uncertainty according to its sources sparks a long-term discussion. One proposition generally considered in dependability studies is that uncertainty is either **aleatory** or **epistemic** [Hacking, 1975]. Aleatory uncertainty is caused by the random variability of natural phenomena or repeatable events. The random varieties can be, for instance, the average temperature of next spring in Toulouse, the likelihood of the head or tail for a rolling coin or the sensor measurements. The epistemic uncertainty is due to the lack of knowledge. It may related to some subjective information like a testimony from a witness.

Dubois [2011] agrees with these two types of uncertainties and names them with more precise terms **randomness** and **incompleteness**. Besides, he identifies a third type of uncertainty: **inconsistency**. It may come from the conflicts among different sources of evidence (as in Shafer's evidence theory [Shafer, 1976]). For example, there are several conflicting testimonies or different opinions by experts on the same subject.

Blockley [2013] considers a third type of uncertainty, that is more complex and calls it **fuzziness** after Zadeh's work on fuzzy set [Zadeh, 1973]. It refers to imprecision or vagueness of definition, which presents as the implicit in the statement such as "the residual stresses in this welded steel structure are considerable". The fuzziness is epistemic in sense, since it can be reduced with more precise information. However, it often comes from the practical compromise between the precision and significance (or relevance) due to our ability while facing high complexity. Based this third type of uncertainty, this author proposes an uncertainty topology with the

Table 1.1: Three types of uncertainties and their sources

Authors	Uncertainties	Sources
Dubois [2011]	- Randomness	- The variability of observed natural phenomena
	- Incompleteness	- The lack of information
	- Inconsistency	- Conflicting testimonies or reports
Blockley [2013]	- Randomness	- The lack of a specific pattern or purpose in some data
	- Incompleteness	- What we do not know
	- Fuzziness	- Imprecision or vagueness of definition

orthogonal separate characteristics of uncertainties FIR (fuzziness, incompleteness and randomness).

Hence, these two authors make contributions to more detailed conceptual distinction of uncertainties according to their origins. This is believed to enrich the dual classification (aleatory and epistemic) for practical uncertainty assessment and decision-makings. The two propositions for uncertainty classification are summarized in Table 1.1.

1.3.1.2 Probability theory and the emerging uncertainty theories

Probabilistic approaches have been once assumed as the universal tools to express uncertainties. The frequentist probabilities is capable for modelling stochastic situation. However, the establishment of probability distributions is data demanding.

Aven [2010] deems the subjective probability is adequate to evaluate such uncertainties for risk analysis and the alternative uncertainty approaches are not necessarily helpful in risk assessment. However, there are fierce debates on the insufficiency of classical probability theory [Blockley, 2013; Dubois, 2010]. Dubois [2010] argues the single probability distribution can hardly express the lack of information and random variability at the same time. In the probabilistic setting, when the available information are scarce, the best strategy is to collect more information. Let us take an example. Equal probabilities are always used to show the total ignorance. Now, we would like to express our belief in whether the event X will happen. The universe set of the possible results is $\Omega = \{x, \bar{x}\}$. The probabilities of x, $\bar{x}$ are $p(x) = p(\bar{x}) = 0.5$. These probabilities can be interpreted as we have no information at all for the event X; or it can be explained as we are fully acknowledged that X is a pure random event. From the equal probabilities, we cannot tell the difference between these two situations. Denœux [1999] considers the precise number normally required in probabilistic theory is too strict, especially when there is little information available. Thus, the motivation to explicitly express uncertain infor-

mation arouses the attention of many researchers to the new emerging uncertainty theories in last decades.

Due to the imperfection of probabilistic theory, there are requirements for new approaches of representing uncertainties (e.g., see the International Journal of Approximate Reasoning). They shall be able to explicitly express incomplete information and more expressive than the unit interval. In addition, these approaches shall be less information demanding, compared with probability distributions. Certainly, they have to address the same issues faced by probability.

1.3.2 Uncertainty measures

The uncertainty theories mentioned previously are subject to measure and propagate the uncertainty in a general sense. As indicated by Smets [Smets and Kennes, 1994], one cannot select the best approach among them: each of the uncertainty theory serves for a specific problem or a domain of application. Therefore, we present various theories with the respect to their ways to measure uncertainty. We firstly clarify several basic principles for uncertainty measures. Then, 5 uncertainty theories are briefly introduced describing the concepts, properties and some operations. These theories include: probability in subjective setting (Section 1.3.2.2), imprecise probability (Section 1.3.2.3), fuzzy set theory (Section 1.3.2.4), possibility theory (Section 1.3.2.5), and belief function theory (Section 1.3.3).

1.3.2.1 Basic principles for uncertainties measures

In order to provide a general foundation for the comparison of different theories, three axioms for the measure of an uncertain event, also called a non-additive uncertain measure (or confidence measure), are specified (referred to [Gacôgne, 1997; Dubois and Prade, 2009]). These are the most basic notions defined on a set Ω of knowledge. In general, we speak of finite sets, in which case the events are all the parts of Ω. We can then denote a function c on [0, 1] as the confidence measure in a subset A or B of Ω. It should at least have the properties [Gacôgne, 1997; Dubois and Prade, 2009]:

$$c(\emptyset) = 0,\ c(\Omega) = 1 \tag{1.1}$$

The monotonicity with respect to inclusion:

$$A \subset B \Longrightarrow c(A) \leq c(B) \tag{1.2}$$

Based on (1.1) and (1.2), the following properties can be also deduced:

$$c(A \cup B) \geq max(c(A), c(B)) \tag{1.3}$$

$$c(A \cap B) \leq min(c(A), c(B)) \tag{1.4}$$

1.3.2.2 Probability in subjective setting

Probability theory is the first mathematical theory to describe and quantify uncertainties. It is developed based on Kolmogorov axioms in the 1930's. Briefly, in a sample space Ω, the probability of event A in Ω satisfies the Kolmogorov axioms:

First axiom:

$$P(\emptyset) = 0,\ P(\Omega) = 1 \tag{1.5}$$

Second axiom:

$$P(A) \in \mathbb{R},\ 0 \leq P(A) \leq 1 \tag{1.6}$$

Third axiom:

Any countable sequence of disjoint sets (synonymous with mutually exclusive events) A_1, A_2,... satisfies (the assumption of σ-additivity):

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i) \tag{1.7}$$

For the non-repeatable events, we may assign the probabilities due to the lack of knowledge, rather than the variability of the random outcomes. In this case, the probability is used in a subjective setting, also know as subjective probability. Although the subjective probability seems another interpretation of frequentist probability framework, De Finetti [1974] and his followers (Coletti and Scozzafava [2002]) indicate that the subjective approach starts from a set of Boolean propositions $\{A_j : j = 1, n\}$, not a sample space. An agent (one source of information) can assign coherent degrees of confidence c_i, and a set of logical constrains among these propositions. In this view, the subjective approach to probability can be considered as a transformation of the logical approach to knowledge representation, and of classical deduction [Adams and Levine, 1975]. Moreover, the difficulties for a single probability in distinguishing the incompleteness and pure randomness are the same for subjective probabilities.

1.3.2.3 Imprecise probability

"Imprecise probability" is a generic term to express mathematical models such as upper and lower probabilities, upper and lower previsions, possibility measures and necessity measures, belief function, plausibility function and other qualitative models. A set of generalized denotation of the imprecise probability is popularly employed [Walley, 2000]:

Suppose that event A is a subset of Ω. The lower and upper probabilities of A are denoted by $\underline{P}(A)$ and $\bar{P}(A)$ respectively, with

$$0 \leq \underline{P}(A) \leq \bar{P}(A) \leq 1$$

The interpretations of the lower probability $\underline{P}(A)$ could be: the measure of evidence in favour of A, and the measure of the belief of A. The interpretations of the upper probability $\bar{P}(A)$ could be: the measure of the lack of evidence against A and the measure of the plausibility of A.

Thus, based on available evidence, our confidence in A can be expressed with an interval. A precise probability of event A is a special case when $\underline{P}(A) = \bar{P}(A)$. Moreover, $\underline{P}(A) = 0$ and $\underline{P}(A) = 1$ represent complete lack of knowledge about A, with a flexible continuum in between. Some of the set-functions are directly used in the imprecise probability theory, such as $\underline{P}(A^c) = 1 - \bar{P}(A)$, where A^c is the complement of A.

The imprecise probability theory is also applied in practice [Ferson et al., 2003; Aughenbaugh and Paredis, 2006]. Authors of the work [Aughenbaugh and Paredis, 2006] extend traditional probability theory and incorporate imprecise probabilities, called probability boxes, or p-boxes (shown in Figure 1.6), which compassed by both of the upper and lower cumulative distribution function (cdf). The p-box using imprecise probabilities is suitable for the confidence measurement of a frequentist probabilistic model lacking for complete information.

1.3.2.4 Fuzzy set theory

Zadeh [1965] proposed fuzzy set theory more than fifty years ago. This theory copes with the flexibility in the meaning of natural language words. The proposer hopes to use the fuzzy set theory to capture the vagueness of the human reasoning. He provides a mathematical representation of fuzzy terms in natural language (e.g., tall person, hot water, etc.). These concepts are not sharp but fuzzy. Let take an example from the work of Werro [2016]. Concerning a concept of *a middle-aged*

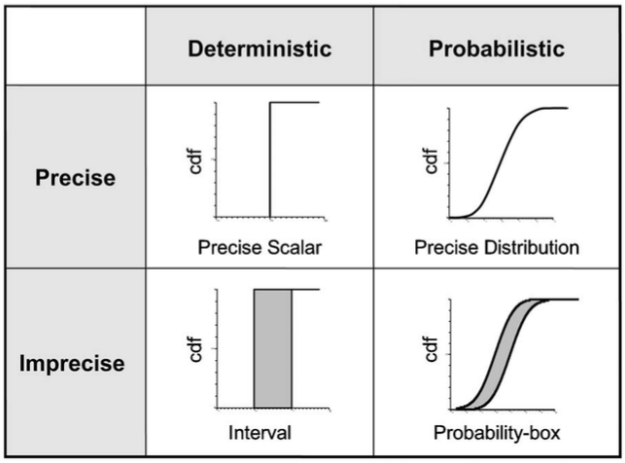
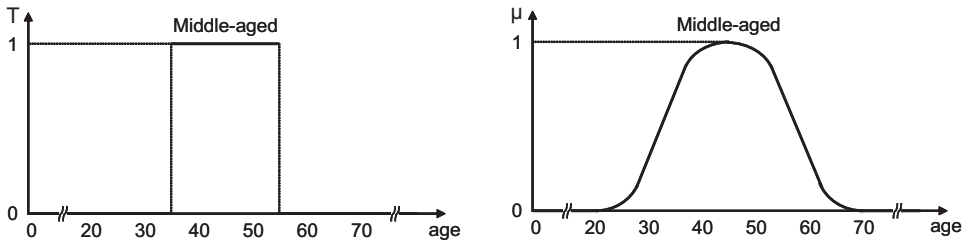


Figure 1.6: Dimensions of uncertainty [Aughenbaugh and Paredis, 2006]



a) Sharp definition b) Definition with a fuzzy set

Figure 1.7: Mathematical definitions of a middle-aged person [Werro, 2016]

person, we are clear with this concept but hard to give a precise definition. In Figure 1.7, two mathematical definitions are presented. The sharp definition shows that a person with age between 35 and 55 is a middle-aged person. The 0 and 1 represent the bivalent condition[3] for the relationship of a certain age and this concept. The second definition provides a partial membership to describe if an age belongs to the middle age. For a 34-year-old person, he/she is not a middle-aged person based on the first definition; and he/she may be considered as a middle-aged person according to the definition with a fuzzy set. Apparently, the definition in Figure 1.7b is more appropriate for expressing such fuzzy concept and closer to the human thinking.

Here, we recall some basic definitions of the fuzzy set theory.

Fuzzy set: A fuzzy set is built from a reference set called *universe of discourse*. The reference set is never fuzzy. Assume that $U = x_1, x_2, ..., x_n$ is the universe of discourse, then a *fuzzy set* A in $U (A \subset U)$ is defined as a set of ordered pairs

[3] An element either belongs or does not belong to the set

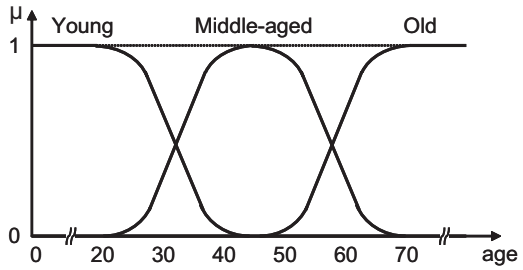


Figure 1.8: Fuzzy partition of the reference set with labelled fuzzy sets [Werro, 2016]

$$\{(x_i, \mu_A(x_i))\}$$

where $x_i \in U$, $\mu_A : U \to [0, 1]$ is the membership function of A and $\mu_A(x) \in [0, 1]$ is the degree of membership of x in A.

Example Consider the universe of discourse $U = 1, 2, 3, 4, 5, 6$. Then a fuzzy set A holding the concept 'large number' can be represented as

$$A = (1, 0), (2, 0), (3, 0.2), (4, 0.5), (5, 0.8), (6, 1)$$

With the considered universe, the numbers 1 and 2 are not 'large numbers', i.e. the membership degrees equal 0. Numbers 3 to 5 partially belong to the concept 'large number' with a membership degree of 0.2, 0.5 and 0.8. Finally number 6 is a large number with a full membership degree. [Werro, 2016]

Linguistic variable: A linguistic variable is characterized by a quintuple

$$(X, T, U, G, M)$$

where X is the name of the variable, T is the set of terms of X, U is the universe of discourse, G is a syntactic rule for generating the name of the terms and M is a semantic rule for associating each term with its meaning, i.e. a fuzzy set defined on U.

Example The linguistic variable represented in Figure 1.8 is defined by the quintuple (X, T, U, G, M) where X is 'age', T is the set {young, middle-aged, old} generated by G and M specifies for each term a corresponding fuzzy set on the universe $U = [0, 100]$.

The fuzzy set allows representing a partial belonging of the element by specifying the membership function. Besides, the properties and operations of the set theory are extended for the fuzzy set. This theory has been effectively applied in

many areas, such as fuzzy control [Zimmermann, 1996], fuzzy diagnosis [Chang et al., 2002], fuzzy data analysis [Denoeux and Masson, 2004; Denœux, 2011], fuzzy classification [Quost and Denoeux, 2016], fuzzy c-means clustering [Bezdek et al., 1984; Pal et al., 2005], etc.

1.3.2.5 Possibility theory

Possibility theory was first proposed by Zadeh [1978] on the basis of fuzzy set theory. Dubois and Prade [1988] made further contribution to the theory development. A pioneer of this theory is an economist Shackle, since he once introduced the min/max algebra to describe degrees of potential surprise [Shackle, 1961]. Like imprecise probability, possibility theory introduces two adjoint set functions to express uncertainty. The two functions are a possibility measure Π that is "maxitive" and a necessity measure N that is "minitive".

In other words, the possibility measure Π provides the most pessimistic or the most conservative measure for the confidence in the truth of events A and B. The definition of the possibility measure Π for the events A and B is:

$$\Pi(A \cup B) = max(\Pi(A), \Pi(B)) \tag{1.8}$$

For the opposite event A^C, we have:

$$max(\Pi(A), \Pi(\neg A)) = 1 \tag{1.9}$$

The necessity measure N is a confidence measure:

$$N(A \cap B) = min(N(A), N(B)) \tag{1.10}$$

Once a measure of possibility Π is defined, then the new measure N is defined by:

$$N(A) = 1 - \Pi(\neg A) \tag{1.11}$$

Conversely, if N is a measure of necessity, Π is defined by

$$\Pi(A) = 1 - N(\neg A) \tag{1.12}$$

These two are therefore dual notions.

Since we have $max(\Pi(A), \Pi(\neg A)) = 1$, this leads to the two properties:

$$\Pi(A) < 1 \Rightarrow N(A) = 0 \quad 0 < N(A) \Rightarrow \Pi(A) = 1 \tag{1.13}$$

These two properties are very important, because they mean that for the pair (N, Π) in [0, 1], one of them is always at one end of the interval (0 or 1) with $N \leq \Pi$.

Possibility theory can be regarded as a qualitative approach for handling uncertain information, which belongs to the non-Bayesian uncertainty calculi depending on probability bounds [Dubois and Prade, 2001].

1.3.3 Focus on belief function theory

The belief function theory, also known as Dempster-Shafer theory (D-S theory) or evidence theory, was developed by Arthur Dempster and Glenn Shafer successively. Dempster's work dealt with sample space probabilities using the upper and lower probabilities and the combination of sources of information [Dempster, 1966, 1967]. Shafer interpreted it in a subjective way. He defined the degree of belief and introduced the belief function [Shafer, 1976]. This theory offers a powerful tool to model human belief in evidence from different sources. Several basic concepts and used operations of this theory are recalled in this section, as they are reused in next chapter.

1.3.3.1 Mass function

Let X be a variable taking values in a finite set Ω representing a *frame of discernment*. Ω is composed of all the possible situations of interest, and 2^Ω is the power set of Ω. For example, let us consider the states of a bulb. We have $\Omega = \{on, off\}$, and its power set is $2^\Omega = \{\{on\}, \{off\}, \{on, off\}, \emptyset\}$. Let us note that $\Omega = \{on, off\}$ represents the ignorance about the state of the bulb.

The *mass function* on Ω (m^Ω) is the mapping of the power set of Ω on the closed interval [0,1] that is, $2^\Omega \rightarrow [0, 1]$. It is also called *basic belief assignment (BBA)*, or *basic probability assignment (BPA)* on the measure space $(\Omega, 2^\Omega)$. Assume that P is a subset of Ω, $P \subseteq \Omega$. Thus, P is an element of 2^Ω, that is, $P \in 2^\Omega$. The mass function must satisfy:

$$\sum_{P \subseteq \Omega} m^\Omega(P) = 1 \tag{1.14}$$

In belief function theory, the mass $m^\Omega(P)$ reflects the degree of belief committed to the hypothesis that the truth lies in P and to no subset of it. A subset P of Ω such as $m^\Omega(P) > 0$ is a *focal set* of belief mass m^Ω.

For instance, we can have the following assignment of belief, considered as one source of information: $m_1(\{on\}) = 0.5$, $m_1(\{off\}) = 0.3$, $m_1(\{on, off\}) = 0.2$. Note

that $m_1(\{on, off\})$ does not represent the belief that the bulb might be in {on} or {off} state, but the degree of belief in the statement "we don't know".

1.3.3.2 Dempster Combination

The combination of evidence called *joint mass* m_{12}^{Ω} aims to aggregate two masses m_1^{Ω} and m_2^{Ω} by Dempster's Rule:

$\forall P, M, N \subseteq \Omega$,

$$m_{12}^{\Omega}(P) = \begin{cases} \sum\limits_{M \cap N = P} \frac{m_1^{\Omega}(M) m_2^{\Omega}(N)}{1-K}, & \text{if } P \neq \emptyset, \\ m_{12}^{\Omega}(\emptyset) = 0, & \text{otherwise.} \end{cases} \tag{1.15}$$

where $K = \sum\limits_{M \cap N = \emptyset} m_1^{\Omega}(M) m_2^{\Omega}(N)$, representing *the degree of conflict.*

For instance, we have another source of information about the state of the bulb: $m_2(\{on\}) = 0.7$, $m_2(\{off\}) = 0.2$, $m_2(\{on, off\}) = 0.1$. We can get m_{12} through combination with $K = 0.31$: $m_{12}(\{on\}) = 0.78$, $m_{12}(\{off\}) = 0.19$, $m_{12}(\{on, off\}) = 0.03$.

Note that when K is nearly 1, this combination rule does not work. Such a situation occurs when 2 experts provide definitively contradictory opinions. Several works focus on this issue and propose other combination rules ([Yager, 1987]'s Rule, [Inagaki, 1991]'s Rule, etc.).

1.3.3.3 Belief function and plausibility function

The *belief function* is the sum of all the masses that support P. The function bel $(2^{\Omega} \rightarrow [0, 1])$ is defined as:

$$bel(P) = \sum_{M \subseteq P, M \neq \emptyset} m^{\Omega}(M) \quad \forall P \subseteq \Omega \tag{1.16}$$

For the bulb example, $bel_{12}(\{on\}) = m_{12}(\{on\}) = 0.78$.

The *plausibility function* is the sum of the masses that *might* support P. The function pl $(2^{\Omega} \rightarrow [0, 1])$ is defined as:

$$pl(P) = \sum_{M \subseteq \Omega, M \cap P \neq \emptyset} m^{\Omega}(M) \quad \forall P \subseteq \Omega \tag{1.17}$$

Following the same example, $pl_{12}(\{on\}) = m_{12}(\{on\}) + m_{12}(\{on, off\}) = 0.81$.

Smets [Smets, 1992] interpreted the belief function bel(P) for all $P \subseteq \Omega$ as *the degree of justified specific support* given to P, and the plausibility function pl(P) for

all $P \subseteq \Omega$ as *the degree of potential specific support* that could be given to P.

1.3.3.4 Useful tools of belief function theory

Many operations were introduced in [Mercier et al., 2005] on belief functions. We focus on three of them, presented hereafter and used later on.

- Discounting operation
  A mass m^Ω can be considered as a piece of information. The source of this information, for instance a person, may not be reliable. Thus, a discounting factor $v \in [0,1]$, representing the reliability of the source, is employed to make the mass m^Ω less informative and to increase the mass allocated to the ignorance Ω, that is, $m^\Omega(\Omega)$. $v = 0$ represents zero reliability of the source; on the contrary, $v = 1$ implies total trust in the source. The discounting operation is conducted as follows:

$$m_v^\Omega(P) = \begin{cases} v \cdot m^\Omega(P), & \text{if } P \neq \Omega, \\ 1 - v \cdot (1 - m^\Omega(\Omega)), & \text{if } P = \Omega. \end{cases} \tag{1.18}$$

  $m_v^\Omega(P)$ can be regarded as a weighted average between ignorance and $m^\Omega(P)$. For example, we estimate that the first source of information is only 80% reliable, that is, $v_1 = 0.8$. Thus, the updated mass is: $m_1(\{on\}) = 0.5 \times 0.8 = 0.4$, $m_1(\{off\}) = 0.3 \times 0.8 = 0.24$, $m_1(\{on, off\}) = 1 - 0.8 \times (1 - 0.2) = 0.36$.

- Vacuous extension
  Consider again the two frames of discernment Ω and Θ and the mass on $\Omega \times \Theta$. Then, a mass on one dimension of frame such as Ω, described as the least committed mass [Smets, 1993], can be extended to $\Omega \times \Theta$ without additional information. Then the definition of *vacuous extension* of m^Ω onto the product frame $\Omega \times \Theta$ is:

$$m^{\Omega\uparrow\Omega\times\Theta}(Q) = \begin{cases} m^\Omega(P), & \text{if } Q = P \times \Theta \text{ for some } P \subset \Omega, \\ 0, & otherwise \end{cases} \tag{1.19}$$

  Following the bulb example, we care about the state of a fan at the same time. The frame of discernment is $\Theta = \{spinning, stopped\}$. Now, we need to do some operations (such as combination, etc.) in the frame $\Omega \times \Theta$, we have to use the vacuous extension. Then, we have, for instance, the extended masses

for bulb states:

$$m_1^{\Omega\uparrow\Omega\times\Theta}(\{on\}\times\{spinning, stopped\}) = m_1^{\Omega}(\{on\}) = 0.5$$
$$m_1^{\Omega\uparrow\Omega\times\Theta}(\{off\}\times\{spinning, stopped\}) = m_1^{\Omega}(\{off\}) = 0.3$$
$$m_1^{\Omega\uparrow\Omega\times\Theta}(\{on, off\}\times\{spinning, stopped\}) = m_1^{\Omega}(\{on, off\}) = 0.2$$

- Marginalization

  $\Omega \times \Theta$ is a product of two frames of discernment Ω and Θ. A mass defined on $\Omega \times \Theta$ can be marginalized on Ω by transferring each mass $m^{\Omega\times\Theta}(Q)$ for $Q \subset \Omega \times \Theta$ to its projection on Ω:

$$m^{\Omega\times\Theta\downarrow\Omega}(P) = \sum_{Q\subset\Omega\times\Theta, Q\downarrow\Omega=P} m^{\Omega\times\Theta}(Q), \forall P \subset \Omega \tag{1.20}$$

  where $Q \downarrow \Omega$ denotes the projection of Q on Ω.

  Assume that we obtain other source of information about the states of the bulb and the fan: $m_3^{\Omega\times\Theta}(\{on\}\times\{spinning\}) = 0.2$, $m_3^{\Omega\times\Theta}(\{on\}\times\{spinning, stopped\}) = 0.4$, $m_3^{\Omega\times\Theta}(\{on, off\}\times\{spinning, stopped\}) = 0.4$. We would like to focus on the state of the bulb and to deduce the mass $m_3^{\Omega}(\{on\})$. The *marginalization* operation is needed.

$$\begin{aligned} m_3^{\Omega}(\{on\}) =& m_3^{\Omega\times\Theta\downarrow\Omega}(\{on\}) \\ =& m_3^{\Omega\times\Theta}(\{on\}\times\{spinning\}) + m_3^{\Omega\times\Theta}(\{on\}\times\{spinning, stopped\}) \\ =& 0.6 \end{aligned}$$

1.3.4 Relationship among uncertainty theories

Five uncertainty theories have been presented. We compare them to select an ideal uncertainty theory for our work in this thesis. Probability theory and imprecise probability possess the frequentist probability nature. The imprecise probability can express only the ignorance about frequency or subjective belief. They can not be described at the same time. Moreover, Dubois explains clearly in [Dubois and Prade, 2009] that the fuzzy set is "a pure logical form", which is not invented as an uncertainty theory to be an alternative of probability theory. Thus, we do not consider it as a tool to formalize the confidence in assurance case and confidence propagation. Furthermore, possibility theory is formally a special case of the belief function theory.

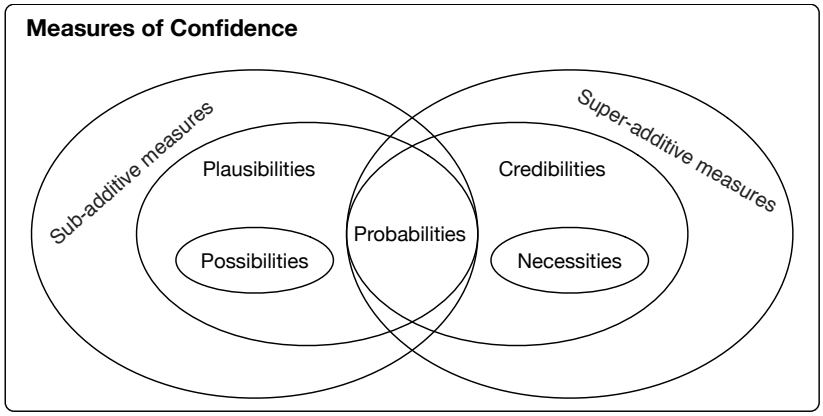


Figure 1.9: Illustration of nested relationship of measure types [Gacôgne, 1997]

Two papers discuss the mathematical generality of the methods for uncertainty measures. On one hand, Gacôgne [1997] provides a nested relationship to illustrate this conclusion, presented in Figure 1.9. As mentioned in Section 1.3.2.1, the uncertainty measures (or confidence measure) are non-addictive measures. The *sub-additive measures* and *super-additive measures* cover the largest area, which indicate the imprecise probabilities in general sense. In the descending order for the generality, the next measures are *plausibility* and *credibility* (belief function). These measures are dedicated to Dempster-Shafer theory. Then, we see the *possibility* and *necessity* from possibility theory. The most specific measure is the classical probability theory requiring precise distributions.

On the other hand, Dubois and Prade [2009] state that the following 3 methods are in the decreasing order of mathematical generality:

- Imprecise probability theory [Walley, 1991]
- Belief function theory (also known as Dempster-Shafer theory, D-S theory) [Dempster, 1967; Shafer, 1976]
- Possibility theory [Dubois and Prade, 1988, 2001]

Considering the expressiveness and the generality of the uncertainty theories, in this thesis, belief function theory is chosen as the mathematical tool to measure the confidence in an argument.

1.4 Confidence assessment of safety cases

Safety case is an important example of the structured arguments that is adopted for critical systems. It is used to present that a system is free from unacceptable risks. This argument often demonstrates the compliance of the system with safety requirements and includes a great deal of convincing evidence. Both developers of critical systems and regulation bodies have to spend a considerable amount of time on evaluating such argumentation, which aims to either produce trustful systems or make a justified decision for certification.

Several works have been done to help this confidence assessment process. They mainly address the problems in two perspectives. One type of approaches focuses on providing more justification with qualitative means. A second trend is the development of quantitative approaches to confidence in an argument. Indeed, excessive growth of argument leads it exhausted to manually estimate confidence in the system assurance. Therefore, quantitative tools might help analysts to estimate the confidence.

In this section, we briefly present the state-of-art of these two kinds of approaches for argument confidence assessment.

1.4.1 Qualitative approaches

Kelly and Weaver [2004] deem that the propositions made in an argument are subjective in nature; and uncertainties may exist in safety argument or supporting evidence [Hawkins et al., 2011]. These factors may impact the justification of the argument for the assurance of a system. The qualitative approaches mainly focus on identifying and reinforcing the factors.

The work of Menon et al. [2009] aims to minimise the uncertainty for goal-based or evidence-based standards. The authors specify the influencing factors from the supporting premises. For one *"leaf claim"*, three factors are concerned: *"replicability, trustworthiness and coverage"*; and for all supporting claims, there are four more factors: *"scope, user-defined importance, independence and reinforcement"*. All of these factors are of great importance for evaluating the supporting claims, but a checklist of influencing factors without formal definitions and propagation methods provides limited aid to the system assurance measurement.

Hawkins et al. [2011] focus on identifying the "defeaters" of an argument and propose to build a secondary argument, called confidence argument. This additional argument is in parallel with the safety case to support the assertions made in the argument. Three *Assurance Claim Points* (ACP) are identified for general

arguments:

- asserted inference (ACP1)
- asserted context (ACP2)
- asserted solution (ACP3)

The proposed confidence argument need to be developed for each ACP. A confidence argument should document the justification of the *appropriateness* and *sufficiency* of the inferences in arguments, the *trustworthiness* and *appropriateness* of the contexts and evidence. It explicitly communicates how to manage the uncertainties related to these argument elements. Following this work, Ayoub et al. [2012] put forward a systematic approach to identifying the assurance deficits results in the construction of confidence arguments. Then, they build another *contrapositive confidence argument* to show that the identified assurance deficits are adequately mitigated. The assurance claim points identified in these works are of great value to improve the soundness or quality of the argument. Nonetheless, it might produce a heavier argument for safety assurance.

1.4.2 Quantitative approaches

Considering that safety cases are more and more complex, it is less feasible to analyse the confidence in a complete safety case with only qualitative methods. According to the survey by Nair et al. [2015b], the quantitative approaches for evidence assessment are *"sometimes"* used in critical domains. Menon et al. [2009] have also mentioned the demand to combine and propagate the confidence measures within an argument. Hence, the issue of quantitatively assessing confidence in an argument has become a research interest over the last years.

Speaking of the approaches to quantitatively modelling the confidence in an argument, we find that most of the works are based on Bayesian Belief Networks (BBN) and Dempster Shafer theory (D-S theory). For instance, Guo [2003], Denney et al. [2011], and Hobbs and Lloyd [2012] propose to construct a BBN to represent an argument; and they assess and propagate the probabilistic confidence in this BBN. Meanwhile, Cyra and Gorski [2011], Ayoub et al. [2013], Duan et al. [2014] and Nair et al. [2015a] adopt D-S theory as the fundamental method to describe the uncertainties in the argument [4]. [Guiochet et al., 2015] put forwards a mixed

[4]The work [Nair et al., 2015a] applies the method of evidential reasoning, which can combine multiple assessments of individual facets of the evidence into a single. This method is developed based on D-S theory.

approach using both of these methods. Yuan et al. [2017] adopt the method of subjective logic, in which the confidence measures, called opinions, are also related to belief representation in D-S theory. Based on a critical study, we find that the BBN model requires too many inputs, and D-S theory have more advantages to explicitly express uncertainty.

In these works mentioned, Nair et al. [2015a] provide a method to extract the expert judgments and propagate these judgments based on belief theory. This method is used to build a confidence argument as proposed in Hawkins et al. [2011]. Nevertheless, they do not address the inference type (called "argument pattern" by Govier [1991]) when aggregating information. They also do not study how the confidence level could be used by the analysts to make a decision regarding the safety case.

Another approach based on D-S theory is presented by [Ayoub et al., 2013]. They introduce four argument types and corresponding formulas to combine confidence in arguments. They do not use the term *"confidence"*, but each goal is explicitly assessed with a *"belief, disbelief and uncertainty"* estimation for the statement. They suggest four argument types: *alternative, disjoint, overlap and containment*, but they provide little justification of the combining formulas. Moreover, no intuitive interpretation of the parameters of the aggregation rules is provided. Like the previous work, the results do not provide any justification for a decision regarding the acceptability of the safety case.

In the work of [Cyra and Gorski, 2011], the authors present a practical method for expert judgement extraction via the decision and confidence estimates and transform these estimates into belief theory parameters (belief, disbelief, and uncertainty). Moreover, they extend the work of [Govier, 1991] and propose six types of arguments, which are complicated for an intuitive identification in a real safety case. According to each of these types, their parameters are not apparent to determine and interpret.

In the approach proposed by Yuan et al. [2017], the number of subjective estimation values to be evaluated by experts is considerable. For an *"One-to-One"* argument, experts have to provide the 9 assessment values in total for 3 factors (*"confidence", "sufficiency", and "necessity"*) (no value for premise weight included). For two-node argument, the number of assessment values increases to 18. Moreover, the way to acquire these values is still an open issue according to the authors.

In a survey paper by Graydon and Holloway [2017], several other approaches are studied for quantitative assessment of safety argument confidence. Whereas these quantitative approaches for confidence assessment are of high interest, the

authors conclude that none of the methods is applicable. This is due to different limitations, such as lack of consideration to *"counterevidence"* [Ayoub et al., 2013], sensitivity to the arbitrary scope of hazards [Ayoub et al., 2013; Cyra and Gorski, 2011], difficulties to extract expert judgement [Nair et al., 2015a].

1.4.3 Identified issues for quantitative approaches

Regarding the existing related work, we find that this subject is of great research value. However, in the meantime, we identify several critical issues within the mentioned research works. Thus, we summarise these issues for developing a practical approach for quantitative confidence assessment of an argument. It should have the following features:

- Comprehensive and formal confidence definition
- Rigorous confidence aggregation rules
- Practical method of expert judgement extraction
- Feasible parameter estimation

1.5 Conclusion

In this chapter, we mainly introduce three important parts of knowledge for our work: 1) the safety argument, its relationship with various functional safety standards and the structuring approaches; 2) various uncertainty theories with the focus on belief function theory; 3) existing work on confidence assessment for the safety argument. These are the theoretical background to open up our research on the confidence assessment. Let's go back to summarise briefly each part.

Firstly, we discuss the definitions of safety cases in safety related standards of different critical domains. The requirements of using the safety argument emphasis its importance in safety assurance and certification. We compare several argument structuring methods and their evolving path. The GSN, presented as the goal-based notation, is suitable for large scale arguments; and it has a certain extent of recognition by several important domains, especially in UK [Kelly and Weaver, 2004]. Hence, in this thesis, we develop our confidence assessment approach based on the safety arguments via GSN method.

Then, the concept and classification of uncertainty are introduced. We make distinction between probabilistic and non-probabilistic uncertainty theories. Several

non-probabilistic methods are briefly presented; and belief function theory (D-S theory) is focused. Considering the generality of the D-S theory for uncertainty measurement, we propose to choose this method to implement our study.

Finally, the existing approaches for the confidence assessment of safety arguments are explored. Based on this work, we identify four critical issues for an ideal assessment approach to measure the confidence in a safety argument.

Chapter 2

Confidence Propagation in Safety Arguments

Contents

2.1 Introduction

The confidence in the system safety is commonly estimated through the safety arguments. Many works [Cyra and Gorski, 2011; Denney et al., 2011; Ayoub et al., 2012; Guiochet et al., 2015] were carried out to assess the confidence in the argument quantitatively. However, Graydon and Holloway [2017] conclude that whereas quantitative approaches for confidence assessment are of high interest, no method is currently fully applicable mainly due to the imperfect adaptability of the proposed methodologies to real safety arguments. Hence, a trustable and practical method or tool dedicated to this issue is still needed. In this chapter, we propose such a confidence assessment method for structured arguments. We formally define and aggregate this confidence consistently using the Dempster-Shafer theory. The definitions of the confidence assessment parameters and aggregation rules are demonstrated for the single argument, double-node argument, and n-node argument, respectively. At the end of this chapter, the sensitivity analysis of the aggregation rules is carried out. Most of the work has been published in papers [Wang et al., 2016b,a].

2.2 Confidence in an argument

Confidence, in common sense, is the feeling or belief that one can have faith in or rely on someone or something. The confidence or trust has been used by the dependability community to define dependability itself [Avižienis et al., 2004]:*the system dependability is the ability to deliver service that can justifiably be trusted.* This definition implies that the confidence (or trust) in the system dependability needs to be justified. The concept of dependability involves the following attributes of a specific system: *availability, reliability, safety, integrity, maintainability, and confidentiality.* These attributes are often justified through the structured argument, such as safety case [Kelly and Weaver, 2004; Bishop and Bloomfield, 1998], assurance case [Bloomfield et al., 2006], trust case [Cyra and Gorski, 2007], dependability case [Bloomfield et al., 2007], etc.

What we address in this thesis is the confidence in the justification of one of the dependability attributes (such as safety, etc.). To do so, we focus on the confidence in the structured argument. Thus, we may define the confidence in an argument as the belief in the truth placed in the top goal, and we will present in this section how this confidence could be formally defined.

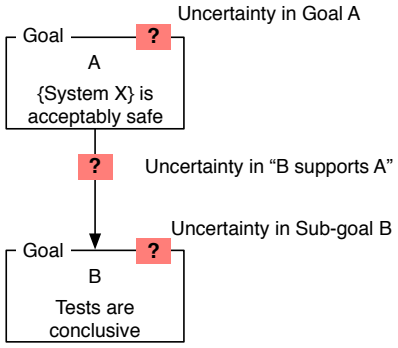


Figure 2.1: Sources of uncertainties in a simple inference modelled with GSN

2.2.1 Sources of uncertainty in an argument

To assess the confidence, we need to identify the potential uncertainties in the arguments. Taking a simple GSN safety argument as an example (shown in Figure 2.1): the top goal *A: {System X} is acceptably safe* is supported by the sub-goal *Tests are conclusive*. Two sources of uncertainties are identified, which are marked on the safety argument:

- Uncertainty in the fact that B is effectively supporting A

  For instance, do we consider that if "Tests are conclusive", then "{System X} is acceptably safe"? We may doubt this inference, that is, to which extent the claim A can be deduced from the claim B. We name a measure as *"appropriateness"* to estimate the degree of this doubt or certainty in the inference. The definitions of the appropriateness depend on the different argument structures. They are introduced in Section 2.3, 2.4 and 2.5, respectively.

- Uncertainty in the fact that B is True

  For instance, do we consider that "Tests are conclusive"? We may doubt this claim after evaluating the available associate evidence. We propose another measure of a sub-goal named *"trustworthiness"*, which assesses the degree of this doubt or certainty in claim B. The definition of the trustworthiness is universal for all claims, it is introduced in the following section (Section 2.2.2).

2.2.2 Formal definition of trustworthiness

A goal in GSN is always expressed by a statement (e.g., "Tests are conclusive"). Here, the assessment of trustworthiness of a goal is generally studied with focusing on a statement. Let's consider, for instance, a statement *A* "{System X} is

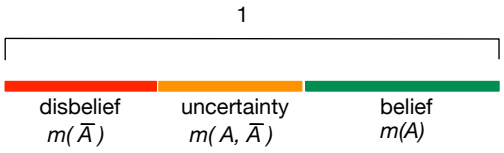


Figure 2.2: The measures of truth of statement A with D-S theory

acceptably safe". The frame of discernment Ω_A for the truth of A is binary: $\{A, \overline{A}\}$ or {True, False}. In Dempster Shafer theory, the *mass function* of set P reflects the degree of belief committed to that the truth is placed in P. Hence, the *trustworthiness* of a statement is formalised through assigning the *mass functions* to sets representing the *belief, uncertainty*, and *disbelief*. An opinion of the truth of this statement can be explicitly expressed with 3 masses represented in Figure 2.2. These measures are:

$$\begin{aligned}
&\textit{Belief in the statement A:} \quad bel_A = m^{\Omega_A}(\{A\}),\\
&\textit{Disbelief in the statement A:} \quad disb_A = m^{\Omega_A}(\{\overline{A}\}),\\
&\textit{Uncertainty in the statement A:} \quad uncer_A = m^{\Omega_A}(\{A, \overline{A}\}).
\end{aligned}$$

According to the constraint of the mass function, this leads to $m(\{A\}) + m(\{\overline{A}\}) + m(\{A, \overline{A}\}) = 1$, i.e., *belief + disbelief + uncertainty = 1*. Hence, we define the trustworthiness of statement of the goal A based on the *belief function and mass function* of Dempster-Shafer theory:

Definition 2.2.1 *The trustworthiness of a statement of the goal A is a three-tuple* $trust_A = (bel_A, uncer_A, disb_A)$*:*

$$trust_A : \begin{cases} bel_A \;\; = bel^{\Omega_A}(\{A\}) = m^{\Omega_A}(\{A\}) \\ disb_A \;\; = bel^{\Omega_A}(\{\overline{A}\}) = m^{\Omega_A}(\{\overline{A}\}) \\ uncer_A \;\; = m^{\Omega_A}(\{A, \overline{A}\}) = 1 - m^{\Omega_A}(\{A\}) - m^{\Omega_A}(\{\overline{A}\}) \end{cases} \tag{2.1}$$

where bel_A, $disb_A$, $uncer_A \in [0, 1]$. bel_A, $disb_A$ and $uncer_A$ denote the degree of our belief in, disbelief in or doubt about statement A. If *A="{System X} is acceptably safe"*, it depends on, for instance, the completeness of the test sequence, the correctness of the test results, the clarity of the evidence, the competence of

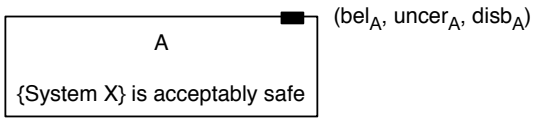


Figure 2.3: Goal A annotated with trustworthiness measures

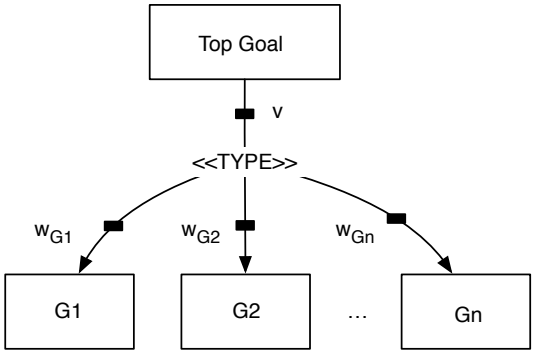


Figure 2.4: An argument annotated with appropriateness measures

the engineers, etc.

In Figure 2.3, the black rectangular annotation represents the trustworthiness of Goal A $(bel_A, uncer_A, disb_A)$.

2.2.3 Formal definition of appropriateness

As introduced in Section 2.2.1, the appropriateness is used to evaluate the inference between two layers of goals. In Figure 2.4, the Top Goal is supported by n sub-goals (G1-Gn). The GSN notation of *"SupportedBy"* arrow is modified to make the annotations for appropriateness measures. We consider three factors of the appropriateness that may influence the propagation of trustworthiness of sub-goals to the top-goal:

- The contributing weight of each sub-goal, $w_{G1}, w_{G2}, ..., w_{Gn}$. As the name indicates, the weight is used to measure the degree of the contribution of each sub-goal to the Top Goal.

- The cooperative contribution of the sub-goals, called *argument type* and annotated with *TYPE*. Govier [1991] presented a proposal for argument types, which are called "patterns in arguments" (see Section 1.2.3.2). They include, for example, the *linked and convergent support patterns*. These argument types aim to explicitly describe how the premises work together. We will

extend this in the next section.

- The overall reliability placed in the sources or the completeness of the sub-goals, denoted by v. As the available premises may be not enough to justify the full confidence in the Top Goal, this parameter provides a possible way to weaken the confidence obtained from the sub-goals. It is also known as a *discounting factor* in D-S theory.

Based on these factors, we propose a general definition of the appropriateness using the argument example in Figure 2.4:

Definition 2.2.2 *The appropriateness of the sub-goals ($appr_{\{G1,...Gn\}\rightarrow A}$, simplified into $appr_A$) regarding the Top Goal is specified with the factors in the following expression:*

$$appr_A = (w_{G1}, w_{G2}, ..., w_{Gn}, w_{TYPE}, v) \tag{2.2}$$

The w_{G1-GN}, w_{TYPE} and v correspond to the three factors that may influence the trustworthiness propagation mentioned above, where w_{TYPE} is a particular parameter to describe the argument type, which is only dedicated to the argument has more than one premise ($n > 1$). Specific definitions of the appropriateness will be given according to different argument structures in the following sections.

2.3 Single argument

The introduction of the confidence propagation starts from the simplest argument with one sub-goal. We call it the *single argument* (one-node argument), as shown in Figure 2.5. A single argument "*A is supported by B*" has only one premise. Note that we adopt the GSN notation *"is supported by"* arrow (from A to B), whereas, in this section, we will focus on the confidence propagation from B to A (bottom-up). The trustworthiness and appropriateness have been introduced in the previous sections. Here, we are going to present how to realise the trustworthiness propagation from B to A. The calculation here is dedicated to the simple case; and the same rationale is applied in the next sections for double-node and n-node arguments.

2.3.1 Appropriateness of the sub-goal

The appropriateness of sub-goals affects the trustworthiness of the top goal. The general definition of appropriateness is given in Section 2.2.3. Here, we specify the

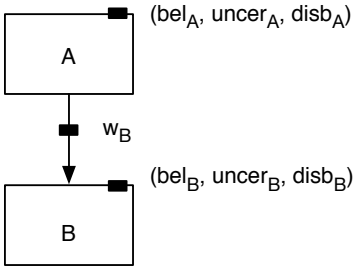


Figure 2.5: A single argument annotated with assessment parameters

corresponding factors for the single argument based the Definition 2.2. In this case, only the contributing weight and completeness of the sub-goal are considered.

Firstly, we start from the definition of the contributing weight, for instance the weight of sub-goal B (w_B) (see Figure 2.5). Based on the D-S theory, masses are used to express the degree of belief in certain states. Because the truth of a statement is binary, we propose to measure the appropriateness according to different states of inferences between B and A. A 2-tuple (X_B, X_A) presents the cross product $\Omega_B \times \Omega_A$, where X_B and X_A are elements of Ω_B and Ω_A, respectively ($\Omega_A = \{A, \overline{A}\}$, $\Omega_B = \{B, \overline{B}\}$). Therefore, the frame of discernment $\Omega_B \times \Omega_A = \{(\overline{B}, \overline{A}), (B, \overline{A}), (\overline{B}, A), (B, A)\}$. Among the elements of the frame, for example, $(\overline{B}, \overline{A})$ represents the situation: when B is false, A is false. In our approach, the appropriateness of B to A is measured with two possible cases:

- Mass assigned to the inferences ($\{(B, A), (\overline{B}, \overline{A})\}$). This mass is called the *contributing weight* of B, denoted with $w_{B\rightarrow A}$ (simplified as w_B). $\{(B, A), (\overline{B}, \overline{A})\}$ indicates the inferences that *A is true* can be inferred from *B is true*, and reversely *B is false* leads to *A is false*.

$$m^{\Omega_B \times \Omega_A}(\{(B, A), (\overline{B}, \overline{A})\}) = w_B, \qquad w_B \in [0, 1] \tag{2.3}$$

- Mass assigned to the ignorance case ($\Omega_B \times \Omega_A$). $\Omega_B \times \Omega_A$ is the simplified expression for $\{(\overline{B}, \overline{A}), (\overline{B}, A), (B, \overline{A}), (B, A)\}$. It represents the uncertainty whether B contributes to demonstrate the truth of A. According to the definition of the mass function in Equation 2.14, the sum of all the masses of focal sets shall equal to 1. In order be in compliance with this constraint, the rest of mass is assigned to $\Omega_B \times \Omega_A$.

$$m^{\Omega_B \times \Omega_A}(\Omega_B \times \Omega_A) = 1 - w_B \tag{2.4}$$

Note that when $w_B = 1$, i.e., B is fully appropriate to support A, then uncertainty $(1 - w_B)$ is equal to zero.

Also, the available premise may be not enough to justify the full confidence in the top goal. The discounting factor (v) is introduced in Definition 2.2. In D-S theory, the use of discounting factor aims to measure the reliability of the source of information (usually called an agent). It is adopted here to represent the reliability of the sources or the completeness of premises. According to the discounting operation presented in Equation 2.18, the support of sub-goal B shall be weakened and more mass is credited to uncertain state ($\Omega_B \times \Omega_A$). Thus, we define the *appropriateness* of the sub-goal as follows:

Definition 2.3.1 *The appropriateness of the sub-goal B to support the top goal A ($appr_{B\to A}$) is specified by the masses m_1 assigned to the subset $(\{(B, A), (\overline{B}, \overline{A})\})$ and $(\Omega_B \times \Omega_A)$ considering the discounting factor v:*

$$appr_{B\to A} : \begin{cases} m_1^{\Omega_B\times\Omega_A}(\{(B, A), (\overline{B}, \overline{A})\}) = w_B v \\ m_1^{\Omega_B\times\Omega_A}(\Omega_B \times \Omega_A) = 1 - w_B v \end{cases} \tag{2.5}$$

We define $w_B \in [0, 1]$ as the *contributing weight* of B, representing the degree that A depends on B; $v \in [0, 1]$ is the discounting factor that is used to evaluate the completeness of the available premises for A. When $v = 1$, it means that B sufficiently supports A and no other premise is needed. When $v = 0$, $m_1^{\Omega_B\times\Omega_A}(\Omega_B \times \Omega_A) = 1$ means that B does not provide any knowledge about A, i.e. a full uncertainty exists in A.

2.3.2 Trustworthiness of the sub-goal

Even if the argument B is appropriate to support A, we have to estimate the trustworthiness of B itself. The trustworthiness of a goal is introduced in the Definition 2.2.1. To combine these two types of confidence assessment measure of sub-goal B to obtain the *trustworthiness* of top goal A, we need to unify the masses assigned to different frames of discernment (Ω_B and $\Omega_B \times \Omega_A$) to the same one ($\Omega_B \times \Omega_A$). Thus, the operation of *vacuous extension*[1] is employed to realize this transformation (It is actually an extension of a mass defined in Ω_B to the frame of discernment

[1]Recall of the vacuous extension operation (detailed in Section 1.3.3):
The definition of *vacuous extension* of m^Ω onto the product frame $\Omega \times \Theta$ is:

$$m^{\Omega\uparrow\Omega\times\Theta}(Q) = \begin{cases} m^\Omega(P), & \text{if } Q = P \times \Theta \text{ for some } P \subset \Omega, \\ 0, & \textit{otherwise} \end{cases}$$

$\Omega_B \times \Omega_A$) . The masses m_2 present the trustworthiness of B extended to the frame $\Omega_B \times \Omega_A$ (represented by the up arrow $\uparrow$):

$$trust_B : \begin{cases} bel^{\Omega_B}(\{B\}) = m_2^{\Omega_B \uparrow \Omega_B \times \Omega_A}(\{B\} \times \Omega_A) = bel_B \\ bel^{\Omega_B}(\{\overline{B}\}) = m_2^{\Omega_B \uparrow \Omega_B \times \Omega_A}(\{\overline{B}\} \times \Omega_A) = disb_B \\ m^{\Omega_B}(\{B, \overline{B}\}) = m_2^{\Omega_B \uparrow \Omega_B \times \Omega_A}(\Omega_B \times \Omega_A) = uncer_B = 1 - bel_B - disb_B \end{cases} \tag{2.6}$$

Where $bel_B, disb_B, bel_B + disb_B \in [0, 1]$. $\{B\} \times \Omega_A$ is used instead of $\{(B, A), (B, \overline{A})\}$ and $\{\overline{B}\} \times \Omega_A$ instead of $\{(\overline{B}, A), (\overline{B}, \overline{A})\}$ to highlight the focus on B.

2.3.3 Confidence propagation of single argument

Our aim is to deduce the *trustworthiness* of A ($bel_A, disb_A, uncer_A$) based on the *trustworthiness* of B ($trust_B$, Equation 2.6) and the *appropriateness* of B to A ($appr_{B \to A}$, Equation 2.5). They can be regarded as two ways of observation for assessing A. These two sources of information can be combined with the help of Dempster's rule[2].

In order to illustrate the combining process of m_1 (Equation 2.5) and m_2 (Equation 2.6), the 6 possible combinations and *focal sets* in the frame $\Omega_B \times \Omega_A$ are shown in Table 2.1. The conflict factor K in this combination rule is 0, due to no conflict in this case. Our aim is to obtain the trustworthiness of A ($bel_A, disb_A, uncer_A$) in the frame Ω_A from the combined results on $\Omega_B \times \Omega_A$. Thus, the *marginalization* operation [3] shall be used. For example, the bel_A is obtained from the focal set $\{(B, A)\}$ underlined in Table 2.1:

$$bel^{\Omega_A}(\{A\}) = m^{\Omega_A}(A) = m^{\Omega_B \times \Omega_A \downarrow \Omega_A}(A) = m_{12}^{\Omega_B \times \Omega_A}(\{(B, A)\}) \tag{2.7}$$

[2] Recall of the Dempster's Rule (detailed in Section 1.3.3):
The *joint mass* m_{12}^{Ω} obtained through aggregating two masses m_1^{Ω} and m_2^{Ω} is:
$\forall P, M, N \subseteq \Omega$,

$$m_{12}^{\Omega}(P) = \begin{cases} \sum_{M \cap N = P} \frac{m_1^{\Omega}(M) m_2^{\Omega}(N)}{1-K}, & \text{if } P \neq \emptyset, \\ m_{12}^{\Omega}(\emptyset) = 0, & \text{otherwise.} \end{cases}$$

where $K = \sum_{M \cap N = \emptyset} m_1^{\Omega}(M) m_2^{\Omega}(N)$, denoting *the degree of conflict*.

[3] Recall of the marginalization operation (detailed in Section 1.3.3):
A mass defined on $\Omega \times \Theta$ can be marginalized on Ω by transferring each mass $m^{\Omega \times \Theta}(Q)$ for $Q \subset \Omega \times \Theta$ to its projection on Ω:

$$m^{\Omega \times \Theta \downarrow \Omega}(P) = \sum_{Q \subset \Omega \times \Theta, Q \downarrow \Omega = P} m^{\Omega \times \Theta}(Q), \forall P \subset \Omega$$

where $Q \downarrow \Omega$ denotes the projection of Q on Ω.

Table 2.1: Focal sets after the combination of $appr_{B\to A}$ and $trust_B$

		$\boldsymbol{m_1}$ $\boldsymbol{(appr_{B\to A})}$	
		$\boldsymbol{m_1^{\Omega_B\times\Omega_A}(\{(\overline{B},\overline{A}),(B,A)\})}$	$\boldsymbol{m_1^{\Omega_B\times\Omega_A}(\Omega_B\times\Omega_A)}$
$\boldsymbol{m_2}$ $\boldsymbol{(trust_B)}$	$\boldsymbol{m_2^{\Omega_B\uparrow\ \Omega_B\times\Omega_A}(\{B\}\times\Omega_A)}$	$\{\underline{(B,A)}\}$	$\{B\}\times\Omega_A$
	$\boldsymbol{m_2^{\Omega_B\uparrow\ \Omega_B\times\Omega_A}(\{\overline{B}\}\times\Omega_A)}$	$\{(\overline{B},\overline{A})\}$	$\{\overline{B}\}\times\Omega_A$
	$\boldsymbol{m_2^{\Omega_B\uparrow\ \Omega_B\times\Omega_A}(\Omega_B\times\Omega_A)}$	$\{(\overline{B},\overline{A}),(B,A)\}$	$\Omega_B\times\Omega_A$

Then, the *belief* in A is calculated according to Dempster's Rule is:

$$\begin{aligned} m_{12}^{\Omega_B\times\Omega_A}(\{B,A\}) =& \frac{m_1^{\Omega_B\times\Omega_A}(\{(\overline{B},\overline{A}),(B,A)\})\times m_2^{\Omega_B\uparrow\Omega_B\times\Omega_A}(\{B\}\times\Omega_A)}{1-K} \\ =& bel_B w_B v \end{aligned} \tag{2.8}$$

Thus, according to Equation 2.7:

$$bel^{\Omega_A}(\{A\}) = m_{12}^{\Omega_B\times\Omega_A}(\{B,A\}) = bel_B w_B v \tag{2.9}$$

where $bel_B, w_B, v \in [0,1]$.

Similarly, the $disb_A$ is obtained from the focal set $\{(\overline{B},\overline{A})\}$; the $uncer_A$ is calculated from the rest four focal sets in the Table 2.1. Therefore, we summarise that the trustworthiness of A $(bel_A, disb_A, uncer_A)$ is:

$$trust_A : \begin{cases} bel^{\Omega_A}(\{A\}) = m^{\Omega_A}(\{A\}) = bel_B w_B v \\ disb^{\Omega_A}(\{A\}) = m^{\Omega_A}(\{\overline{A}\}) = dis_B w_B v \\ uncer^{\Omega_A}(\{A\}) = m^{\Omega_A}((\{\overline{A},A\})) = 1-(bel_B+dis_B)w_B v \end{cases} \tag{2.10}$$

where $bel_B, disb_B, w_B, v \in [0,1]$.

2.4 Double-node argument

Besides the simplest argument with one premise, most arguments have a more complex structure with two or more premises. In this section, we employ the same approach to extend the confidence assessment of the single argument to the double-node argument. A symbolic double-node argument is presented in Figure 2.6: goal A is supported by two sub-goals B and C. Similarly, the confidence assessment

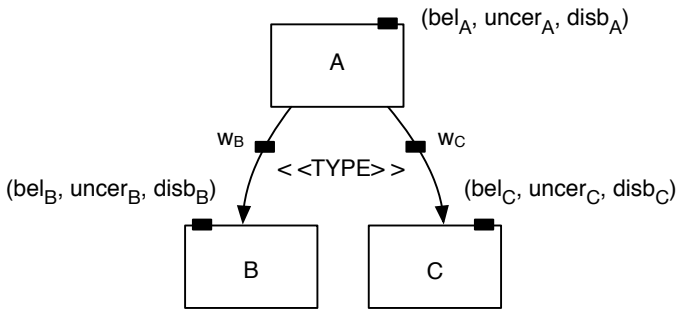


Figure 2.6: A double-node argument annotated with assessment parameters

parameters are annotated on this argument.

These parameters are:

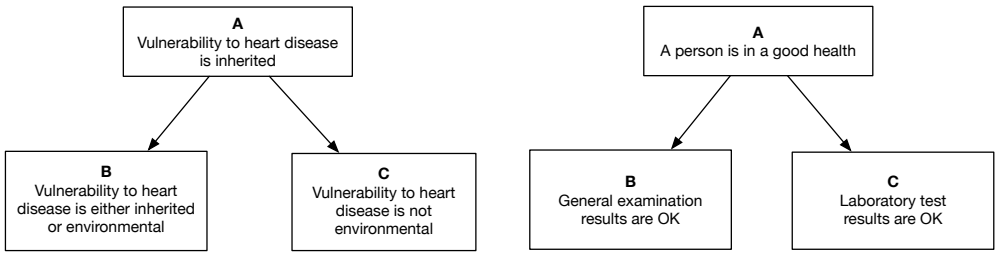
- The *appropriateness* of the sub-goals B and C (see the expression below), including the *contributing weights*, ($w_{B\to A}$, $w_{C\to A}$, simplified as w_B, w_C), the operator for the argument type (w_{TYPE}) and the completeness of the premises (discounting factor v). The argument types will be defined in this section.

$$appr_A = (w_B, w_B, w_{TYPE}, v) \tag{2.11}$$

- The *trustworthiness* of goals A: $trust_A = (bel_A, disb_A, uncer_A)$, B: $trust_B = (bel_B, disb_B, uncer_B)$ and C: $trust_C = (bel_C, disb_C, uncer_C)$.

2.4.1 Evolution of argument types

Initially, Govier [1991] emphasises the importance of the cooperative contribution of the premises in argument assessment. She proposes three argument patterns: *linear sequential pattern, linked support pattern ("pure AND") and convergent support pattern ("pure OR").* These patterns are considered regarding the logic reasoning. The statement can only be *true* and *false*. Referring to some other related works of Ayoub et al. [2013], Cyra and Gorski [2011] and Guiochet et al. [2015] discussing the types of arguments, most arguments are not always *"pure AND"* nor *"pure OR"* to infer a statement. Cyra and Gorski [2011] extend Govier's patterns by considering the argument with more complex inferences. In Figure 2.7, the inference of (a) is more obvious than (b). The good results of *general examination* and *laboratory test* can only contribute to the confidence in the person's health rather than completely justify it. This may be due to the existence of uncertainties, such as the undetectable diseases.



(a) Govier [1991]'s example (b) Cyra and Gorski [2011]'s example

Figure 2.7: Argument examples with different inference complexities

Thus, Cyra and Gorski [2011] define two types of argument and several sub-types:

- Type 1: the falsification of a single premise leads to the rebuttal of the conclusion, including NSC-argument (Necessary and Sufficient Condition list argument), SC-argument (Sufficient Condition list argument) and the combination of NSC-argument and SC-argument

- Type 2: the falsification of one of the premises decreases, but not nullifies, the support for the conclusion, including C-argument (Complementary argument), A-argument (Alternative argument) and the combination of C-argument and A-argument

The proposal of the complementary and alternative arguments are very appropriate for naming the mutual contribution of the premises. However, there are only the descriptive definitions for these argument types; and the two combination cases are not clearly explained regarding both definition and treatments. In consequence, the corresponding confidence aggregation rules are proposed with little justification and consistency for different types.

Then, Ayoub et al. [2013] introduce characterisation of the argument types based on the degree of the overlap between premises. The argument types are *alternative, disjoint, overlap and containment.* It is similar to Cyra and Gorski [2011]'s classification except for the type of *overlap.*

By these related works on the argument type classification, in next section, we propose an approach to formally define the argument types as a part of the appropriateness of sub-goals. The definitions are under the framework of D-S theory, which will provide the convenience for developing the confidence aggregation rule in the further step. These argument types include all the argument types mentioned

above. A comparison table of different types of argument is presented after the introduction of the appropriateness.

2.4.2 Appropriateness of the sub-goals

This subsection presents how sub-goals of a double-node argument support a top goal. For the single argument, we evaluate the contribution of the sub-goal alone. The *contributing weight* of the sub-goal is introduced in the previous subsection. When there are multiple sub-goals, their mutual influence to the higher-level goal shall be examined at the same time. The distinction between the disjoint and joint contributions of sub-goals B and C is to be made. Thus, we consider that sub-goals B and C support the top goal A in four different ways.

We propose to assign masses to model these four ways based on D-S theory. Since three statements A, B and C are concerned, a three-dimension frame of discernment is adopted $\Omega = \Omega_B \times \Omega_C \times \Omega_A$ (following the order of inference $B, C \rightarrow A$). It is equivalent to $\{(\overline{B}, \overline{C}, \overline{A}), (B, \overline{C}, \overline{A}), (\overline{B}, C, \overline{A}), (\overline{B}, \overline{C}, A), (B, C, \overline{A}), (\overline{B}, C, A), (B, \overline{C}, A)(B, C, A)\}$. The subsets of Ω are used to denote the possible inferences among A, B and C: e.g. $\{(\overline{B}, \overline{C}, \overline{A})\}$ stands for "when both B and C are false, A is false".

All frames of discernment of trustworthiness and appropriateness need to be unified to $\Omega = \Omega_B \times \Omega_C \times \Omega_A$ with the help of *the vacuous extension*. For instance, the mass in the frame $\Omega_B \times \Omega_A$, $m_1^{\Omega_B \times \Omega_A}(\{(B, A), (\overline{B}, \overline{A})\})$, in Definition 2.3.1 turns to $m_1^{\Omega_B \times \Omega_A \uparrow \Omega}(\{B\} \times \Omega_C \times \{A\} \cup \{\overline{B}\} \times \Omega_C \times \{\overline{A}\})$, which is simplified into $m_1^{\Omega}(\{B\} \times \Omega_C \times \{A\} \cup \{\overline{B}\} \times \Omega_C \times \{\overline{A}\})$.

In our approach, the different ways that B and C support A correspond to four "pure" cases: *pure B alone, pure C alone, pure AND and pure OR*. They will be used as basic elements to describe the "mixed" cases of complex arguments described hereafter. The appropriateness of sub-goals for these "pure" cases are respectively formalised as follows (the discounting factor v will be discussed later):

- *Pure B alone*: A exclusively depends on B. This case is equivalent to the single argument with the weight of the sub-goal equal to 1.

$$m^{\Omega}(\{B\} \times \Omega_C \times \{A\} \cup \{\overline{B}\} \times \Omega_C \times \{\overline{A}\}) = w_B = 1 \tag{2.12}$$

  w_B is the *contributing weight* of B, denoting the degree that A depends on B.

- *Pure C alone*: A exclusively depends on C.

$$m^{\Omega}(\Omega_B \times \{C\} \times \{A\} \cup \Omega_B \times \{\overline{C}\} \times \{\overline{A}\}) = w_C = 1 \tag{2.13}$$

w_C is the *contributing weight* of C, denoting the degree that A depends on C.

- *Pure AND*: B and C contribute to A with an AND logic gate.

$$m^{\Omega}(\{(\overline{B}, \overline{C}, \overline{A}), (\overline{B}, C, \overline{A}), (B, \overline{C}, \overline{A}), (B, C, A)\}) = w_{B \times C \to A} = 1 \tag{2.14}$$

$w_{B \times C \to A}$ denotes the degree of AND gate relation between B and C when they contribute to A.

- *Pure OR*: B and C contribute to A with an OR logic gate.

$$m^{\Omega}(\{(\overline{B}, \overline{C}, \overline{A}), (\overline{B}, C, A), (B, \overline{C}, A), (B, C, A)\}) = w_{B+C \to A} = 1 \tag{2.15}$$

$w_{B+C \to A}$ denotes the degree of OR gate relation between B and C when they contribute to A.

As discussed in the Section 2.4.1, not all the arguments belong to the "pure" cases. In fact, most arguments are not. Therefore, we propose two "mixed" types for common arguments (*complementary and redundant arguments*). Then, three particular types (*fully complementary and fully redundant arguments and disparate argument*) are derived based on these two types for some limit cases. These argument types are formally distinguished by the different definitions of the appropriateness of the sub-goals to the top goal. Similarly, the discounting factor v, evaluating the completeness of the available premises for A, is taken into account in the following definitions.

- *Complementary argument (C-Arg)* is the combination of *"Pure B alone"* (Equation 2.12), *"Pure C alone"* (Equation 2.13) and *"Pure AND"* (Equation 2.14). With the constraint that the sum of all the masses of focal sets is equal to 1, $w_B + w_C + w_{B \times C \to A} = 1$ (before the consideration of discounting factor v). Thus, $w_{B \times C \to A} = 1 - w_B - w_C$, denoting the degree of the complementarity between sub-goals.

Definition 2.4.1 *The appropriateness of sub-goals for complementary argument (C-Arg) is defined as:*

$appr_{\{B,C\}\to A}$:

$$\begin{cases} m_1^{\Omega}(\{B\}\times\Omega_C\times\{A\}\cup\{\overline{B}\}\times\Omega_C\times\{\overline{A}\}) = w_B\cdot v \quad (Pure\ B\ alone) \\ m_1^{\Omega}(\Omega_B\times\{C\}\times\{A\}\cup\Omega_B\times\{\overline{C}\}\times\{\overline{A}\}) = w_C\cdot v \quad (Pure\ C\ alone) \\ m_1^{\Omega}(\{(\overline{B},\overline{C},\overline{A}),(\overline{B},C,\overline{A}),(B,\overline{C},\overline{A}),(B,C,A)\}) = w_{B\times C\to A}\cdot v \quad (Pure\ AND) \\ m_1^{\Omega}(\Omega) = 1-v \end{cases} \tag{2.16}$$

where $v, w_B, w_C \in [0,1]$, and $w_{B\times C\to A} = 1 - w_B - w_C \geq 0$.

- *Redundant argument (R-Arg)* is the combination of *"Pure B alone"* (Equation 2.12), *"Pure C alone"* (Equation 2.13) and *"Pure OR"* (Equation 2.15). Similarly, $w_B + w_C + w_{B+C\to A} = 1$. $w_{B+C\to A} = 1 - w_B - w_C$, denoting the degree of the redundancy between sub-goals.

Definition 2.4.2 *The appropriateness of sub-goals for redundant argument (R-Arg) is defined as:*

$appr_{\{B,C\}\to A}$:

$$\begin{cases} m_1^{\Omega}(\{B\}\times\Omega_C\times\{A\}\cup\{\overline{B}\}\times\Omega_C\times\{\overline{A}\}) = w_B\cdot v \quad (Pure\ B\ alone) \\ m_1^{\Omega}(\Omega_B\times\{C\}\times\{A\}\cup\Omega_B\times\{\overline{C}\}\times\{\overline{A}\}) = w_C\cdot v \quad (Pure\ C\ alone) \\ m_1^{\Omega}(\{(\overline{B},\overline{C},\overline{A}),(\overline{B},C,A),(B,\overline{C},A),(B,C,A)\}) = w_{B+C\to A}\cdot v \quad (Pure\ OR) \\ m_1^{\Omega}(\Omega) = 1-v \end{cases} \tag{2.17}$$

where $v, w_B, w_C \in [0,1]$, and $w_{B+C\to A} = 1 - w_B - w_C \geq 0$.

- *Fully complementary argument (FC-Arg).* When $w_{B\times C\to A} = 1$ for the *complementary argument (C-Arg)*, we call this argument as *Fully complementary argument (FC-Arg).* It corresponds to the *"Pure AND"* case (Equation 2.14). $w_{B\times C\to A} = 1$ denotes the full complementarity between sub-goals. The appropriateness of sub-goals for *fully complementary argument (FC-Arg)* is:

$appr_{\{B,C\}\to A}$:

$$\begin{cases} m_1^{\Omega}(\{(\overline{B},\overline{C},\overline{A}),(\overline{B},C,\overline{A}),(B,\overline{C},\overline{A}),(B,C,A)\}) = w_{B\times C\to A}\cdot v = v \\ m_1^{\Omega}(\Omega) = 1 - w_{B\times C\to A}\cdot v = 1 - v \end{cases} \tag{2.18}$$

where $v \in [0,1]$

- *Fully redundant argument (FR-Arg).* When $w_{B+C\to A} = 1$ for the *redundant argument (R-Arg)*, we call this argument as *Fully redundant argument (FR-Arg).* It corresponds to the *"Pure OR"* case (Equation 2.15). $w_{B+C\to A} = 1$ denotes the full redundancy between sub-goals. The appropriateness of sub-goals for *fully redundant argument (FR-Arg)* is:

  $appr_{\{B,C\}\to A}$:

$$\begin{cases} m_1^{\Omega}(\{(\overline{B},\overline{C},\overline{A}),(\overline{B},C,A),(B,\overline{C},A),(B,C,A)\}) = w_{B+C\to A}\cdot v = v \\ m_1^{\Omega}(\Omega) = 1 - w_{B\times C\to A}\cdot v = 1 - v \end{cases} \tag{2.19}$$

  where $v \in [0,1]$

- *Disparate argument (D-Arg)* is the combination of *"Pure B alone"* (Equation 2.12) and *"Pure C alone"* (Equation 2.13). It can be seen as the limit case of redundant with $w_{B\times C\to A} = 0$ or the limit case of complementary with $w_{B+C\to A} = 0$. Then, $w_B + w_C = 1$. The appropriateness of sub-goals for *disparate argument (D-Arg)* is:

  $appr_{\{B,C\}\to A}$:

$$\begin{cases} m_1^{\Omega}(\{B\}\times\Omega_C\times\{A\}\cup\{\overline{B}\}\times\Omega_C\times\{\overline{A}\}) = w_B\cdot v \\ m_1^{\Omega}(\Omega_B\times\{C\}\times\{A\}\cup\Omega_B\times\{\overline{C}\}\times\{\overline{A}\}) = w_C\cdot v \\ m_1^{\Omega}(\Omega) = 1 - v \end{cases} \tag{2.20}$$

  where $v, w_B, w_C \in [0,1]$, and $w_B + w_C = 1$.

We have presented all the argument types proposed. In general, there are two types of argument: *redundant argument* and *complementary argument.* The *fully redundant/complementary argument* can actually evolve to *disparate argument* by

Table 2.2: Comparison of different proposals for argument classification

	Govier [2013]	Cyra and Gorski [2011]		Ayoub et al. [2013]	Guiochet et al. [2015]	Proposal in this thesis
Argument Types	Convergent	-		-	Alternative	FR-Arg
	-	Type2	A-argument/ Combination of A,C-argument	Alternative/ Overlap/ Containment		R-Arg
	-		C-argument/ Combination of A,C-argument	Disjoint	-	D-Arg
	-	-		-	Complementary	C-Arg
	Linked	Type1		-		FC-Arg
	Linear sequential	-		-	Simple argument	Simple argument

continuously changing the values of the weights. For completeness, the single argument is included. We make a comparison among different proposals of classification of argument types in Table 2.2. As shown in this table, our proposal of the argument type classification encloses all argument types mentioned in related work.

2.4.3 Trustworthiness of the sub-goals

The trustworthiness of a goal is defined in the Definition 2.2.1. In order to aggregate the two types of confidence assessment measures of sub-goals B and C, the trustworthiness shall be in the frame of discernment $\Omega = \Omega_B \times \Omega_C \times \Omega_A$. With the help of the operation of vacuous extension, the trustworthiness of sub-claims B and C are:

$$trust_B : \begin{cases} bel^{\Omega_B}(\{B\}) = m_2^{\Omega_B \uparrow \Omega}(\{B\} \times \Omega_C \times \Omega_A) = bel_B \\ bel^{\Omega_B}(\{\overline{B}\}) = m_2^{\Omega_B \uparrow \Omega}(\{\overline{B}\} \times \Omega_C \times \Omega_A) = disb_B \\ m^{\Omega_B}(\{B, \overline{B}\}) = m_2^{\Omega_B \uparrow \Omega}(\Omega) = uncer_B = 1 - bel_B - disb_B \end{cases} \tag{2.21}$$

$$trust_C : \begin{cases} bel^{\Omega_C}(\{C\}) = m_3^{\Omega_C \uparrow \Omega}(\Omega_B \times \{C\} \times \times \Omega_A) = bel_C \\ bel^{\Omega_C}(\{\overline{C}\}) = m_3^{\Omega_C \uparrow \Omega}(\Omega_B \times \{\overline{C}\} \times \Omega_A) = disb_C \\ m^{\Omega_C}(\{C, \overline{C}\}) = m_3^{\Omega_C \uparrow \Omega}(\Omega) = uncer_C = 1 - bel_C - disb_C \end{cases} \tag{2.22}$$

Where $bel_B, disb_B, bel_B + disb_B, bel_C, disb_C, bel_C + disb_C \in [0, 1]$.

2.4.4 Confidence aggregation for complementary arguments

All confidence assessment parameters of sub-goals have been identified for the double-node argument. Different argument types have varying behaviours in terms of their contribution to the confidence in A. In this section, the confidence aggregation for complementary arguments is considered. The aim is still to calculate the *trustworthiness* of top goal A ($bel_A, disb_A, uncer_A$) based on the combination of the *appropriateness* of sub-goals to A (Equation 2.16) and the *trustworthiness* of sub-goals (Equation 2.21 and 2.22). This combination is realised by employing the Dempster Rule (see Footnote 2). Regarding the definitions of masses for the confidence assessment parameters, the issue of the combination conflict is avoided due to the way to define the masses.

The masses of assessment parameters m_1, m_2 and m_3 are considered as independent pieces of evidence. According to the Dempster's Rule, only two pieces of evidence can be combined at the same time. However, referring to the commutativity of Dempster's rule Shafer [1976], the sequence of combinations does not make difference to the result. As the equations for trustworthiness of B (m_2 in Equation 2.21) and C (m_3 in Equation 2.22) have similar form, their combinations are performed first ($m_{23} = m_2 \oplus m_3$); then we combine the intermediate results with the appropriateness of sub-goals for complementary argument m_1 (Equation 2.16).

There are 9 possible focal sets from combining the masses m_2 and m_3 for trustworthiness B and C. They lead to 9 focal sets for mass m_{23}. For all the combinations, the conflict factor K (see footnote 2) calculated is 0. This intermediate combined results are presented in the Table 2.3. Because all the intermediate results are useful in the next step, all the masses are calculated and presented in this table.

Then, we combine the masses for the appropriateness of sub-goals (m_1 in Equation 2.16) with the intermediate masses (m_{23}). The focal sets of all possible combinations of m_1 and m_{23} are presented in Table 2.4. The combined masses are denoted with m_{1-3}. As some obtained subsets of several combinations are the same, they contribute to the same new focal elements. The masses of these subsets shall be added up according to Dempster's Rule ($K = 0$). For instance, the mass of the focal set $\{B, C, A\}$ is calculated as follows:

Table 2.3: Intermediate combination results (m_{23}) of trustworthiness of B and C

		$\boldsymbol{m_2\ (trust_B)}$		
		$\boldsymbol{m_2^\Omega(\{B\} \times \Omega_C \times \Omega_A) = bel_B}$	$\boldsymbol{m_2^\Omega(\{\overline{B}\} \times \Omega_C \times \Omega_A) = disb_B}$	$\boldsymbol{m_2^\Omega(\Omega) = 1 - bel_B - disb_B}$
$\boldsymbol{m_3\ (trust_C)}$	$\boldsymbol{m_3^\Omega(\Omega_B \times \{C\} \times \Omega_A) = bel_C}$	$m_{23}(\{B\} \times \{C\} \times \Omega_A) = bel_B \cdot bel_C$	$m_{23}(\{\overline{B}\} \times \{C\} \times \Omega_A) = disb_B \cdot bel_C$	$m_{23}(\Omega_B \times \{C\} \times \Omega_A) = (1 - bel_B - disb_B)bel_C$
	$\boldsymbol{m_3^\Omega(\Omega_B \times \{\overline{C}\} \times \Omega_A) = = disb_C}$	$m_{23}(\{B\} \times \{\overline{C}\} \times \Omega_A) = = bel_B \cdot disb_C$	$m_{23}(\{\overline{B}\} \times \{\overline{C}\} \times \Omega_A) = = disb_B \cdot disb_C$	$m_{23}(\Omega_B \times \{\overline{C}\} \times \Omega_A) = = (1 - bel_B - disb_B)disb_C$
	$\boldsymbol{m_3^\Omega(\Omega) = 1 - bel_C - disb_C}$	$m_{23}(\{B\} \times \Omega_C \times \Omega_A) = bel_B(1 - bel_C - disb_C)$	$m_{23}(\{\overline{B}\} \times \Omega_C \times \Omega_A) = disb_B(1 - bel_C - disb_C)$	$m_{23}(\Omega) = (1 - bel_B - disb_B)(1 - bel_C - disb_C)$

$$
\begin{aligned}
m^{\Omega}_{1-3}(\{B,C,A\}) &= m_1 \oplus m_{23} \\
&= m^{\Omega}_1(\{B\}\times\Omega_C\times\{A\}\cup\{\overline{B}\}\times\Omega_C\times\{\overline{A}\})\cdot m^{\Omega}_{23}(\{B\}\times\{C\}\times\Omega_A)+ \\
&\quad m^{\Omega}_1(\Omega_B\times\{C\}\times\{A\}\cup\Omega_B\times\{\overline{C}\}\times\{\overline{A}\})\cdot m^{\Omega}_{23}(\{B\}\times\{C\}\times\Omega_A)+ \\
&\quad m^{\Omega}_1(\{(\overline{B},\overline{C},\overline{A}),(\overline{B},C,\overline{A}),(B,\overline{C},\overline{A}),(B,C,A)\})\cdot m^{\Omega}_{23}(\{B\}\times\{C\}\times\Omega_A) \\
&= w_B\cdot v\cdot bel_B\cdot bel_C + w_C\cdot v\cdot bel_B\cdot bel_C + (1-w_B-w_C)\cdot v\cdot bel_B\cdot bel_C \\
&= bel_B\cdot bel_C\cdot v
\end{aligned}
\tag{2.23}
$$

The belief in A (bel_A) is deduced by adding up all the masses of the focal sets (underlined in Table 2.4) that contribute to the mass $m^{\Omega\downarrow\Omega_A}(\{A\})$ after the marginalization operation:

$$
\begin{aligned}
bel_A &= bel^{\Omega_A}(\{A\}) = m^{\Omega\downarrow\Omega_A}(\{A\}) = \sum_{Q\subset\Omega, Q\downarrow\Omega_A=\{A\}} m^{\Omega}(Q) \\
&= m^{\Omega}_{1-3}(\{\overline{B},C,A\}) + m^{\Omega}_{1-3}(\{B,\overline{C},A\}) + m^{\Omega}_{1-3}(\{B,C,A\}) + \\
&\quad m^{\Omega}_{1-3}(\{(B,\overline{C},A),(B,C,A)\}) + m^{\Omega}_{1-3}(\{(\overline{B},C,A),(B,C,A)\}) \\
&= [bel_B\cdot w_b + bel_C\cdot w_C + bel_B\cdot bel_C(1-w_B-w_c)]v
\end{aligned}
\tag{2.24}
$$

Similarly, the disbelief ($disb_A$) and uncertainty ($uncer_A$) in A are calculated as follows:

$$
\begin{aligned}
disb_A &= bel^{\Omega_A}(\{\overline{A}\}) = m^{\Omega\downarrow\Omega_A}(\{\overline{A}\}) = \sum_{Q\subset\Omega, Q\downarrow\Omega_A=\{\overline{A}\}} m^{\Omega}(Q) \\
&= m^{\Omega}_{1-3}(\{\overline{B},\overline{C},\overline{A}\}) + m^{\Omega}_{1-3}(\{\overline{B},C,\overline{A}\}) + m^{\Omega}_{1-3}(\{B,\overline{C},\overline{A}\}) + \\
&\quad m^{\Omega}_{1-3}(\{(\overline{B},\overline{C},\overline{A}),(\overline{B},C,\overline{A})\}) + m^{\Omega}_{1-3}(\{(\overline{B},\overline{C},\overline{A}),(B,\overline{C},\overline{A})\}) \\
&= [disb_B(1-w_C) + disb_C(1-w_B) - disb_B\cdot disb_C(1-w_B-w_C)]v \\
&= \{disb_B\cdot w_B + disb_C\cdot w_C + [1-(1-disb_B)(1-disb_C)](1-w_B-w_C)\}v
\end{aligned}
\tag{2.25}
$$

Table 2.4: Combination results (m_{1-3}) of confidence assessment parameters

		m_1 ($appr_{\{B,C\}\to A}$)		
		$m_1^{\Omega}(\{B\}\times\Omega_C\times\{A\}\cup \overline{B}\}\times\Omega_C\times\{\overline{A}\})$	$m_1^{\Omega}(\Omega_B\times\{C\}\times\{A\}\cup \Omega_B\times\{\overline{C}\}\times\{\overline{A}\})$	$m_1^{\Omega}(\{(\overline{B},\overline{C},\overline{A}),(\overline{B},C,\overline{A}),(B,\overline{C},\overline{A}),(B,C,A)\})$
m_{23} ($trust_B \oplus trust_C$)	$m_{23}^{\Omega}(\{B\}\times\{C\}\times\Omega_A)$	$\{\underline{B,C,A}\}$	$\{\underline{B,C,A}\}$	$\{\underline{B,C,A}\}$
	$m_{23}^{\Omega}(\{\overline{B}\}\times\{C\}\times\Omega_A)$	$\{\overline{B},C,\overline{A}\}$	$\{\underline{\overline{B},C,A}\}$	$\{\overline{B},C,\overline{A}\}$
	$m_{23}^{\Omega}(\{B\}\times\{\overline{C}\}\times\Omega_A)$	$\{\underline{B,\overline{C},A}\}$	$\{B,\overline{C},\overline{A}\}$	$\{B,\overline{C},\overline{A}\}$
	$m_{23}^{\Omega}(\{\overline{B}\}\times\{\overline{C}\}\times\Omega_A)$	$\{\overline{B},\overline{C},\overline{A}\}$	$\{\overline{B},\overline{C},\overline{A}\}$	$\{\overline{B},\overline{C},\overline{A}\}$
	$m_{23}^{\Omega}(\{\overline{B}\}\times\Omega_C\times\Omega_A)$	$\{(\overline{B},\overline{C},\overline{A}),(\overline{B},C,\overline{A})\}$	$\{(\overline{B},\overline{C},\overline{A}),(\overline{B},C,A)\}$	$\{(\overline{B},\overline{C},\overline{A}),(\overline{B},C,\overline{A})\}$
	$m_{23}^{\Omega}(\{B\}\times\Omega_C\times\Omega_A)$	$\{\underline{(B,\overline{C},A),(B,C,A)}\}$	$\{(B,\overline{C},\overline{A}),(B,C,A)\}$	$\{(B,\overline{C},\overline{A}),(B,C,A)\}$
	$m_{23}^{\Omega}(\Omega_B\times\{\overline{C}\}\times\Omega_A)$	$\{(\overline{B},\overline{C},\overline{A}),(B,\overline{C},A)\}$	$\{(\overline{B},\overline{C},\overline{A}),(B,\overline{C},\overline{A})\}$	$\{(\overline{B},\overline{C},\overline{A}),(B,\overline{C},\overline{A})\}$
	$m_{23}^{\Omega}(\Omega_B\times\{C\}\times\Omega_A)$	$\{(\overline{B},C,\overline{A}),(B,C,A)\}$	$\{\underline{(\overline{B},C,A),(B,C,A)}\}$	$\{(\overline{B},C,\overline{A}),(B,C,A)\}$
	$m_{23}^{\Omega}(\Omega)$	$\{B\}\times\Omega_C\times\{A\}\cup \overline{B}\}\times\Omega_C\times\{\overline{A}\}$	$\Omega_B\times\{C\}\times\{A\}\cup \Omega_B\times\{\overline{C}\}\times\{\overline{A}\}$	$\{(\overline{B},\overline{C},\overline{A}),(\overline{B},C,\overline{A}),(B,\overline{C},\overline{A}),(B,C,A)\}$

Table 2.5: Aggregation rules for complementary arguments

Types	Aggregation rules
C-Arg	$\begin{cases} bel_A & = [bel_B \cdot w_B + bel_C \cdot w_C + bel_B \cdot bel_C(1 - w_B - w_C)]v \\ disb_A & = \{disb_B \cdot w_B + disb_C \cdot w_C + [1 - (1 - disb_B)(1 - disb_C)](1 - w_B - w_C)\}v \\ uncer_A & = 1 - bel_A - disb_A \end{cases}$

$$
\begin{aligned}
uncer_A =& m^{\Omega_A}(\{A, \overline{A}\}) = \sum_{Q \subset \Omega, Q \downarrow \Omega_A = \{A, \overline{A}\}} m^{\Omega}(Q) \\
=& 1 - [bel_B \cdot w_b + bel_C \cdot w_C + bel_B \cdot bel_C(1 - w_B - w_c)]v - \\
& [disb_B(1 - w_C) + disb_C(1 - w_B) - disb_B \cdot disb_C(1 - w_B - w_C)]v \\
=& 1 - bel^{\Omega_A}(\{A\}) - bel^{\Omega_A}(\{\overline{A}\})
\end{aligned}
\tag{2.26}
$$

The confidence aggregation rules for the complementary argument with two sub-goals are developed (presented in Equation 2.24, 2.25 and 2.26). The *trustworthiness* of top goal A (bel_A, $disb_A$, $uncer_A$) can be deduced based on these aggregation rules. We summarise them in the Table 2.5.

2.4.5 Confidence aggregation for redundant arguments

Due to various ways of supporting top goal by the sub-goals, the confidence propagation in redundant arguments is different from the complementary arguments. However, the approach to calculate the confidence aggregation rules shares the same procedure of previous section. Now, assume that the double-node argument in Figure 2.6 is a redundant argument. The *trustworthiness* of top goal A ($bel_A, disb_A, uncer_A$) is calculated based on the combination of the *appropriateness* of sub-goals to A (m_1 in Equation 2.17) and the *trustworthiness* of sub-goals (m_2 in Equation 2.21 and m_3 in Equation 2.22). Since the masses m_1 and m_2 are combined, we can use the results shown in Table 2.3.

Then, the calculation process is similar to the one used for the complementary argument presented in the previous section. We directly give the confidence aggregation rules for the redundant argument in Table 2.6.

Table 2.6: Aggregation rules for redundant arguments

Types	Aggregation rules
R-Arg	$\begin{cases} bel_A = \{bel_B \cdot w_B + bel_C \cdot w_C + [1-(1-bel_B)(1-bel_C)](1-w_B-w_C)\}v \\ disb_A = [disb_B \cdot w_B + disb_C \cdot w_C + disb_B \cdot disb_C(1-w_B-w_C)]v \\ uncer_A = 1 - bel_A - disb_A \end{cases}$

2.4.6 Aggregation rules for particular argument types

In Section 2.4.2, three particular argument types are introduced along with the two basic argument types. They are, respectively, *fully complementary argument (FC-Arg)*, *fully redundant argument (FR-Arg)* and *Disparate argument (D-Arg)*. As mentioned in that section, these three argument types are the particular cases when the weights of sub-goals equal to some limit values. In this subsection, we determine the confidence aggregation rules based on the rules for complementary and redundant arguments.

- *Fully complementary argument (FC-Arg)*:

  For the fully complementary argument, $w_{B\times C\to A} = 1$, i.e. $w_B = w_C = 0$. The trustworthiness of A, $trust_A = (bel_A, disb_A, uncer_A)$ can be calculated with the formula:

$$trust_A : \begin{cases} bel_A = bel_B \cdot bel_C \cdot v \\ disb_A = [1-(1-disb_B)(1-disb_C)]v \\ uncer_A = 1 - bel_A - disb_A \end{cases} \tag{2.27}$$

  In this case, the way that the sub-goals B and C contribute to the belief in goal A turns to a *pure AND*. In contrast, the disbelief propagates in a manner of OR logic gate from sub-goals to top goal. These characteristics are, in turn, compliant with the initial definition based on AND logic gate.

- *Fully redundant argument (FR-Arg)*:

  For the fully redundant argument (FR-Arg), $w_{B+C\to A} = 1$, i.e. $w_B = w_C = 0$. The trustworthiness of A, $trust_A = (bel_A, disb_A, uncer_A)$ can be calculated with the formula:

$$trust_A : \begin{cases} bel_A = [1 - (1 - bel_B)(1 - bel_C)]v \\ disb_A = disb_B \cdot disb_C \cdot v \\ uncer_A = 1 - bel_A - disb_A \end{cases} \tag{2.28}$$

In this case, the way that the sub-goals B and C contribute to the belief in goal A turns to a *pure OR*. In contrast, the disbelief propagates in a manner of AND logic gate from sub-goals to top goal. These characteristics are also compliant with the initial definition based on OR logic gate.

- *Disparate argument (D-Arg)*:

  For both complementary and redundant arguments, if the $w_{B\times C\to A}$ and $w_{B+C\to A}$ decrease (i.e. w_B and w_C increase) to $w_{B\times C\to A} = 0$ and $w_{B+C\to A} = 0$ (i.e. $w_B + w_C = 1$), the aggregation rules of the complementary and redundant arguments become the same formula:

$$trust_A : \begin{cases} bel_A = (bel_B w_B + bel_C w_C)v \\ disb_A = (disb_B w_B + disb_C w_C)v \\ uncer_A = 1 - bel_A - disb_A \end{cases} \tag{2.29}$$

  In this case, B and C contribute independently to the top goal A with their own weights. The confidence aggregation rules are the weighted sum of the trustworthiness of the sub-goals.

2.5 N-node argument

It is common for an argument to have more than two premises. The confidence aggregation rules for double-node (an argument with 2 premises) have been introduced in Section 2.4. The development process is relatively complex due to the twice combination of the masses for assessment parameters. With the growth of the premise number, the required calculation will increase exponentially. Specifically, this calculation includes the combination of masses and the expression simplification for the non-linear polynomials. Regarding the former, for an argument with 2 premises to n premises, the number of possible combination is shown in Table 2.7. Thus, it would better to have the general confidence aggregation rules for n-node argument of both types of arguments.

Before the development of the aggregation rules, a requirement for the n-node argument structure has to be stated first.

Table 2.7: Number of possible combinations to develop aggregation rules

N	$\#C_1$	$\#C_2$	$\#C_{total}$
2	9	36	45
3	27	135	162
		...	
n	3^n	$3^n(n+2)$	$3^n(n+3)$

C_1: combination of the trustworthiness of sub-goals, C_2: combination of the appropriateness of sub-goals to top goal with the results of C_1, C_{total}: total combinations

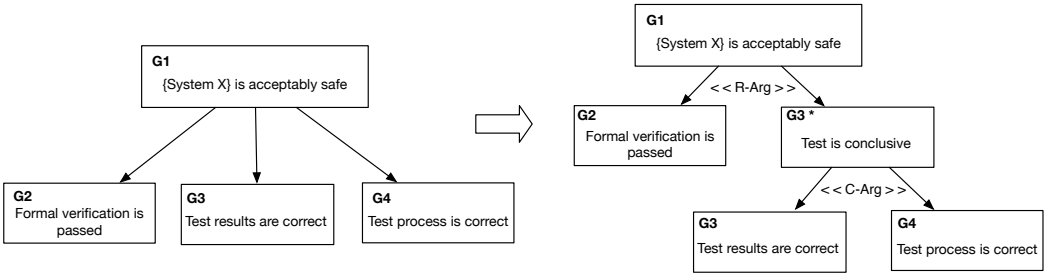


Figure 2.8: Re-structuring an argument for confidence propagation

2.5.1 Re-structuring n-node argument

In order to employ the same approach to develop the aggregation rules and to avoid, to the maximum extent, introducing new uncertainties, we require that every branch in the n-node argument shall belong to only one argument type. No complementary premises and redundant premises are mixed to support one goal. On the contrary, the argument needs to be modified. For example, in Figure 2.8, the top goal *"G1: System is acceptably safe"* is supported by three sub-goals. They are, respectively, *"G2: Formal verification is passed"*, *"G3: Test results are correct"* and *"G4: Test process is correct"*. The formal verification and test are two different techniques to validate and verify, for instance, the compliance of system safety requirements. The evidence produced through these two techniques may have some degree of redundancy (R-Arg). However, G3 and G4 are the premises related to the test and are typically complementary (C-Arg). In this case, these premises shall be regrouped and new intermediate goals need to be proposed (G3* in the right figure of Figure 2.8).

2.5.2 Confidence aggregation for n-node arguments

We propose to use an inductive approach to deduce the confidence aggregation rules. The three-step process for single and double-node is performed again for an argument with three sub-goals: B, C and D. The frame of discernment is $\Omega' = \Omega_B \times \Omega_C \times \Omega_D \times \Omega_A$. The appropriateness and trustworthiness of sub-goals are given in the Equation 2.30-2.34.

The appropriateness of sub-goals for the three-node *complementary argument* is:

$$\begin{cases} m_1^{\Omega'}(\{B\} \times \Omega_C \times \Omega_D \times \{A\} \cup \{\overline{B}\} \times \Omega_C \times \Omega_D \times \{\overline{A}\}) = w_B \cdot v \\ m_1^{\Omega'}(\Omega_B \times \{C\} \times \Omega_D \times \{A\} \cup \Omega_B \times \{\overline{C}\} \times \Omega_D \times \{\overline{A}\}) = w_C \cdot v \\ m_1^{\Omega'}(\Omega_B \times \Omega_C \times \{D\} \times \{A\} \cup \Omega_B \times \Omega_C \times \{\overline{D}\} \times \{\overline{A}\}) = w_D \cdot v \\ m_1^{\Omega'}(\{(\overline{B},\overline{C},\overline{D},\overline{A}),(\overline{B},C,\overline{D},\overline{A}),(B,\overline{C},\overline{D},\overline{A}),(\overline{B},\overline{C},D,\overline{A}), \\ \quad (B,C,\overline{D},\overline{A}),(B,\overline{C},D,\overline{A}),(\overline{B},C,D,\overline{A}),(B,C,D,A)\}) = w_{B\times C\times D\to A} \cdot v \\ m_1^{\Omega'}(\Omega) = 1 - v \end{cases} \tag{2.30}$$

where $v, w_B, w_C, w_D \in [0,1]$, and $w_{B\times C\times D\to A} = 1 - w_B - w_C - w_D \geq 0$.

The appropriateness of sub-goals for the three-node *redundant argument* is:

$$\begin{cases} m_1^{\Omega'}(\{B\} \times \Omega_C \times \Omega_D \times \{A\} \cup \{\overline{B}\} \times \Omega_C \times \Omega_D \times \{\overline{A}\}) = w_B \cdot v \\ m_1^{\Omega'}(\Omega_B \times \{C\} \times \Omega_D \times \{A\} \cup \Omega_B \times \{\overline{C}\} \times \Omega_D \times \{\overline{A}\}) = w_C \cdot v \\ m_1^{\Omega'}(\Omega_B \times \Omega_C \times \{D\} \times \{A\} \cup \Omega_B \times \Omega_C \times \{\overline{D}\} \times \{\overline{A}\}) = w_D \cdot v \\ m_1^{\Omega'}(\{(\overline{B},\overline{C},\overline{D},\overline{A}),(\overline{B},C,\overline{D},A),(B,\overline{C},\overline{D},A),(\overline{B},\overline{C},D,A), \\ \quad (B,C,\overline{D},A),(B,\overline{C},D,A),(\overline{B},C,D,A),(B,C,D,A)\}) = w_{B+C+D\to A} \cdot v \\ m_1^{\Omega'}(\Omega) = 1 - v \end{cases} \tag{2.31}$$

where $v, w_B, w_C, w_D \in [0,1]$, and $w_{B+C+D\to A} = 1 - w_B - w_C - w_D \geq 0$.

The trustworthiness of sub-goals for the three-node argument is:

$$\begin{cases} bel^{\Omega_B}(\{B\}) = m_2^{\Omega_B\uparrow\Omega'}(\{B\} \times \Omega_C \times \Omega_D \times \Omega_A) = bel_B \\ bel^{\Omega_B}(\{\overline{B}\}) = m_2^{\Omega_B\uparrow\Omega'}(\{\overline{B}\} \times \Omega_C \times \Omega_D \times \Omega_A) = disb_B \\ m^{\Omega_B}(\{B,\overline{B}\}) = m_2^{\Omega_B\uparrow\Omega'}(\Omega') = uncer_B = 1 - bel_B - disb_B \end{cases} \tag{2.32}$$

$$\begin{cases} bel^{\Omega_C}(\{C\}) = m_3^{\Omega_C \uparrow \Omega'}(\Omega_B \times \{C\} \times \Omega_D \times \Omega_A) = bel_C \\ bel^{\Omega_C}(\{\overline{C}\}) = m_3^{\Omega_C \uparrow \Omega'}(\Omega_B \times \{\overline{C}\} \times \Omega_D \times \Omega_A) = disb_C \\ m^{\Omega_C}(\{C, \overline{C}\}) = m_3^{\Omega_C \uparrow \Omega'}(\Omega') = uncer_C = 1 - bel_C - disb_C \end{cases} \tag{2.33}$$

$$\begin{cases} bel^{\Omega_D}(\{D\}) = m_4^{\Omega_D \uparrow \Omega'}(\Omega_B \times \Omega_C \times \{D\} \times \Omega_A) = bel_D \\ bel^{\Omega_D}(\{\overline{D}\}) = m_4^{\Omega_D \uparrow \Omega'}(\Omega_B \times \Omega_C \times \{\overline{D}\} \times \Omega_A) = disb_D \\ m^{\Omega_D}(\{D, \overline{D}\}) = m_4^{\Omega_D \uparrow \Omega'}(\Omega') = uncer_D = 1 - bel_D - disb_D \end{cases} \tag{2.34}$$

Where $bel_B, disb_B, bel_B + disb_B, bel_C, disb_C, bel_C + disb_C, bel_D, disb_D, bel_D + disb_D \in [0, 1]$.

Since the formula development is the same, the calculation process will not be detailed again. The confidence aggregation rules for three-node arguments is directly provided in Table 2.8.

Table 2.8: Aggregation rules for three-node arguments supporting A

Types	Aggregation rules
C-Arg	$\begin{cases} bel_A = & [bel_B \cdot w_B + bel_C \cdot w_C + bel_D \cdot w_D \\ & + bel_B \cdot bel_C \cdot bel_D(1 - w_B - w_C - w_D)]v \\ disb_A = & \{disb_B \cdot w_B + disb_C \cdot w_C + disb_D \cdot w_D \\ & + [1 - (1 - disb_B)(1 - disb_C)(1 - disb_D)](1 - w_B - w_C - w_D)\}v \\ uncer_A = & 1 - bel_A - disb_A \end{cases}$
R-Arg	$\begin{cases} bel_A = & \{bel_B \cdot w_B + bel_C \cdot w_C + bel_D \cdot w_D \\ & + [1 - (1 - bel_B)(1 - bel_C)(1 - bel_D)](1 - w_B - w_C - w_D)\}v \\ disb_A = & [disb_B \cdot w_B + disb_C \cdot w_C + disb_D \cdot w_D \\ & + disb_B \cdot disb_C \cdot disb_D(1 - w_B - w_C - w_D)]v \\ uncer_A = & 1 - bel_A - disb_A \end{cases}$

We find the regular pattern of the aggregation rules for double-node and three-node argument. The aggregation rules for any n-node argument (for n>1) are supposed to be developed based on the same approach. Thus, we prove that the general confidence aggregation rules for n-node complementary arguments and redundant arguments are the formulas shown in Table 2.9.

Table 2.9: Aggregation rules for n-node arguments supporting A

Types	Aggregation rules
C-Arg	$\begin{cases} bel_A = & [\sum_{i=1}^{n} bel_i w_i + (1 - \sum_{i=1}^{n} w_i) \prod_{i=1}^{n} bel_i]v \\ disb_A = & \{\sum_{i=1}^{n} disb_i w_i + (1 - \sum_{i=1}^{n} w_i)[1 - \prod_{i=1}^{n} (1 - disb_i)]\}v \\ uncer_A = & 1 - bel_A - disb_A \end{cases}$
R-Arg	$\begin{cases} bel_A = & \{\sum_{i=1}^{n} bel_i w_i + (1 - \sum_{i=1}^{n} w_i)[1 - \prod_{i=1}^{n} (1 - bel_i)]\}v \\ disb_A = & [\sum_{i=1}^{n} disb_i w_i + (1 - \sum_{i=1}^{n} w_i) \prod_{i=1}^{n} disb_i]v \\ uncer_A = & 1 - bel_A - disb_A \end{cases}$

Where $n > 1$, $bel_i, disb_i, w_i, v \in [0, 1]$, and $\sum_{i=1}^{n} w_i \leq 1$

2.6 Sensitivity analysis of confidence aggregation rules

In this section, we propose to carry out the sensitivity analysis to discover the behaviours of the confidence aggregation rules. These behaviours are to be analysed to judge whether they are in line with the rationale about the corresponding argument types and to validate the propagation operators.

2.6.1 Sensitivity analysis with Tornado graph

We suggest performing a sensitivity analysis using a tornado graph. It is a simple statistical tool, which shows the positive or negative influence of basic elements on main function. Considering a function $f(x_1, ...x_n)$, where values $X_1, ..., X_n$ of the variables x_i have been estimated, the tornado analysis consists in the estimation (for each $x_i \in [X_{min}, X_{max}]$) of the values $f(X_1, ..., X_{i-1}, X_{min}, X_{i+1}, ...X_n)$ and $f(X_1, ..., X_{i-1}, X_{max}, X_{i+1}, ...X_n)$, where X_{min} and X_{max} are the maximum and minimum admissible values of variables x_i. Hence for each x_i, we get an interval of possible variations of function f. The tornado graph is a visual presentation with ordered intervals. In our case, we estimate the confidence in A, $m(A)$, with corresponding intervals for v, bel_B, bel_C, $disb_B$, $disb_C$, w_B and w_C.

We take the example of the double-node argument to analyse the confidence aggregation rules for both complementary (see Table 2.5) and redundant (see Table 2.6) arguments. The basic values (X_i) and intervals $[X_{min}, X_{max}]$ for each parameter are shown in Table 4.7. The basic values (X_i) are given arbitrarily and the intervals $[X_{min}, X_{max}]$ are deduced, according to the requirements for the pa-

rameters in the formulas: $bel_i, disb_i, w_i, v \in [0,1]$, and $\sum_{i=1}^{n} w_i \leq 1$. For instance, the interval for w_B is [0,0.9], because $w_C = 0.1$ and the sum of them should not be more than 1.

Table 2.10: Values and intervals chosen for the sensitivity analysis

	v	bel_B	bel_C	$disb_B$	$disb_C$	w_B	w_C
Basic value X_i	0.9	0.5	0.8	0.2	0.1	0.4	0.1
$[X_{min}, X_{max}]$	[0,1]	[0,0.8]	[0,0.9]	[0,0.5]	[0,0.2]	[0,0.9]	[0,0.6]

With the basic values above, the trustworthiness of A are $(bel_A, disb_A, uncer_A) = (0.432, 0.207, 0.361)$ for complementary argument and $(bel_A, disb_A, uncer_A) = (0.657, 0.09, 0.253)$ for redundant argument. These values are set as the positions of vertical axis in corresponding tornado graphs. To determine the sensitivity to bel_B, we keep the basic values for all other variables and only calculate the values bel_A for $bel_B = 0$ and $bel_B = 0.8$: we obtain the values of the confidence in A [0.072, 0.648] for complementary argument and [0.432, 0.792] for redundant argument. The same approach is applied for other parameters. The analysis results are presented in Figure 2.9.

2.6.2 Result analysis

All graphs show that v is the most influencing factor. When $v = 0$, the belief in A is 0. Observing the structure of the aggregation formulas, this parameter v remains as the common factor of the formulas after multiple combinations. This is in compliance with the original idea of using a discounting factor. Thus, v is the most sensitive point for these formulas. In terms of interpretation, v is used to measure the overall reliability of the sources or the completeness of the premises. Proposing this factor aims to provide a possibility to criticise all sub-goals as a whole. Generally, we assume that $v = 1$, which means that full confidence in sub-goals will lead to full confidence in the top goal of an argument. In the inverse case ($v \neq 1$), it should be very cautious to determine the value of v. In next chapter, we will discuss some possible situations in an safety argument to lower this value.

The trustworthiness of B has more impact on the trustworthiness of A than that of C in all six graphs. This is consistent to the higher weight of B than C. Comparing the impacts of B and C for two types of arguments, the impact difference between B and C for complementary argument is greater than redundant argument. It signifies that the confidence in the top goal of a complementary argument relies more on the confidence of each sub-goals. Only to increase both of the $trust_B$ and

Complementary argument

Redundant argument

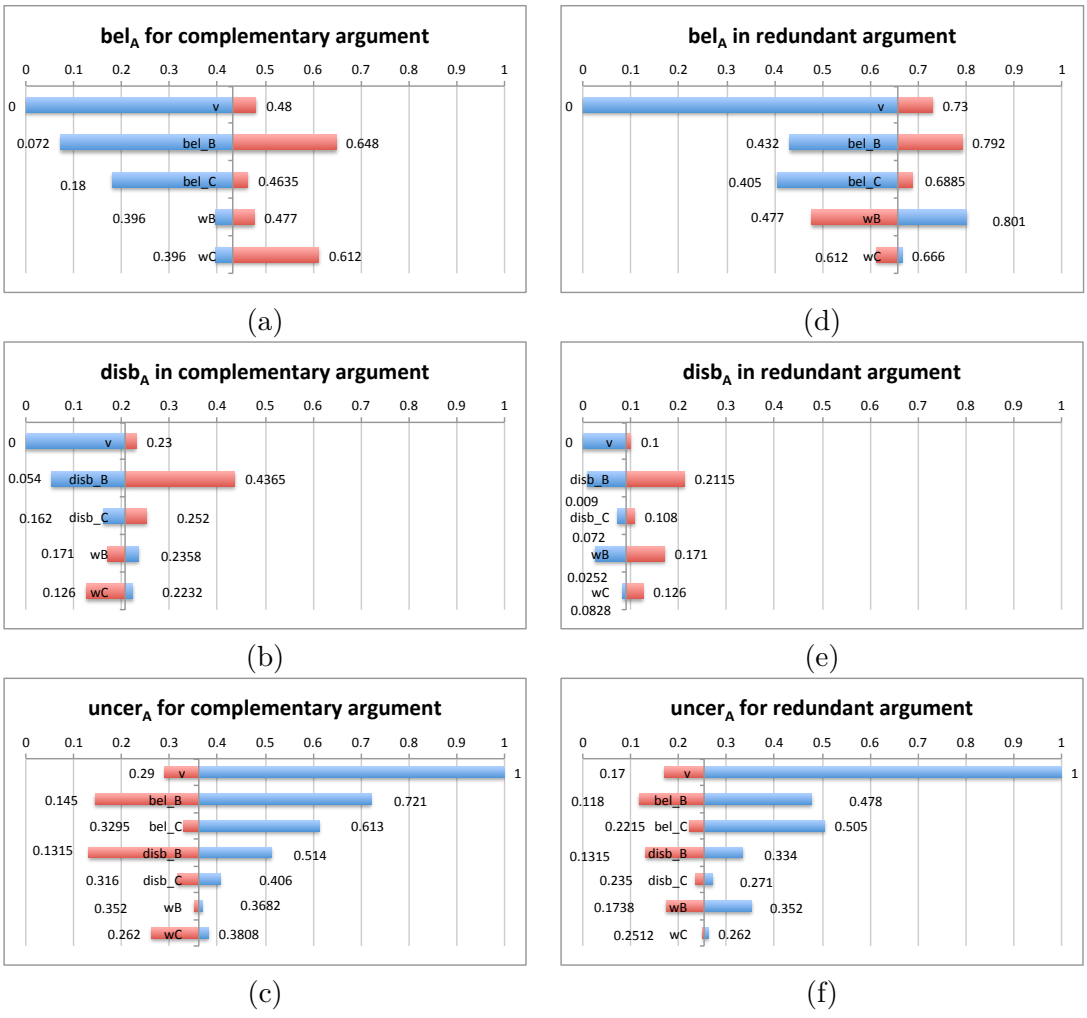


Figure 2.9: Tornado graphs for two types of double-node arguments

$trust_C$ can effectively maximize the $trust_A$.

Furthermore, an interesting consequence for redundant argument is that when the weight w_B increases, the belief in A decreases (see Figure 2.9 (d)). When $w_B = 0.9$, then $bel_A = 0.432$. This is due to the constraint that $w_{B+C\to A} = 1 - w_B - w_C$, that is, increasing w_B leads to less redundancy. Therefore, the belief in A declines. It implies that, for redundant arguments, increasing the redundancy of B and C (i.e. decreasing w_B and w_C) will contribute to higher confidence in A. This result shows that the right interpretation of the weight w_B or w_C is not only the impact of $trust_B$ or $trust_C$ on $trust_A$, but also a representation of the degree of redundancy (or complementarity for the complement argument).

2.6.3 Analysis conclusion

According to this sensitivity analysis, the behaviours of the aggregation rules are consistent with our expectation regarding the influence of each parameters on the trustworthiness of top goal ($trust_A = (bel_A, uncer_A, disb_A)$). The different impacts of the appropriateness of sub-goals on $trust_A$ intuitively distinguish ways of the trustworthiness propagation between the complementary and redundant arguments. Generally, the complementary argument is more sensitive to the variation of the assessment measures. When taking the same values of all the measures, the *belief* (bel_A) in the top goal of a complementary argument is lower than a redundant argument, whereas the *disbelief* ($disb_A$) and *uncertainty* ($uncer_A$) are always higher than a redundant argument.

A particular difference of the impacts of w_B and w_C shown in graphs (a) and (d) is discovered. It indicates that the variation of weights of sub-goals can also strengthen or weaken an *"AND gate"* or an *"OR gate"* of an argument (the complementarity or redundancy of sub-goals). This is actually a reminder of the original idea of defining the mixed propagation operators: *B alone, C alone and pure AND/OR.*

2.7 Conclusion

In this chapter, we propose a confidence propagation model for safety argument of different inference types. In fact, we put forward a systematic approach based on D-S theory to develop the confidence propagation model and generalise it for n-node arguments. In the mean time, a preliminary validation is carried out with the sensitivity analysis. Firstly, we identify the influencing factors of the confidence in an ar-

gument; then they are formally defined as *trustworthiness:* $trust = (bel, uncer, disb)$ and *appropriateness:* $appr = (w_i, w_{TYPE}, v)$. Then, these definitions are further specified according to different argument structures (simple argument and multi-node argument) and different inference types (complementary and redundant). Corresponding confidence aggregation rules are developed; and they are finally generalised into the aggregations rules of n-node arguments. However, applying this model requires the values of considerable parameters (e.g. 10 variables to be determined for a 2-node argument). In addition, they seem not obvious to be obtained. Thus, we will deal with this issue of feasibility of our proposed approach in next chapter.

Chapter 3

Confidence assessment framework for Safety Arguments

Contents

3.1 Introduction

The definition and propagation of the confidence in arguments were studied in previous chapter. This work realises the formalisation of assessment measures of the safety arguments, which provides a mathematical model for the uncertain information fusion. However, it is still an incomplete approach for the practical application to assess the confidence in a safety case. In order to make this approach feasible, we

have to fill the gaps between the theoretical models and expert judgement. Moreover, many parameters of the theoretical model have to be determined. Therefore, an integrated confidence assessment framework, which can bridge these gaps, is needed. Most of this work is published in the publication [Wang et al., 2017a].

In the previous chapter, two measures are proposed in the theoretical model: the *trustworthiness* and the *appropriateness* of premises in arguments. Thus, the issue regarding to the determination of the values of these measures emerges. The safety assessors or engineers (called experts below) are supposed to provide their opinions on the arguments as the raw data for these measures. However, it is not obvious for the experts to evaluate directly the degree of *"belief"*, *"disbelief"* and *"uncertainty"* in one statement (*trustworthiness*), nor the contributing weight of certain evidence (*appropriateness*). For the *trustworthiness*, we adopt a practical method to extract the expert decision and confidence in this decision. Then, these judgements are transferred to the trustworthiness $(bel, disb, uncer)$, which are the notations used in the theoretical model. For the *appropriateness*, we propose a method to reuse the framework itself to derive the corresponding parameters based on the collected expert judgements. Moreover, some supplementary considerations for the evaluation of the other elements (contexts, justifications and assumptions) in a safety case are given.

Based on these solutions, in this chapter, we propose an integrated framework realising the quantitative assessment of the confidence in safety arguments. A sensitivity analysis is followed to show the behaviours of the updated assessment model.

3.2 Framework Overview

The proposed confidence assessment framework of the safety argument is illustrated in Figure 3.1. It aims to evaluate the confidence in the top goal and help the final decision-making for the acceptance of the corresponding system. The whole assessment process involves the following four steps:

1) Building the structured safety case for the system under consideration.

2) Estimating the parameters affecting the confidence propagated upwards in the argumentation

3) Assessing confidence in sub-goals based on available safety evidence.

4) Aggregating confidence and making decision on the claim of the *Top Goal*.

An overview of these steps is given below, and they are defined in Section 3.3.

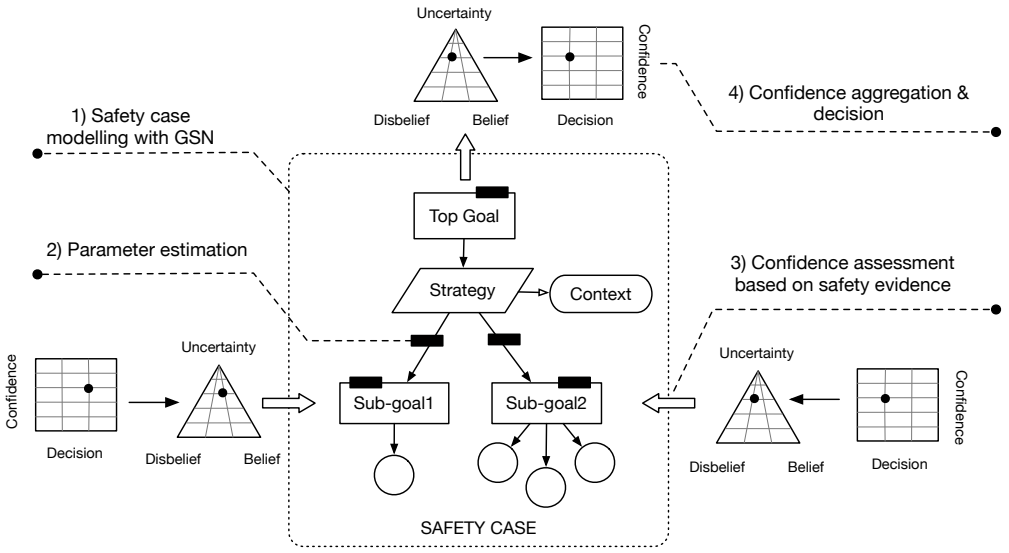


Figure 3.1: Overview of confidence assessment framework for the safety argument

3.2.1 Building the structured safety case

How to develop a structured safety argument is not in the scope of this thesis. As discussed in the Chapter 2, the goals structuring notation is a comprehensive extension of Toulmin's notation; and this goal-based notation is suitable for large scale safety argument. In this step, the development of GSN safety argument shall refer to the guideline of GSN specifications GSN Standard [2011] and related domain-specific standards. In the GSN specification, both a top-down method for building new safety case and a bottom-up method for working from available evidence are provided with specific steps. It also indicates how to avoid the common errors in terms of language and structural issues while constructing goal structures. It is essential to review the argument model to verify the completeness of evidence in accordance to standards and the hierarchical structure of the safety argument. The reviewing process is suggested to follow a 4-step approach. Modifications of the safety arguments should be implemented when necessary.

3.2.2 Estimating the parameters

Based on a complete GSN safety case, the next step of the assessment is to estimate the parameters of the *appropriateness*, which affect the confidence propagated upwards in the argumentation. The formal expression of the *appropriateness* is recalled:

$$appr_A = (w_{G1}, w_{G2}, ..., w_{Gn}, w_{TYPE}, v) \tag{3.1}$$

where w_{G1-GN} are the *contributing weights* of sub-goals; w_{TYPE} is the *degree of complementarity/redundancy* among sub-goals depending on argument types; v is the discounting factor. The estimation of these parameters requires data from experts.

After this step, we will obtain a parametrised safety argument model, which is ready to undergo the quantitative confidence assessment. Once this generic model is built for a given system, it is also reusable for similar systems with same safety level.

3.2.3 Assessing confidence in sub-goals

The assessment process follows a bottom-up approach as shown in Figure 3.1. The confidence assessment starts from the lowest level of sub-goals based on associated safety evidence. A scaled evaluation matrix is utilised to extract the experts' judgement of a sub-goal based on the supporting evidence. It is presented in Figure 3.1 with the two axes marked with *"decision"* and *"confidence"* and detailed in Figure 3.2 b). This judgement is assessed based on these two values: the *Decision* (*dec*) on the statement of the goal and the *Confidence (conf)* in this decision. By anology, it could be compared with the review process in most of the confidences; where reviewers are asked to decide if the paper is accected or rejected, and the associated confidence in their decision. They are then transformed into the 3-tuple *(bel,uncer,disb)* of the trustworthiness in this goal, as shown with the scaled triangles in Figure 3.1 marked with *"belief"*, *"uncertainty"* and *"disbelief"*. The 3-tuple is presented in this three-dimension coordinate system named Jøsang triangle Jøsang [2001]. This transformation from the experts' judgement to *(bel,uncer,disb)* refers to a related work Cyra and Gorski [2011]. With the help of this transformation to the 3-tuple *(bel,uncer,disb)*, the expert judgements can then be propagated upwards in the structured safety argument.

3.2.4 Confidence aggregation and decision-making

The last step of the assessment is to aggregate the confidence and make decision on the claim of the *Top Goal*. Confidence aggregation is realized through the combination of the 3-tuple *(bel,uncer,disb)* of lower-level goals to obtain the *(bel,uncer,disb)* of the higher-level goals. This step is based on the proposed confidence assessment approach derived from the Dempster-Shafer theory (see Chapter 3). As shown

in Figure 3.1, the trustworthiness *(bel,uncer,disb)* of *Sub-goal1* and *Sub-goal2* are aggregated to produce the *(bel,uncer,disb)* of the *Top Goal*. This aggregation requires parameter values of the argumentation estimated in Step 2). Finally, the decision is derived from the inverse transformation of the experts' judgement and *(bel,uncer,disb)*. This aims to generate the final judgement on the *Top Goal*, i.e. the *decision* and the *confidence in this decision (dec, conf)* of the *Top Goal*.

3.3 Framework implementation

In this section, we introduce an approach of the expert judgement extraction for argument assessment (Section 3.3.1). This approach is then integrated with the confidence assessment methodology proposed in the previous chapter (Section 3.3.2). With the help of this judgement extraction approach, the confidence assessment framework for safety argument becomes complete. In Section 3.3.3, a simple example of the judgement estimation and propagation is presented to show the calculation process of this proposed framework. We continue the sensitivity analysis introduced in the previous chapter for the confidence assessment framework (Section 3.3.4).

3.3.1 Judgement extraction approach

While assessing an argument, an expert has to evaluate all the elements of this argument, i.e. goals, evidence, contexts, etc. In Figure 3.2 a), a goal G1: *"Low-level requirements coverage is achieved"* has to be assessed. It is supported by the evidence S1: *"Low-level requirement coverage verification reports"*, which records the coverage verification of low-level requirements. We adopt an evaluation matrix as proposed by Cyra and Gorski [2011] to assess G1 with two criteria: the *decision* on the goal and the *confidence in the decision (dec, conf)*. In Figure 3.2 b), there are 4 levels for decision scale from *"rejectable"* to *"acceptable"* and 6 levels for Confidence Scale from *"lack of confidence"* to *"for sure"*. The solid dot in the matrix represents the evaluation of this goal by an expert. Here, the expert *"accepts"* this goal *"with very high confidence"*. The decision *"acceptable"* indicates that the assessor believes that all the low-level requirements were actually covered. Moreover, the *"very high confidence"* comes from a relatively high coverage rate and thorough explanation of discrepancies in evidence S1.

In order to propagate the expert judgement in the safety case model, we have to quantify these levels into numeric values. Then, we may aggregate the judgements to obtain the assessment results of the higher-level goals. As mentioned in the Section

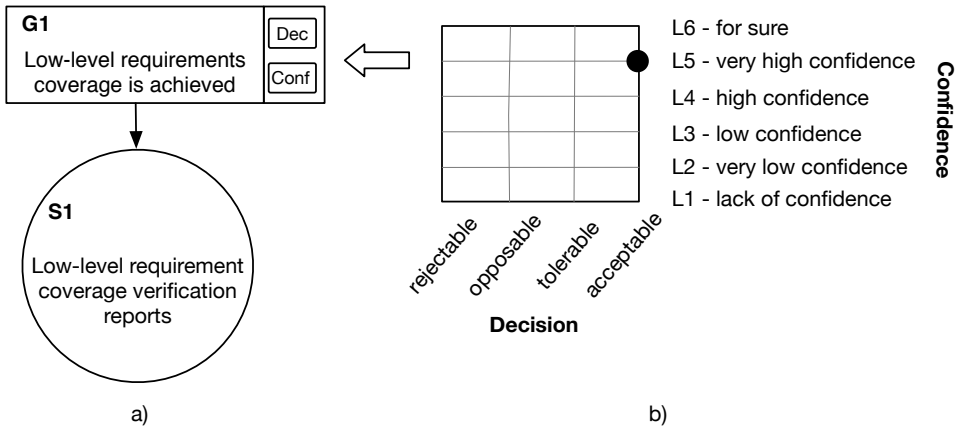


Figure 3.2: An evaluation matrix for safety argument

Table 3.1: Values of the decision on a goal

Decision	Rejectable	Opposable	Tolerable	Acceptable
[0,1]	0	0.33	0.67	1

Table 3.2: Values of the confidence in the decision of a goal

Confidence level	L1	L2	L3	L4	L5	L6
[0,1]	0	0.2	0.4	0.6	0.8	1

3.2.3, the expert judgements expressed by *(dec, conf)* will be transformed into the 3-triple $(bel, disb, uncer)$. Thus, the judgements fit the input of the argument assessment model proposed in previous chapter based on D-S theory. In fact, this step is used to formalise the expert judgements into mass functions in order to take advantage of the D-S Theory to combine uncertain information. As elaborated in the previous chapter, this uncertainty theory offers a powerful tool to explicitly model and process information with uncertainty.

We firstly assume that, for both dimensions of *"decision"* and *"confidence in decision"*, the levels are evenly and linearly distributed. With respect to the value limits of parameters based on D-S theory, the value range for both *"decision"* and *"confidence level"* are $[0, 1]$. The values of the scales are illustrated in Table 3.1 and 3.2.

Then, we would like to transfer the expert judgements $(dec, conf)$ into the 3-triple $(bel, disb, uncer)$. This transformation is intuitively illustrated in Figure 3.3. The left figure a) presents the evaluation matrix; and the right figure b) introduces the Jøsang triangle. This opinion triangle is proposed by Jøsang Jøsang [2001] in his subjective logic for reasoning about trust propagation in secure information systems. The opinion of an agent (or an expert) about the trust is graphically

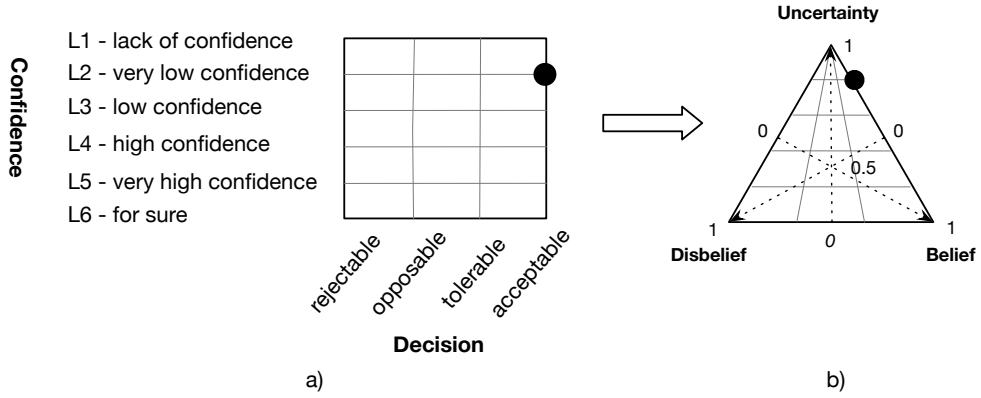


Figure 3.3: Transfering the expert opinions to belief functions

described in this triangular coordinate system with 3 axes: *belief, uncertainty, disbelief*. They are compatible with the functions in D-S theory. Thus, the Jøsang triangle is considerably suitable to show the measures for the trustworthiness of a goal used in our approach. A change of the order for the *"confidence"* levels in the evaluation matrix aims to be consistent with the direction of *"uncertainty"* axis of the opinion triangle.

The formalised transformation method is based on the adopted definitions of *decision* (dec_A) and *confidence in the decision* ($conf_A$) for a goal A from the work of Cyra and Gorski Cyra and Gorski [2011]. These two measures are also defined in compliance with the belief functions theory. The definitions in [Cyra and Gorski, 2011] use the *belief function* and *plausibility function* of D-S theory, as shown in Equations (3.2) and (3.3).

$$conf_A = bel_A + 1 - pl_A, \qquad conf_A \in [0,1] \tag{3.2}$$

$$\begin{cases} dec_A = \dfrac{bel_A}{bel_A + 1 - pl_A}, & bel_A + 1 - pl_A \neq 0 \\ dec_A = 1, & bel_A + 1 - pl_A = 0 \end{cases} \tag{3.3}$$

Since in our proposed method, the 3-triple *(bel,uncer,disb)* is adopted to measure the trustworthiness of a goal, the plausibility function pl_A in Equation 3.2 and 3.3 has to be transformed into *"belief"* (bel), *"uncertainty"* ($uncer$) and *"disbelief"* ($disb$). Let's recall the definitions of the trustworthiness of a goal with mass function (shown in Equation 3.3.1).

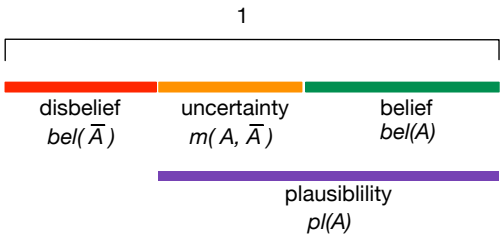


Figure 3.4: The measures of truth of statement A with D-S theory

$$trust_A : \begin{cases} bel_A \ = bel^{\Omega_A}(\{A\}) = m^{\Omega_A}(\{A\}) \\ disb_A \ = bel^{\Omega_A}(\{\overline{A}\}) = m^{\Omega_A}(\{\overline{A}\}) \\ uncer_A \ = m^{\Omega_A}(\{A, \overline{A}\}) = 1 - m^{\Omega_A}(\{A\}) - m^{\Omega_A}(\{\overline{A}\}) \end{cases} \tag{3.4}$$

Compared with the definition of plausibility function[1], the plausibility function of goal A, $pl_A = m(\{A\}) + m(\{A, \overline{A}\})$. In Figure 3.4, the relationship among these measures are illustrated. Thus, $m(\{\overline{A}\}) = 1 - pl_A$. This is the mass for the degree of disbelief that we place in goal A, $disb_A$ (see Equation 3.3.1). So, we slightly change these definitions (Equation 3.2 and 3.3) to be in accordance with the notation of our approach by replacing $1 - pl_A$ with $disb_A$. Moreover, in the original definition of the *decision* (Equation 3.3), when $bel_A + 1 - pl_A = 0$, the $dec_A = 1$, that is, "acceptable". However, $bel_A + 1 - pl_A = 0$ is equivalent to $m(\{A\}) + m(\{\overline{A}\}) = 1 - m(\{A, \overline{A}\}) = 0$. Then, $m(\{A, \overline{A}\}) = uncer_A = 1$, which means maximum uncertainty, or a complete lack of knowledge. We consider that it would be more reasonable to assign the decision "rejectable" ($dec_A = 0$) for fully uncertain case. Therefore, the modified definitions are presented in Definition 3.3.1.

Definition 3.3.1 *The expert decision in a statement and the corresponding confidence in this decision are defined as:*

$$conf_A = bel_A + disb_A = m(\{A\}) + m(\{\overline{A}\}) \tag{3.5}$$

[1]Recall of the definition of plausibility function (detailed in Section 1.3.3):
The *plausibility function* is the sum of the masses that *might* support P. The function pl ($2^{\Omega} \rightarrow [0, 1]$) is defined as:

$$pl(P) = \sum_{M \subseteq \Omega, M \cap P \neq \emptyset} m^{\Omega}(M) \quad \forall P \subseteq \Omega$$

$$\begin{cases} dec_A = \dfrac{bel_A}{bel_A + disb_A} \\ \qquad = \dfrac{m(\{A\})}{m(\{A\}) + m(\{\overline{A}\})}, & bel_A + disb_A \neq 0 \\ dec_A = 0, & bel_A + disb_A = 0 \end{cases} \tag{3.6}$$

Where $conf_A$, $dec_A \in [0, 1]$.

Furthermore, the inverse functions from $(dec_A, conf_A)$ to $(bel_A, uncer_A, disb_A)$ are given in the Equation 3.7.

$$\begin{cases} bel_A = conf_A * dec_A \\ disb_A = conf_A * (1 - dec_A) \\ uncer_A = 1 - bel_A - disb_A \end{cases} \tag{3.7}$$

When the $(bel_A, uncer_A, disb_A)$ needs to be transformed into $(dec_A, conf_A)$ after the propagation of trustworthiness in the argument model, calculated values may not be exactly the ones of 4 decision levels and 6 confidence levels. If so, these numbers should be rounded off to the nearest levels. A conservative way is to choose the low level close to the calculated number. An example is given later on in Section 3.3.3.

3.3.2 Integrated confidence assessment model

With the help of the approach for the judgement extraction presented in Section 3.3.1, our proposed confidence assessment framework can be updated by integrating this approach. Concerning the double-leaf argument of *"sub-goals B and C support top goal A"*, our framework can be regarded as a function f:

$$(dec_A, conf_A) = f[(dec_B, conf_B), (dec_C, conf_C), w_B, w_C, v, < type >] \tag{3.8}$$

where inputs are the evaluation of sub-goals B and C and their weights, and the output is the assessed results of top goal A. More generally, this framework can be applied to a safety argument with multiple sub-goals and more hierarchical levels, due to the general version of aggregation rules.

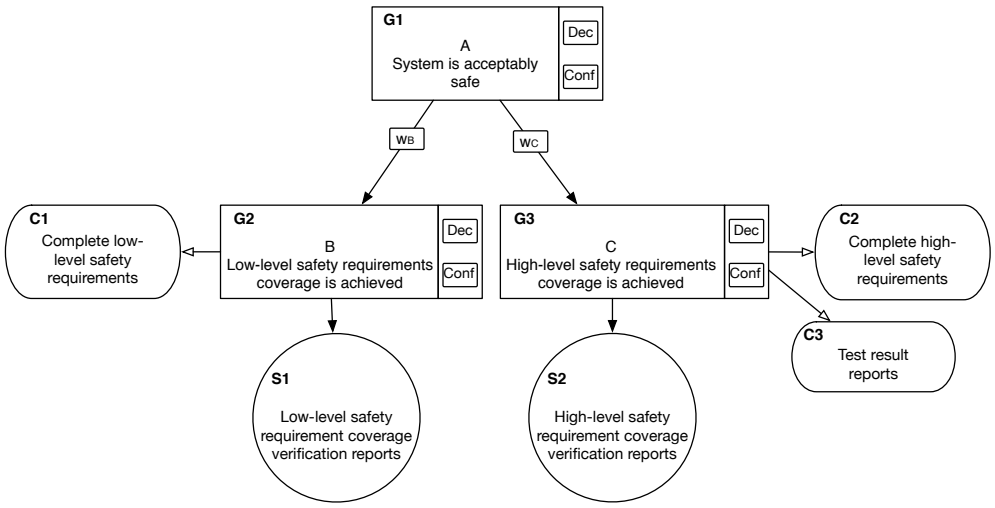


Figure 3.5: A safety argument example to be estimated

3.3.3 Example of judgement estimation and propagation

In this subsection, we use a fragment of GSN argument (shown in Figure 3.5) as an example to present the calculation process of the proposed confidence assessment model. In this argument fragment, we assume that *"G1: system is acceptably safe"* (top-goal A), if *"G2: Low-level safety requirements coverage is achieved"* (sub-goal B) and *"G3: High-level safety requirements coverage is achieved"* (sub-goal C) are fulfilled. The confidence in B is based on the assessment of sub-goals B and C. The purpose of this example is to simply illustrate the propagation calculation, rather than to deduce the relating parameters. Thus, we utilise some arbitrary values of the assessment results: sub-goal B $(dec_B, conf_B)$ =*("opposable","L2-very low confidence")* and sub-goal C $(dec_C, conf_C)$ =*("acceptable", "L5-very high confidence")*. For sub-goal B, the assessor *weak rejects* it; this decision is based on insufficient evidence (very low confidence). For sub-goal C, the assessor accepts it because of the sufficient positive evidence for sub-goal C (*very high confidence*). The assessment results are marked in the corresponding evaluation matrices for B and C in Figure 3.6.

Considering the appropriateness parameters, we choose the values $w_{B\times C\to A} = 0.5$ (complementary argument) and the equal contributing weights $w_B = w_C = (1 - w_{B\times C\to A})/2 = 0.25$. The low-level requirements coverage is verified through the structural coverage analysis of the testing results; the high-level requirements coverage is verified based on testing results. B and C are linked to each other, but they also cover two different aspects. Thus, they are considered as partial

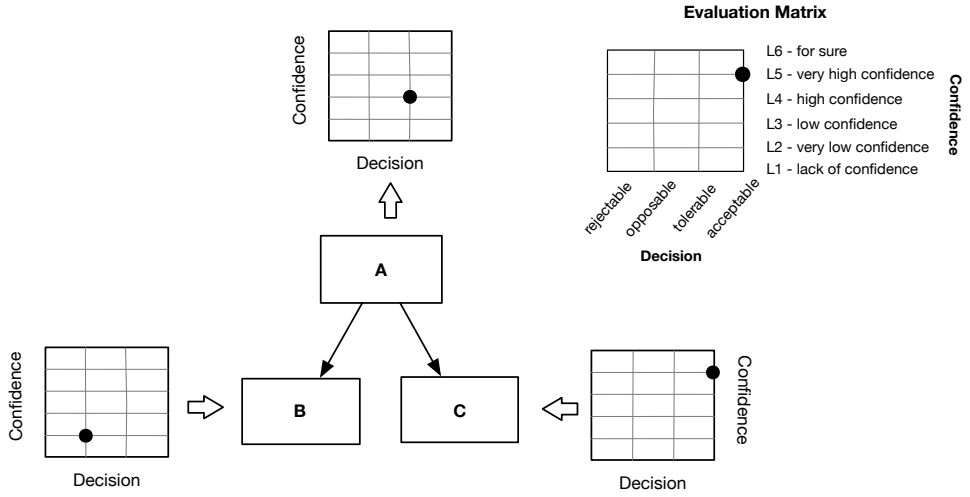


Figure 3.6: Illustration of the confidence propagation

Table 3.3: Confidence aggregation rules for complementary arguments

Types	Aggregation rules
C-Arg	$bel_A = [bel_B \cdot w_B + bel_C \cdot w_C + bel_B \cdot bel_C(1 - w_B - w_C)]v$ $disb_A = \{disb_B \cdot w_B + disb_C \cdot w_C + [1 - (1 - disb_B)(1 - disb_C)](1 - w_B - w_C)\}v$ $uncer_A = 1 - bel_A - disb_A$

complementary arguments. A strategy for estimating these parameters will be briefly introduced in next section. In next chapter, an experimental application of this strategy presents the parameters estimation with the help of a survey.

A three-step process of the calculation presents the assessment of confidence in A, based on the chosen values for the judgement opinions and appropriateness of B and C:

- Transforming the evaluation $(dec, conf)$ of B and C to $(bel, uncer, disb)$ using Equation 3.7. Considering some arbitrary values $(dec_B, conf_B) = (0.33, 0.2)$ and $(dec_C, conf_C) = (1, 0.8)$, then we calculate that $(bel_B, uncer_B, disb_B) = (0.066, 0.8, 0.134)$ and $(bel_C, uncer_C, disb_C) = (0.8, 0.2, 0.0)$.

- Aggregating the trustworthiness of B and C with the aggregation rules of complementary argument. In Table 3.3, we recall these rules previously presented in Section 2.4. $(bel_B, uncer_B, disb_B) = (0.243,\ 0.657, 0.101)$.

- Calculating the decision on A and the confidence in the decision $(dec_B, conf_B) = (0.707, 0.343)$ according to Definition 3.3.1. The level of decision and confidence in this decision are selected by the nearest value of the results. Thus,

the assessment results for top goal B is transformed to ($dec_B, conf_B$)= *("tolerable", "L3-with low confidence")* (see Figure 3.6). The obtained *decision* and *confidence* in the decision for top goal B are located in the middle of sub-goals B and C. This is reasonable considering the values chosen for the appropriateness for the sub-goals.

In this section, we present the calculating process to propagate the available judgements of sub-goals for a parametric argument model. In next section, we will analysis how the variation of the parameters affects the propagation results.

3.3.4 Sensitivity analysis

We did the sensitivity analysis for the original confidence assessment model in the previous chapter, which helps to identify the characteristics and behaviours of this model. In this chapter, the introduction of the judgement extraction to the confidence assessment model may lead to the change of the performance of the assessment model. Thus, the framework function (see Equation 3.8) need to be further analysed. We adopt the tornado graph again to implement the sensitivity analysis. This method is introduced in Section 2.6.

For a better readability, we recall the principle of this sensibility analysis. Considering a function $f(x_1, ...x_n)$, where values $X_1, ..., X_n$ of the variables x_i have been estimated, the tornado analysis consists in the estimation (for each $x_i \in [X_{min}, X_{max}]$) of the values $f(X_1, ..., X_{i-1}, X_{min}, X_{i+1}, ...X_n)$ and $f(X_1, ..., X_{i-1}, X_{max}, X_{i+1}, ...X_n)$, where X_{min} and X_{max} are the maximum and minimum admissible values of variables x_i. Hence for each x_i, we get an interval of possible variations of function f. The tornado graph is a visual presentation with ordered intervals. In our case, we estimate the decision on A (dec_A) and confidence in the decision on A ($conf_A$) with corresponding intervals for v, dec_B, $conf_B$, dec_C, $conf_C$, w_B and w_C.

Taking the example of the argument *"sub-goals B and C support top goal A"*, we analyse the evaluation and propagation of the expert judgements within the framework of both complementary and redundant arguments. The basic values (X_i) and intervals $[X_{min}, X_{max}]$ for each parameter are shown in Table 3.4. The basic values (X_i) are arbitrarily provided; and the intervals $[X_{min}, X_{max}]$ are then deduced according to the requirements for the parameters in the formulas: $dec_i, conf_i, w_i, v \in [0, 1]$ and $\sum_{i=1}^{n} w_i \leq 1$. For instance, the interval for w_B is [0,0.9], because $w_C = 0.1$ and the sum of them should not be more than 1.

Table 3.4: Example of values and intervals for sensitivity analysis

	v	dec_B	$conf_B$	dec_C	$conf_C$	w_B	w_C
Basic value X_i	-	*tolerable*	*L5*	*opposable*	*L4*	-	-
	0.9	0.67	0.8	0.33	0.6	0.4	0.1
$[X_{min}, X_{max}]$	[0,1]	[0,1]	[0,1]	[0,1]	[0,1]	[0,0.9]	[0,0.6]

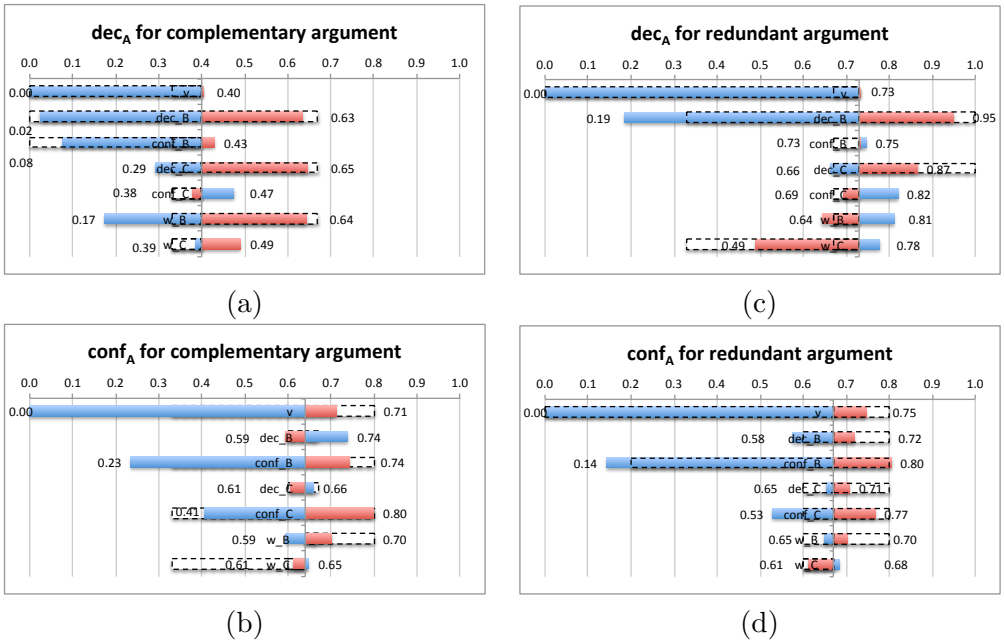


Figure 3.7: Tornado graphs of two argument types for assessment framework

The values of the decision and confidence in this decision on A $(dec_A, conf_A)$ are calculated based on the assessment measures of B and C. The calculation processes are elaborated in Section 3.3.3. With the basic values in Table 3.4, the values of $(dec_A, conf_A)$ equal to $(0.40, 0.64)$ for complementary argument and $(0.73, 0.67)$ for redundant argument. These values are set as the positions of vertical axes in corresponding tornado graphs. To determine the sensitivity to dec_B, for instance, we keep the basic values for all other variables and calculate the values dec_A with $dec_B = 0$ and $dec_B = 1$: the obtained values of the decision on A are $[0.02, 0.63]$ for complementary argument and $[0.19, 0.95]$ for redundant argument. The same approach is applied for other parameters. The analysis results are presented in Figure 3.7. The dashed bars in the graphs represent the rounded values corresponding to the discrete *decision* or *confidence* level.

All graphs show that v is the most influencing factor on the left side of the vertical axis (due to the nature of a discounting factor). The decision on A is 0 (*"rejectable"*) when $v = 0$. This result can also be deduced from the structure of

the aggregation formulas (see Table 3.4 and 3.5), as v is a common factor. Thus, v is a sensitive point for these formulas. We may discover from figures (a) and (c) that the impact on A (dec_A) of decision on B (dec_B) is greater than C (dec_C). This is due to the higher weight of B (w_B) than C (w_C). These findings are the same with the conclusion of the sensitivity analysis for the trustworthiness propagation in last chapter, which further validates the proposed framework.

Compared with decision of sub-goals (dec_B, dec_C), the confidence ($conf_B$,$conf_C$) have relatively less impacts on the decision of A (dec_A); they mostly changes the confidence in A ($conf_A$). Thus, we may conclude that the decision of A (dec_A) and the confidence the decision $conf_A$ are influenced separately by the decision of sub-goals (dec_B, dec_C) and the corresponding confidence. This also implies that if we want to increase the dec_A, we need to focus on increase dec_B and dec_C, and similarly for $conf_A$.

3.4 Parameter estimation

In this proposed assessment framework, a very important step is to determine the argument types (complementary or redundant) and to estimate the weights of sub-goals (e.g. w_B, w_C) in order to complete the assessment model. Due to the subjectivity of the determination of these parameters, it is not possible to deduce them from the model itself. It would be reasonable to estimate these parameters depending on the data from the expert's decision-making. Our plan is to derive the values of these parameters from the expert judgement on some generic argument fragments. Taking an example of a double-node argument, we propose to firstly provide some pre-determined judgements of sub-goals as inputs to the arguments (see Figure 3.8). Safety experts are asked to make their decision on the top goal according to the inputs for each argument. Their decisions are then input in our assessment framework (function f referring to Equation 3.8) together with the initial judgements of sub-goals in order to estimate the parameters under interest. This proposal for parameter estimation is based on two hypotheses:

- H1: the experts under investigation have preference for a certain argument type for a given argument. They can also distinguish the different degrees of contribution of sub-goals to the top goal. These opinions may be implicit. However, they are conveyed in the decision that the experts make.

- H2: our confidence assessment framework is able to describe the mental model of the experts for the confidence propagation.

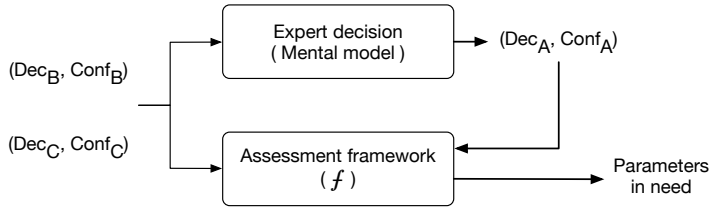


Figure 3.8: Parameter estimation for confidence assessment model

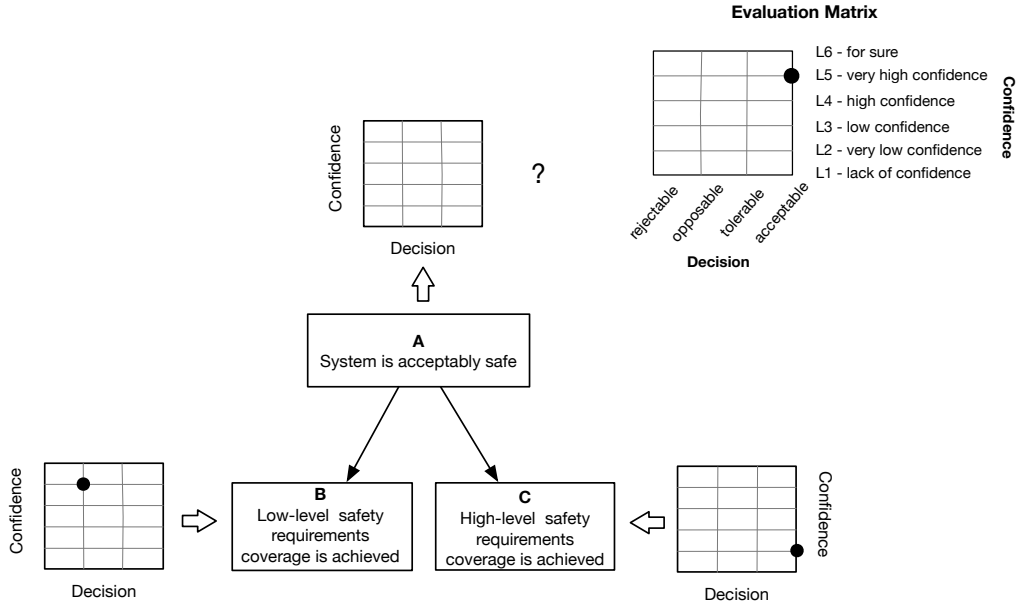
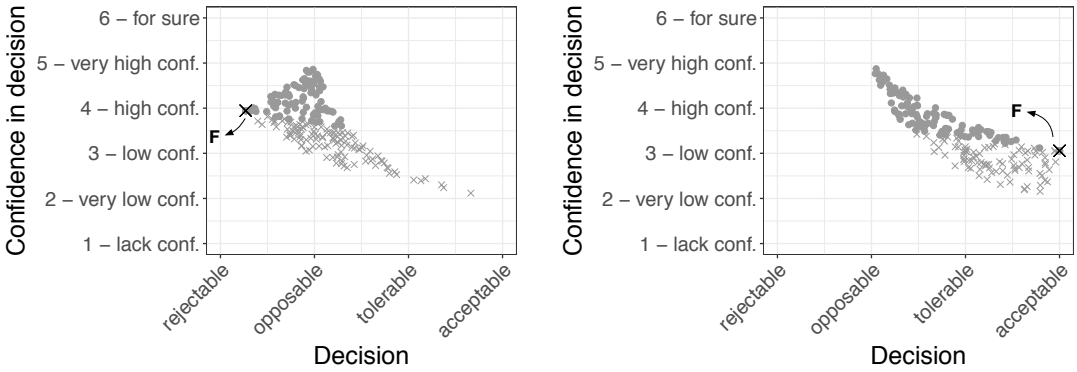


Figure 3.9: An example of question to collect expert opinion

Figure 3.9 presents one question example for parameter estimation. The predetermined opinions for two sub-goals are provided. For sub-goal B, the opinion is *"opposable"* (weak reject) with *"very high confidence"* $(Dec_B, Conf_B)$; and for sub-goal C, the opinion is *"acceptable"* with *"very low confidence"* $(Dec_C, Conf_C)$. The respondent is about to provide his/her *decision* on the top goal and the *confidence* in this decision, based on the experience and understanding of this argument fragment.

After the information collected from safety experts, we need to analyze the data and deduce the parameters. In order to have a intuitive understanding of the confidence propagation of framework, we propose to illustrate this propagation results in the evaluation matrix (see Figure 3.10). This also helps to easily compare them with the answers by experts. The results are calculated based on the inputs of sub-goals according to the aggregation rules in Table 2.5 and 2.6 with the Definition 3.3.1. We call these calculating results *"theoretical data"*. The theoretical data



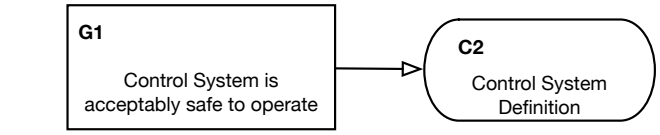
a) Complementary argument b) Redundant argument

Figure 3.10: Theoretical data calculated based on pre-determined inputs

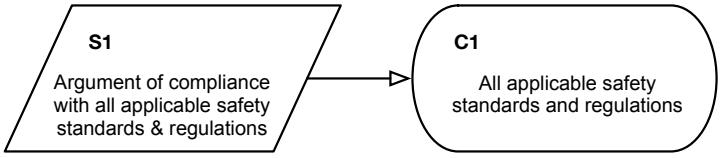
of complementary and redundant arguments are shown as a cluster of grey dots derived from all possible the values of w_B and w_C. Figures a) and b) correspond to the same inputs of B and C in Figure 3.9. According to the proposed assessment approach, we calculate the values of $(dec_A, conf_A)$ from inputs $(dec_B, conf_B)$ and $(dec_C, conf_C)$. The values are then plotted in the evaluation matrix. The *solid dots* represent the values with the constraint that $w_B > w_C$; whereas the *crosses* represent the values of $w_B \leq w_C$. In the figures, the "F" letters represent the output of a special cases: *fully complementary argument (FC-Arg) and fully redundant argument (FR-Arg).*

We can clearly discover the different behaviours between the confidence propagation of the complementary and redundant arguments. For the former, most of the calculating results trend to place on the left of the decision *"opposable"*; for the limit situation (FC-Arg), the decision approaches to *"rejectable"* with *"high confidence"*. These opinions are even lower than the opinion of B. This shows the mutual contribution of sub-goals of the complementary argument is less than each sub-goal. It is due to the AND-gate influence among these sub-goals. On the contrary, for the latter, most of the calculating results concentrate on the right of the decision *"tolerable"*; for the limit situation (RC-Arg), the decision is *"acceptable"* with *"low confidence"*, which inherits the higher level of opinion between the sub-goals.

Once the expert opinions are obtained, we can directly compare the two sources of data in the evaluation matrix. Some rough information, such as the preliminary judgement of the validity of the proposed assessment framework, the argument types and the relative importance of the weights among sub-goals, might be determined. Then, more accurate values of these parameters are to be estimated by statistical analysis. Here, we briefly introduce the process of parameter estimation. In next



a) Linked with a Goal



b) Linked with a Strategy

Figure 3.11: The usage of a context

chapter, a case study focusing on this work is implemented.

3.5 Discussion on context elements in GSN

In a GSN safety case, the contextual and explanatory elements to the *goal* and *strategy* are of extreme importance for the argumentation. As introduced in Chapter 2, these elements include: *context*, *justification* and *assumption*. Amongst these three elements, the context is the most commonly used. Let us discuss how to assess the context element with considering its intended roles according to the GSN standard GSN Standard [2011].

In simple terms, a context is a reference to contextual information or a statement. It can be connected to a goal or to a strategy (see Figure 3.11). The purpose of using a context is to:

- *"declare supplementary information related to the claim made in Goal G1"* [GSN Standard, 2011, Section 1.3.6] or
- *"declare supplementary information related to the definition or explanation of terms used in the strategy"* [GSN Standard, 2011, Section 1.3.11].

This normative document specifies that a context or a contextual statement aims to define the scope over which the claim in Goal G1 is made or the Strategy S1 is applied. Hence, the supporting arguments for the goal or derived from the strategy should be compliant with the context.

On one hand, the goals and solutions are the main compositions of a structured argument. The validity of a goal should be ensured by sub-goals or solutions. Thus,

the "supplementary information" (i.e. contexts) for a goal has to be reviewed to verify the conformity of the overall argument.

On the other hand, *"strategy is the description of how to realise a goal decomposition"*, which can be regarded as descriptive information or justification in the argument inference. Therefore, we assume that the contexts linked to strategies do not impact the validity of to higher-level goal.

In the confidence assessment framework, we consider that the parameter v can be used to evaluate the contexts. v, as a discounting factor, serves to decrease the aggregated certainty (*belief* and *disbelief*), that is, to increase the uncertainty. For the contexts linked to a goal, we consider $v \leq 1$. The value of v depends on the assessment of all contexts, regarding to the appropriateness and the sufficiency of the supplementary information. Similar scalable method for the judgement extraction as in Section 3.3.1) may be adopted. For the the contexts linked to a strategy, we consider $v = 1$. This is a naive proposition of the assessment of contextual information. It will not be included in the scope of this thesis.

3.6 Conclusion

In this chapter, we propose a 4-step confidence assessment framework for the safety case of a critical system. This work is on the basis of the quantitative model of the confidence assessment for the safety argument proposed in Chapter 3. A method for the judgement extraction is integrated to the mathematical model, which makes the model more practical for a real engineering application. With this assessment framework, an assessor of an argument needs only to provide his/her *decision* and the *confidence in the decision* for the lowest level of sub-goals of the argument based on the available evidence. Then, these opinions will automatically be combined to generate the *decision* and the *confidence in the decision* of the top goal. Sensitivity analysis is carried out to identify the characteristics and behaviours of the new integrated model. We also present the principles for implementing the parameter estimation (Step 2 of the framework). Furthermore, we open up a supplementary discussion about the treatment of the contextual elements in GSN.

Chapter 4

Case study of railway safety cases

Contents

4.1 Introduction

In the previous chapters, an integrated confidence assessment framework for safety arguments is proposed. In the meantime, multiple aspects regarding this framework are to be validated, such as the feasibility of the judgement extraction, parameter estimation, reusable parametric argument fragments, possible applications of this framework, etc. In this chapter, we carried out a case study based on railway safety cases for both validation and application of this framework. This is an extension of the published work [Wang et al., 2017b].

This case study aims to apply the proposed assessment framework on safety cases and to provide a solution to realise the parameter estimation. In railway domain, the European standard EN50129 [2003] gives guidance on the establishment of safety cases. Thus, our study starts from the analysis of the relating railway standards (Section 4.2). This study might also be a direction to generate general and reusable safety case models based on our approach, which may facilitate the quantitative argument assessment. Thus, we build a structured argument base on the part of safety requirements extracted from the standards (Section 4.3). Then, the parameter estimation of a safety argument is realised with the help of a survey amongst safety experts (presented in Section 4.4). Meanwhile, the feasibility of the judgement extraction is also tested. In Section 4.5, we analyse the survey results. Section 4.6 describes how to apply this framework to a simplified example of WSP system (Wheel Slide Protection, railway equipment like an Anti-lock Braking System for automotive).

4.2 The railway safety standards for signalling systems

For the railway system in Europe, the European Railway Agency (ERA) proposes a framework of Common Safety Method (CSM) [ERA, 2015] to standardise risk evaluation and assessment process. The CSM is a general safety regulation on very high level aiming to European nations. It suggests using the EN5012x standards to harmonised design targets for railway technical systems. This series of standards provide a guideline for ensuring the functional safety of safety-related electronic systems for railway signalling applications.

The EN5012x standards are derived from the general standard IEC61508 [2010]. EN50126 [1999] is mainly used to manage the railway system RAMS (Reliability, Availability, Maintainability, and Safety) throughout the life-cycle process. EN50128 [2011] focuses on the control and protection software applications, which must meet the software safety integrity requirements. EN50129 aims to provide the conditions for the acceptance and approval of safety-related systems. The evidence of satisfying these conditions is explicitly required to be documented in a safety case. In this standard, safety case is defined as *the documented demonstration that the product complies with the specified safety requirements.*

EN50129 introduces a high-level structure for any safety case of the railway signalling system. It provides documented evidence that justifies the rigorous development processes and safety life-cycle activities, ensuring adequate confidence in the critical system safety. The structure is mainly based on the acceptance con-

ditions: 1) evidence of quality management, 2) evidence of safety management, 3) evidence of functional and technical safety. We choose to study this standard as a basis for our case study because the concept of safety case is explicitly mentioned.

Amongst the evidence, functional and technical safety is of the utmost importance. It is required to explain the technical safety principles for design and to reference all the available evidence. The concept of Safety Integrity Level originated from IEC61508 [2010] is used in the EN5012x series standards. Four levels (SIL1-4) are associated with the four severity classes (Insignificant, Marginal, Critical, Catastrophic). The SIL4 is the highest safety integrity level. The recommended safety assurance techniques are differentiated according to these SILs. In Section 4.3, we further discuss the technical safety arguments proposed by EN50129.

4.3 Safety Case Modelling based on EN50129

EN50129 provides a high-level argument structure as the guideline for building safety cases for railway signalling systems. Meanwhile, the required techniques and measures to avoid systematic faults are listed for system life-cycle activities. However, the rationale behind how these techniques serve the objectives is not indicated in the argument structure. In this section, the high-level safety case structure and technique checklists are translated with the Goal Structuring Notation presented in Section 1.2.3.4. Then, a proposal for the necessary but missing inference between them is given based on the analysis of standards and engineering experience.

4.3.1 Modelling the Standard with GSN

EN50129 presents a clear high-level structure for the safety case. This structure is reflected, in the GSN argument model, as multiple layers of goals. In fact, the goals are interpreted from the headings of the parts or sections of the safety case indicated by this standard. The reasoning behind the sub-goals is also explicitly given. We translate the sub-goals and reasoning into GSN models (presented in Figure 4.1 and Figure 4.2). These models provide a more intuitive presentation of the sub-goals and inference processes. They also contribute to the consistency of the analysis in the following sections.

In Figure 4.1, the first two layers of the GSN model (left) are designed based on the *safety case structure* (right) provided in EN50129. Part 1 of the safety case is the definition of the system. It is considered as contextual information (*C1*) for the entire safety argument. Another context (*C2*) is the existing international and

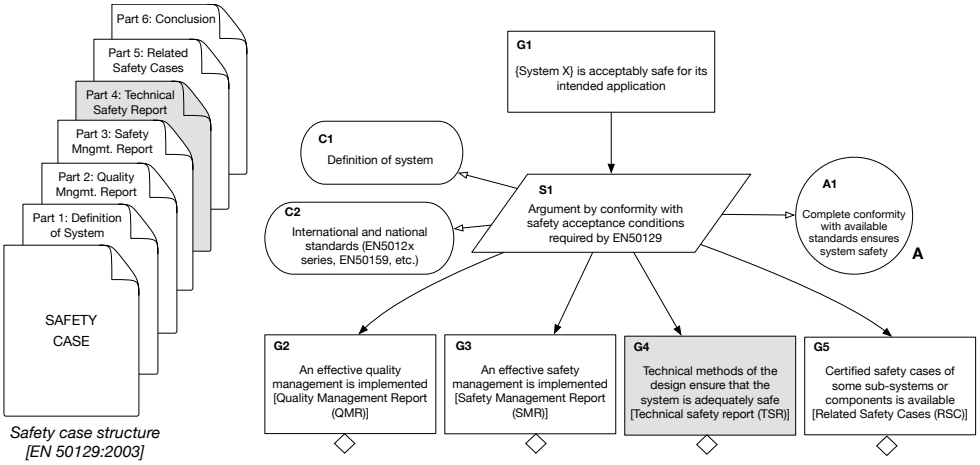


Figure 4.1: The highest level model of Safety Case

national standards for railway electronic system, i.e EN5012x, EN50159 [2010] and the related regulations. Moreover, this safety case is supposed to be built based on the assumption (*A1*) that the conformity with the available standards leads to an acceptable safety of the system. The top goal *G1: {System X} is acceptably safe for its intended application* is broken down into 4 sub-goals according to the *Part 2-5* of the Safety Case. They are the claims of achievement of effective *quality management* (G2), *safety management* (G3), *safety technical methods* (G4) and the availability of *related certified safety cases* (G5) for sub-systems or components.

On the basis of these sub-goals, the standard provides more guidelines for formulating the safety case. Goal G4 is taken as an example to represent the third layer of sub-goals. In Figure 4.2, G4 is supported by 5 sub-goals (G6-G10) as the trustworthy technical evidence to ensure system safety (S2). They correspond to the *Section 2-6* of the *technical safety report* required by EN50129. These sub-goals concern respectively the requirement-assured functionality (G6), the hardware fault effect analysis (G7), the assurance of functionality and safety considering external influence (G8), the definition of rules, conditions or constraints to be complied with during other phases of the system life-cycle (G9), and finally, the safety qualification test under operational conditions (G10).

Then, the goal G6 can be further broken down into goals G11-G14 following the strategy (S3) that the fulfilment of system and safety requirements can guarantee the correct functional operation of systems. As shown in Figure 4.2, the description of system architecture (C3) and system interface (C4) is the context for this part of the argument. Then, the four sub-goals are the claims for the fulfilment of the system (G11) and the safety (G12) requirements, as well as the correct functionality

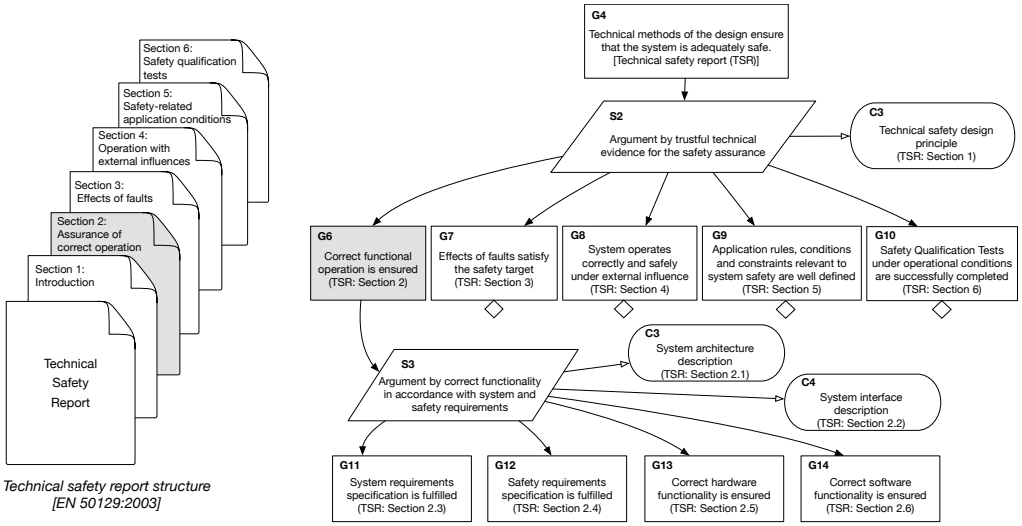


Figure 4.2: The main structure of the Technical Safety Report

of hardware (G13) and software (G14).

Briefly, this tree-structured model from G1 to G14 illustrates one complete branch of the high-level safety case in compliance with EN50129. Compared with pure textual argumentation, this model focuses exclusively on the key objectives to be achieved and the relationship inferred amongst them.

4.3.2 Technical Safety Evidence

Besides the argumentation structure of the safety case, the safety evidence supporting the goals is of equal importance. In the EN50129 and EN50128, the recommended techniques or measures are provided as normative information. For each technique or measure, different requirement degrees are prescribed according to different Safety Integrity Levels (SIL). There are 5 degrees: *Mandatory (M)*, *Highly Recommended (HR)*, *Recommended (R)*, *no suggestion for or against being used (-)* and *Not Recommended (NR)*. These prescriptions are obtained based on years of engineering experience and discussion with relevant experts. For instance, the use of *simulation* is *recommended (R)* for the verification and validation of the functions or systems with SIL2,3,4, not necessarily for SIL1 (see Table 4.1). It indicates that the adoption of *simulation* can increase our confidence in the functional or system safety. Taking another example, in Table A.3 of EN50128: Software Architecture, the technique *artificial intelligence for fault correction* is not recommended (NR) (which is actually coming from the IEC61508). Once this kind of technique is adopted in system design without reasonable explanation, our confidence in system

Table 4.1: Techniques required in the V&V process of system design in EN50129 (excerpt)

Techniques/Measures	SIL 1	SIL 2	SIL 3	SIL 4
1 Checklists	R: prepared checklists, concentration on the main safety issues		R: prepared detailed checklists	
2 Simulation		R	R	
3 Functional testing of the system	HR: functional tests, reviews should be carried out to demonstrate that the specified characteristics and safety requirements have been achieved		HR: comprehensive functional tests should be carried out on the basis of well defined test cases to demonstrate the specified characteristics and safety-requirements are fulfilled	

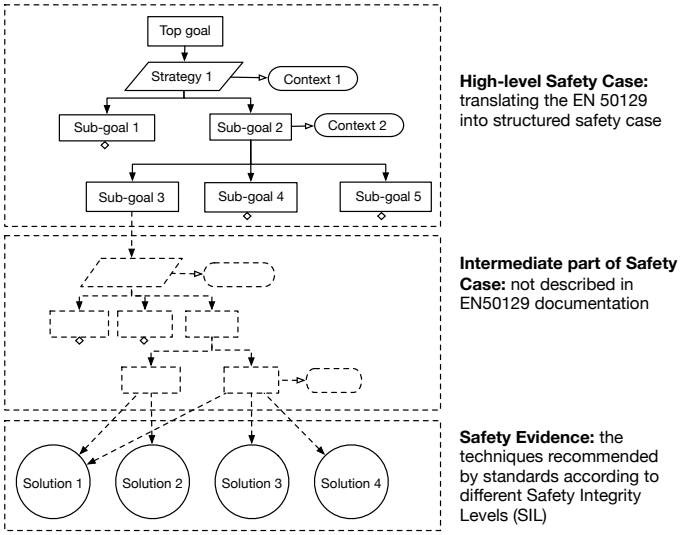


Figure 4.3: Inference gap between high-level goals and safety evidence required by the EN50129

safety may decrease. Therefore, the use or non-use of a technique mentioned in the standards is considered as the evidence for safety assessment. This evidence always appears in the output documents of life-cycle activities, e.g. hazard log, test results, etc.

4.3.3 Intermediate Argument Development for Goal G12

In the previous two sections, we translate the high-level safety argument structure and the safety evidence into the GSN argument models. However, the rationale of how the high-level goals are based on this safety evidence is not directly given in the standards. The organization of arguments is left to engineers to develop. In Figure 4.3, the inference gap is presented with the dashed schema between the *High-level Safety Case* and *Solutions* (also called evidence).

B.2.4 Fulfilment of safety requirements specification

This shall demonstrate how the specified safety functional requirements are fulfilled by the design. All relevant evidence shall be included (or referenced).

EXAMPLE
- design principles and calculations;
- test specifications and results;
- safety analyses and results.

Figure 4.4: The excerpt from EN50129 relating to goal G12

In order to make the rationale of the standard explicit, a fragment of the railway safety case is deduced for the goal *G12: Safety requirements specification is fulfilled by the design* subjected to a system required to reach SIL4. It is broken down into sub-goals and finally supported by the related technical safety evidence required by EN50129. This intermediate part of the arguments is based on the analysis of the EN50129 and our engineering experience for safety assurance. The GSN model is shown in Figure 4.6.

There is little guidance to justify the fulfilment of *G12* (see Figure 4.4). In another section of this standard (5.3.9), verification and validation of safety requirements is developed, and a list of techniques is given in a table (see Table 4.2). In this table, there are 11 recommended techniques and measures for the V&V process. Thus, the supporting *solutions* (Sn12.1-Sn12.11 in Figure 4.6) for goal G12 correspond to these techniques and measures. Note that these techniques are adopted for achieving the SIL4. For the systems with a lower SIL requirement, less evidence is required. For instance, according to EN50129 [2003], *Sn12.4* (audit) and *Sn12.6* (simulation) are not required when the required SIL is less than 4. Additionally, the degree of independence among validators, verifiers, designers and project managers shall be in accordance with the expected SIL of the system under assessment, as shown in the standard (see Figure 4.5). The verification of this independence is actually a part of the safety management activities (i.e. goal G3). Thus, we propose to consider it as a context of the V&V activities in the goal G12 (see context C12.1 in Figure 4.6).

EN50129 provides no or little information for some techniques, such as checklists, simulation, inspection of documentation, etc. In fact, all deliverables related to G12 should also be included as solutions in this part of the argument. We suggest regrouping the techniques by two strategies (corresponding to Figure 4.6):

- Argument by the traceability and satisfaction of all safety requirements (*S12.1*);
- Argument by high confidence demonstrated by actual use (*S12.2*).

Table 4.2: Recommended techniques/measures for V&V process [EN50129, 2003]

Techniques/Measures	SIL 1	SIL 2	SIL 3	SIL 4
1 Checklists	R: prepared checklists, concentration on the main safety issues		R: prepared detailed checklists	
2 Simulation		R	R	
3 Functional testing of the system	HR: functional tests, reviews should be carried out to demonstrate that the specified characteristics and safety requirements have been achieved		HR: comprehensive functional tests should be carried out on the basis of well defined test cases to demonstrate the specified characteristics and safety-requirements are fulfilled	
4 Functional testing under environmental conditions	HR: the testing of safety-related functions and other functions under the specified environmental conditions should be carried out		HR: the testing of safety-related functions and other testing under the specified environmental conditions should be carried out	
5 Surge immunity testing	HR: surge immunity should be tested to the boundary values of the real operational conditions	HR: surge immunity should be tested higher / higher limit than the boundary values of the real operation conditions		
6 Inspection of documentation	HR			
7 Ensure design assumptions are not compromised by manufacturing process			HR: specify manufacturing requirements and precautions, plus audit of actual manufacturing process by safety organisation	
8 Test facilities	R: designer of the test facilities should be independent from the designer of the system or product		HR: designer of the test facilities should be independent from the designer of the system or product	
9 Design review	HR: reviews should be carried out at appropriate stages in the life-cycle to confirm that the specified characteristics and safety requirements have been achieved		HR: reviews should be carried out at appropriate stages in the life-cycle to confirm that the specified characteristics and safety requirements have been achieved	
10 Ensure design assumptions are not compromised by installation and maintenance processes	HR: specify installation and maintenance requirements and precautions		HR: specify installation and maintenance requirements and precautions, plus audit of actual installation and maintenance processes by safety organisation	
11 High confidence demonstrated by use (optional where some previous evidence is not available)	R: 10 000 hours operation time, at least 1 year experience with equipments in operation		R: 1 million hours operation time, at least 2 years experience with different equipments including safety analysis, detailed documentation also of minor changes during operation time	

NOTE Checklists, computer aided specification tools and Inspection of the specification can be used in the verification activity of a phase.

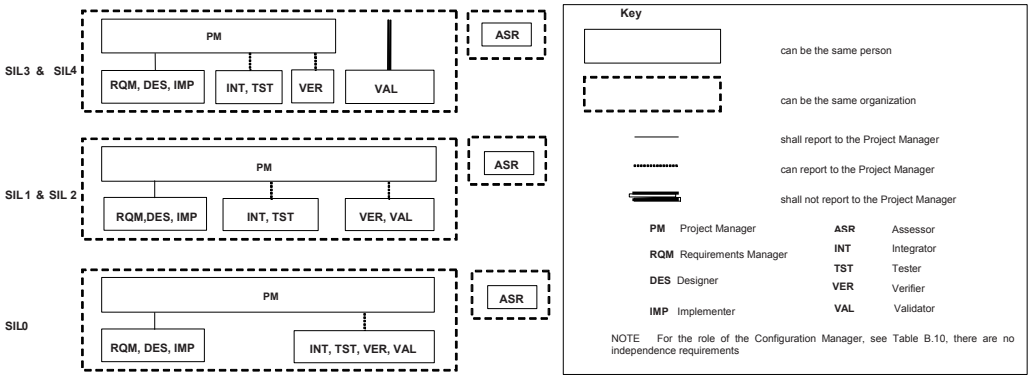


Figure 4.5: Independent requirements for personnel and organisational structure (updated version presented in [EN50128, 2011])

Concerning the strategy *S12.1*, for a newly developed system, the high-level safety requirements are ensured through 1) a vertical view: all safety requirements are traced in each life-cycle phases and 2) a horizontal view: the conclusive validation of the satisfaction of safety requirements. This strategy is actually from the "V-model" representation of the life-cycle introduced in EN50126. This branch of the safety argument is built following the goal-based structure below:

- All safety requirements are traced in each life-cycle phases (*G12.1*)
  - Assurance of document traceability (*G12.4*)
    - Checklists, e.g., checklist for the deliverable required in each life-cycle phase (*Sn12.1*)
    - Inspection of documentation, e.g., coverage verification between high-level and low-level requirements (*Sn12.2*)
    - Design review, e.g., verification of conformity between specification and design implementation, code review (*Sn12.3*)
  - Design assumptions are ensured in manufacturing, installation and maintenance process (*G12.5*)
    - Requirements and precautions, audit report of manufacturing, installation and maintenance processes (*Sn12.4* and *Sn12.5*)
- The satisfaction of all safety requirements are validated by test and simulation (*G12.2*)
  - Validation by simulation (*G12.6*)
  - Validation by functional testing (*G12.7*)
    - Validation by (internal) functional testing (*G12.9*)
    - Validation by functional testing under specified environmental conditions (*G12.10*)
  - *Validation by robustness testing (G12.8)*

Based on the strategy *S12.2*, if the system under consideration is a re-use of a previous system with minor changes, the safe operation history can also contribute to confidence in the fulfilment of safety requirements (G12.3). For SIL4 systems or functions, the safe operational duration is required to exceed 1 million hours, at least a 2-year experience (Sn12.11).

A proposal for the intermediate part of the safety case is presented based on the analysis of EN50129. However, we reached some limitations when following

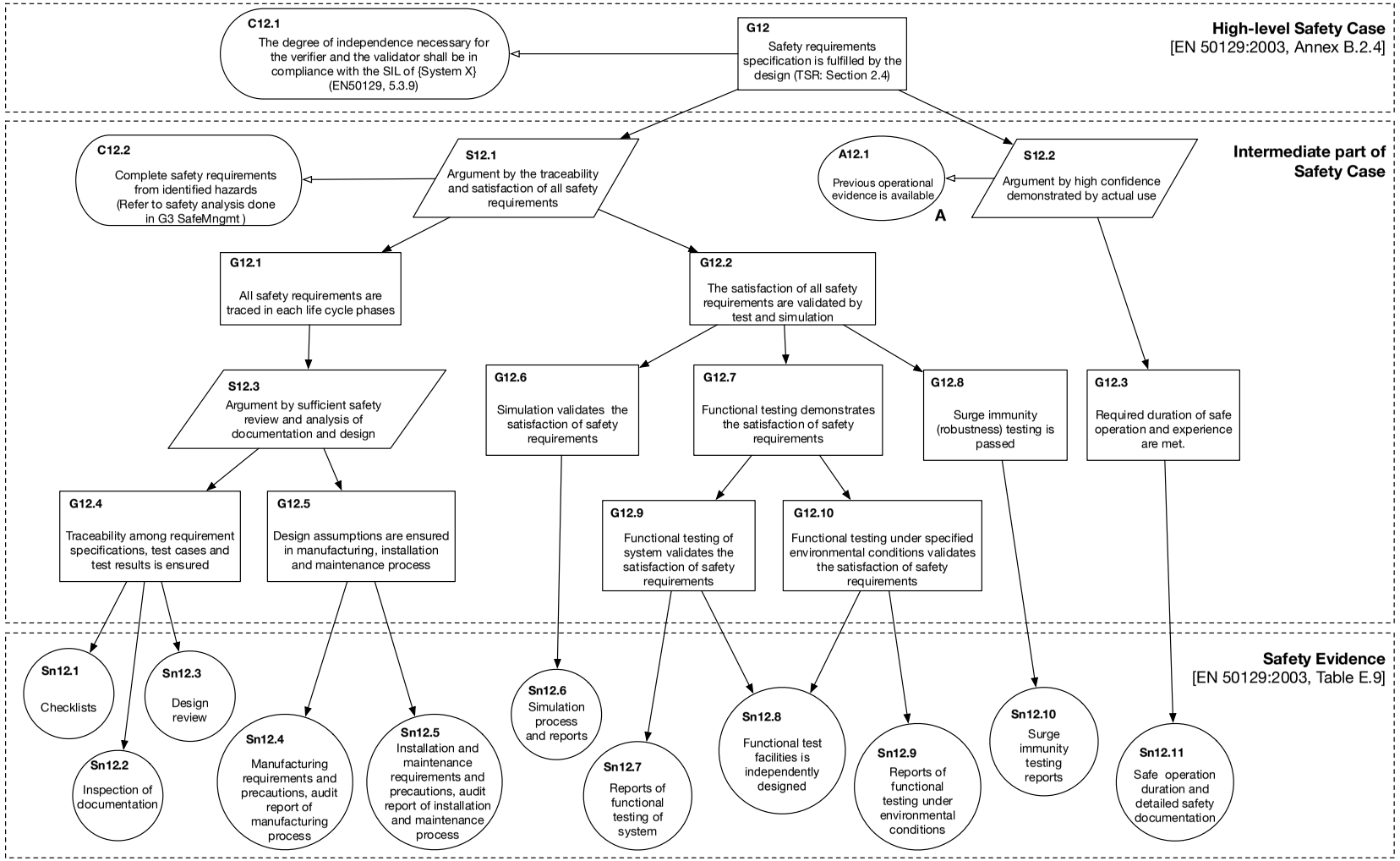


Figure 4.6: GSN presenting the rationale between the goal G12 and relating safety evidence

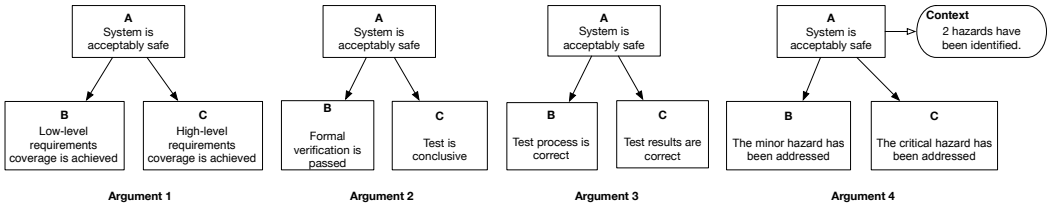


Figure 4.7: Argument fragments questioned in the survey

the guidance to build a reasonable safety case. The techniques requirements in Annex E of the standard are a mix of techniques and objectives. Even for SIL3 and SIL4, some techniques are described in such a general way that the effectiveness for safety assurance might vary over a wide range. In addition, the requirements tables can be used not only in the referred sections but also in other sections without reference. The proposed diagram Figure 4.6 only presents what should be done for one sub-goal (G12), but similar analyses are needed for other sub-goals (see Figure 4.1 and 4.2).

4.4 Capture expert judgement

In the previous chapter, we introduce the methods of expert judgement extraction and parameter estimation (see Section 3.3) used in the assessment framework. In this section, we implement an experimental application by a survey among experts to evaluate these mentioned methods and the framework.

While designing the survey questionnaire, the prior requirement is to make it as clear and simple as possible to avoid misunderstandings and to gather the actual opinions of respondents. At the same time, our objective is to validate our method and not to exploit the full railway standards. Thus, in order to obtain relatively accurate answers, this study focuses on four simple and general safety argument patterns. They are presented in Figure 4.7. These four arguments have the same form with an identical top goal A and two sub-goals B and C. To ensure enough data for parameter estimation, three pairs of inputs are provided for each argument in the evaluation matrix (see Figure 4.8). We associate these inputs to three questions (Q1-Q3). The respondents are asked to make the decision on the acceptance of goal A, and the confidence in this decision based on each pair of inputs.

The decision levels are *rejectable, opposable, tolerable, acceptable*, and the confidence levels in the decision vary from *1-lack of confidence* to *6-for sure*. For a better understanding of the assessment process, an introduction of the evaluation matrix

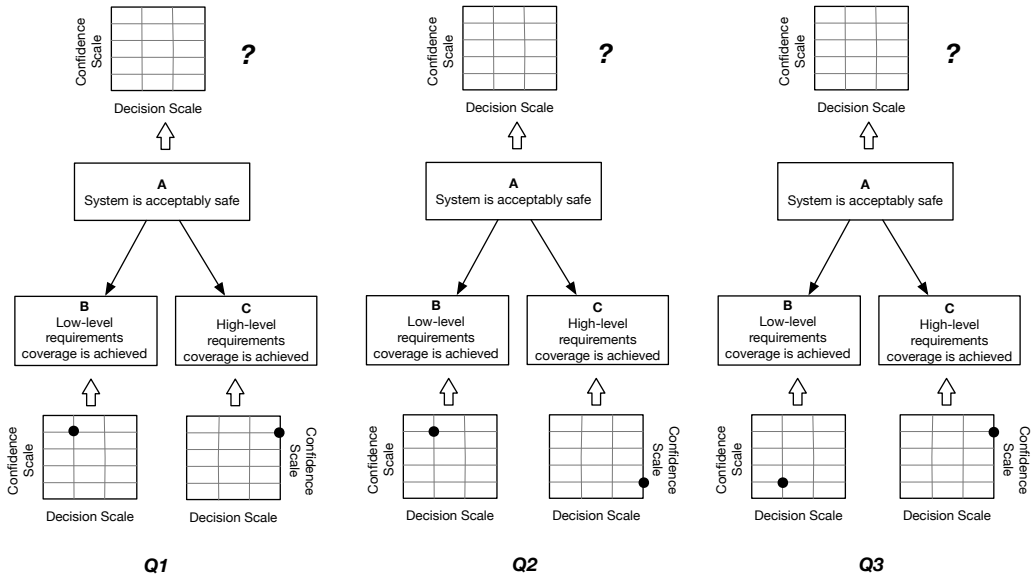


Figure 4.8: Three pairs of inputs for Argument 1

is given at the beginning of the questionnaire; and explanations and assumptions of the 4 arguments are also provided. Furthermore, an extra question follows each argument asking respondents for their understanding degree of the argument. The degrees are "to great extent", "somewhat", "very little" and "not at all". This question is used to remove those less valued answers due to the lack of understanding of certain arguments. The complete version of this questionnaire is presented in Appendix A.

35 experts answered this questionnaire: system safety engineers, safety managers, other engineers of critical system fields, and researchers from the system dependability domain; 2/3 of the respondents are from the railway domain.

4.5 Results and analysis of the expert judgement

The case study aims to estimate the weights and argument types of sub-goals implicitly considered by the experts. In this section, we are going to analyse the collected data (*expert data*) in two successive steps. The first step focuses on a graphical analysis. In Section 3.4, the results calculated based on the proposed framework (*theoretical data*) are presented in the evaluation matrix. Here, we are going to compare them with the expert data. In this step, some rough information can be extracted from the expert data, such as the preliminary judgement of the validity of the proposed assessment framework, the argument types, the relative

importance of the weights among sub-goals. The second step aims to derive more accurate values of the parameters under estimation with a statistical method.

4.5.1 Graphical analysis

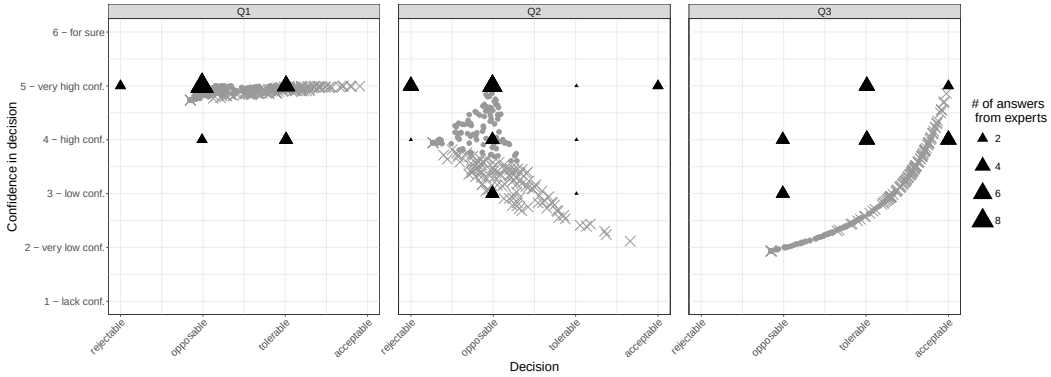
In this subsection, we present the results of the graphical analysis of the collected data. Some calculations for the parameter estimation are used to have a rough idea of the argument features. Another statistical analysis provides more reliable values of these estimated parameters in next subsection.

In Figure 4.9, we present the theoretical data for the 3 questions (Q1-Q3) for Argument 1 as an example. They are derived from the corresponding inputs presented in Figure 4.8 and the possible weights of sub-goals B and C ($w_B, w_C, w_B+w_C \in [0,1]$); and the calculation are realised based on the aggregation rules in Table 2.5 (complementary argument), Table 2.6 (redundant argument) and the Definition 3.2.2. The calculating results are plotted with grey dots ($w_B > w_C$) and crosses ($w_B \leq w_C$). The behaviours of two aggregation rules have opposite trends. The complementary rule trends to produce "negative" results as the grey cluster locates towards *rejectable*; whereas the redundant rule is subject to more "positive" results, as the grey cluster locates towards *acceptable*.

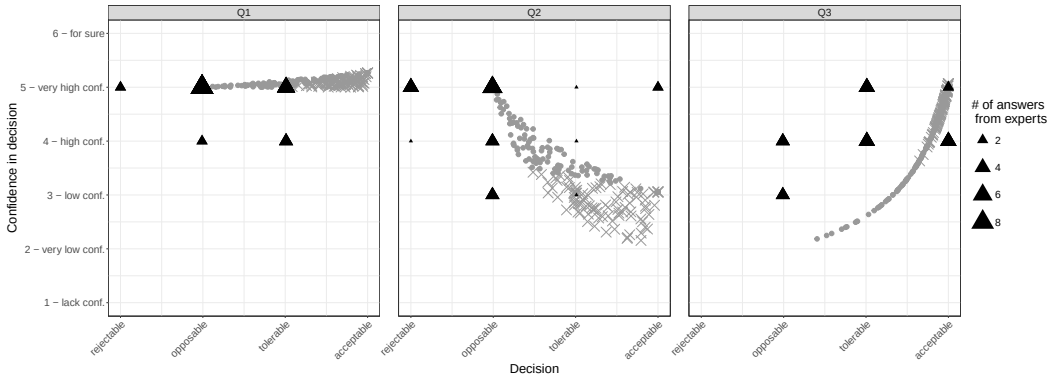
Concerning the expert data, the results collected from the questionnaire are preprocessed to remove the outliers (such as answers with the understanding degree of *"not at all"*). Then, they are plotted with triangles in the evaluation matrix (Figure 4.9) together with the theoretical data. The size of each triangle indicates the number of respondents for the corresponding opinion.

From the observation, the expert data is consistent with the theoretical data of Q1 and Q2 of the complementary argument. We therefore infer that experts have potential preference over the argument parameters. Concerning the argument types, their decisions trend towards *opposable* similar to the theoretical data of the complementary argument rather than redundant argument. Due to the over concentrated distribution of the theoretical data for Q3, no more information can be deduced in this step.

Specifically, we think that the experts' answers located exactly in the cluster of complementary argument ("overlapped" answers) and the ones at the left side of the cluster ("negative" answers) can be interpreted as the preference for the complementary argument. Similarly, the experts' answers situating exactly in the cluster of redundant argument ("overlapped" answers) and the ones at the right side of the cloud ("positive" answers) are considered as the preference for the redundant



a. Comparison between expert data and theoretical data for complementary argument



b. Comparison between expert data and theoretical data for redundant argument

Figure 4.9: Experts responses of Argument 1 and theoretical data

Table 4.3: Calculation results for rudimentary parameter estimation

Arg.	Mean of dec_A	Mean of $conf_A$	Complementary				Redundant				Validated arg. types
			w_B	w_C	$w_{B\times C\to A}$	Type preference	w_B	w_C	$w_{B+C\to A}$	Type preference	
Arg1	0.36	0.68	0.8	0.2	0.0	81.8%	0.8	0.2	0.0	77.3%	C-Arg.
Arg2	0.41	0.66	-	-	-	82.6%	0.7	0.2	0.1	82.6%	R-Arg.
Arg3	0.33	0.68	0.7	0.1	0.2	83.3%	-	-	-	70.8%	C-Arg.
Arg4	0.36	0.49	0.3	0.4	0.3	88.0%	-	-	-	64.0%	C-Arg.

argument. We calculate the rate over the total number of the valid answers (see Equation 4.1), and propose to consider this rate as the *type preference* of the experts.

$$type\ preference = \begin{cases} \frac{N_{overlapped}+N_{negative}}{N_{total\ \ answer}} & for\ complementary\ argument \\ \frac{N_{overlapped}+N_{positive}}{N_{total\ \ answer}} & for\ redundant\ argument \end{cases} \tag{4.1}$$

As the expert data for Q2 distribute relatively scattered, it is easier to identify the overlapped answers. The lesson learned for the choice of input values is discussed in Section 4.5.3.1. Taking the example of the answers for Q2, the calculation results for the *type preference* are presented in the Table 4.3.

Regarding w_B and w_C, we propose to use the mean values of experts data to have a first look at the expert opinion on the weights. Continuing using the data for Q2, the mean values of experts *decision* (dec_A) and *confidence in this decision* ($conf_A$) are calculated (see Table 4.3). Then, we can deduce the corresponding weights based on our proposed framework. The used formulas are recalled in Equations 4.2 and Table 4.4. We assume that $v = 1$. The dash (-) in the table are the solutions not satisfying the constraint: $w_B, w_C, w_B + w_C \in [0, 1]$, which indicates the mean values are not in the distribution cluster of the corresponding argument type.

$$\begin{cases} bel_A = conf_A * dec_A \\ disb_A = conf_A * (1 - dec_A) \\ uncer_A = 1 - bel_A - disb_A \end{cases} \tag{4.2}$$

The deduction of the argument type is based on the weight values and the type preference. For Argument 1, the w_B and w_C have valid values based on both complementary and redundant arguments. As introduced in Section 2.4, $w_{B\times C\to A}$

Table 4.4: Recall of the aggregation rules for the double-node arguments

Types	Aggregation rules
C-Arg	$\begin{cases} bel_A & = [bel_B \cdot w_B + bel_C \cdot w_C + bel_B \cdot bel_C(1 - w_B - w_C)]v \\ disb_A & = \{disb_B \cdot w_B + disb_C \cdot w_C + [1 - (1 - disb_B)(1 - disb_C)](1 - w_B - w_C)\}v \\ uncer_A & = 1 - bel_A - disb_A \end{cases}$
R-Arg	$\begin{cases} bel_A = \{bel_B \cdot w_B + bel_C \cdot w_C + [1 - (1 - bel_B)(1 - bel_C)](1 - w_B - w_C)\}v \\ disb_A = [disb_B \cdot w_B + disb_C \cdot w_C + disb_B \cdot disb_C(1 - w_B - w_C)]v \\ uncer_A = 1 - bel_A - disb_A \end{cases}$

represents the degree of the complementarity or redundancy of an argument. Especially, when $w_{B\times C\to A} = 0$, the argument is a disparate argument (a special case for both complementary and redundant arguments). But the type preference for complementary argument is higher than redundant argument. Thus, we conclude in this step that, from their answers, the experts express that this argument is a complementary argument. The validated argument types for other arguments are also presented in the Table 4.3. Although this is a naive trial to estimate the parameters, the obtained results are, to a great extent, consistent with our expectations.

The confidence assessment and decision-making are believed to be subjective. However, the collected answers from experts are more gathered than we expected. It implies that the experts have some degree of consensus on the rationale of safety justification based on arguments and safety evidence. More precisely, they agree with the variation of the contributions by different techniques or sub-goals to the top goal and also the way that the sub-goals support the top goal. The results derived by the comparison between two sources of data appear reasonable, which can be regarded as a first validation of our assessment framework. Furthermore, based on the above analysis of the survey data, we realise the parameter estimation of the 4 argument examples including argument types and the contributing weights.

4.5.2 Statistical analysis

As mentioned at the beginning of this section, the second step of the parameter estimation will be implemented with the statistical approach. This step is based on the results of the graphical analysis. It aims to obtain more accurate values of the parameters under estimation. The method of least square is adopted. It is usually

used to perform a regression analysis to find the overall solution which minimizes the sum of the squares of the residuals. We explain the parameter estimation method first with Argument 1, and then the results for four arguments are presented.

4.5.2.1 Parameter estimation: an example with Argument 1

Let us take the example of Argument 1 again to illustrate the parameter estimation. Based on the results of the preliminary analysis in the previous section, we assume that Argument 1 is a complementary argument. We recall the aggregation rule of belief function for a double-node complementary argument:

$$bel_A = [bel_B \cdot w_B + bel_C \cdot w_C + bel_B \cdot bel_C(1 - w_B - w_C)]v \tag{4.3}$$

where bel_A (*belief in A*) is the response, bel_B and bel_C are the predictors, and w_B and w_C are the parameters to be estimated. In order to conduct the parameter estimation, we reformulate the function (see below). The introduced notation $f(\mathbf{bel_X}, \mathbf{w_X})$ is the mean function, and ε is the error. In the mean function $f(\mathbf{bel_X}, \mathbf{w_X})$, $\mathbf{bel_X}$ and $\mathbf{w_X}$ are the vectors of the predictors and the parameters to be estimated, respectively. Here, $\mathbf{bel_X} = (bel_B, bel_C)$, and $\mathbf{w_X} = (w_B, w_C)$.

$$\begin{aligned} bel_A =& f(\mathbf{bel_X}, \mathbf{w_X}) + \varepsilon \\ =& [bel_B \cdot w_B + bel_C \cdot w_C + bel_B \cdot bel_C(1 - w_B - w_C)]v + \varepsilon \end{aligned} \tag{4.4}$$

As this is a nonlinear function, the parameter estimation should be implemented via the nonlinear least square. The principle is to minimise the residual sum of squares:

$$s(\mathbf{w_X}) = \sum [bel_A - f(\mathbf{bel_X}, \mathbf{w_X})]^2 \tag{4.5}$$

The notation $\hat{\mathbf{w_X}}$ will be used for the expected values, which renders the minimised residual sum of squares.

Conveniently, we use to *nls* (*nonlinear least square*) function in **R** [Fox and Weisberg, 2011]. The starting values for the parameters are required at the beginning of the estimation. The preliminary estimated values of w_B and w_C in Table 4.3 are set as the starting values. For the default values ("-"), $w_B = w_C = 0$ will be applicable.

The aggregation rule describes the relation among bel_A, bel_B, and bel_C.They are all calculated according to the Equation 3.7. bel_A comes from the collected expert data of 3 questions. bel_B and bel_C are derived from the input. The estimation and statistical test results for Argument 1 are copied from the console from **R**:

```
Formula: bel_A ~ w_B * bel_B + w_C * bel_C + (1 - w_B - w_C) * bel_B *
bel_C
Parameters:
Estimate   Std. Error   t value   Pr(>|t|)
wB 0.53179    0.23501    2.263    0.0273 *
wC 0.36833    0.05562    6.622    1.1e-08 ***
—
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.2253 on 60 degrees of freedom
Number of iterations to convergence: 1
Achieved convergence tolerance: 1.651e-08
```

Looking at the *Parameters* part, the estimates of w_B and w_C are: $w_B = 0.53179$ and $w_C = 0.36833$. It indicates that the experts consider the contributing weight of B (*low-level requirements coverage is achieved*) is a little higher than C (*high-level requirements coverage is achieved*). The degree of complementarity is $w_{B\times C\to A} = 1 - w_B - w_C = 0.0999$. *T value* and *Pr(>|t|)* (p-value) are the results of the T-test. The significance level are given in the *Signif. codes*, where '***' to ' ' mean *extreme significant* to *not significant* in terms of statistics. The *significant* (*) and *extreme significant* (***) levels of the estimates indicate that the estimated parameters are relatively trustable. The *residual standard error* reflects the large variance of the expert data.

Since the estimated values of the parameters are obtained, we can use the parametric framework to predict the propagation results. Based on the same input, the decision of top goal and the confidence in the decision are calculated and plotted in the *evaluation matrix* (see Figure 4.10). The opinions for Q1 and Q2 are within the range of expert data; and the opinion for Q3 is more pessimistic than the experts' responses.

4.5.2.2 Estimation results for four argument fragments

Via the same strategy, we repeat the parameter estimation for each argument fragment. Out of rigorous consideration, these estimations are implemented for the aggregation rules of both complementary and redundant arguments. The results are presented in Table 4.5, which include the contributing weight (w_B, w_C), the complementary/redundant degree $(w_{B\times C\to A}/w_{B+C\to A})$ and the argument types.

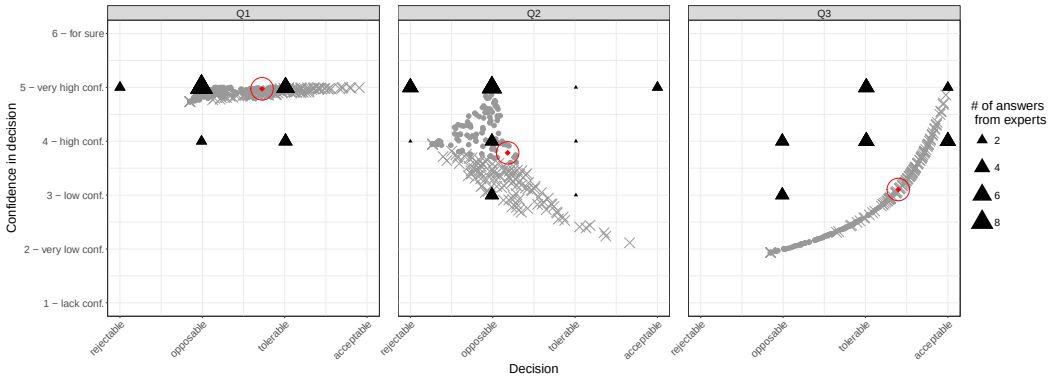


Figure 4.10: Predicting results by the parametric model

Table 4.5: Estimation results for 4 argument fragments via statistical method

Arg.	Complementary			Redundant			Validated arg. types
	w_B	w_C	$w_{BxC->A}$	w_B	w_C	$w_{B+C->A}$	
Arg1	0.5318	0.3683	0.0999	0.6317	0.4682	-0.0999	C-Arg.
Arg2	0.5103	0.2354	0.2542	0.7646	0.4897	-0.2542	C-Arg.
Arg3	0.3700	0.3173	0.3128	0.6827	0.6300	-0.3128	C-Arg.
Arg4	0.1051*	0.5475	0.3474	0.4525	0.8949	-0.3475	C-Arg.

According the definition of appropriateness, we have the constraints: $w_B, w_C \in [0, 1]$ and $w_{B\times C\to A}/w_{B+C\to A} = 1 - w_B - w_C \geq 0$. In Table 4.5, all $w_{B\times C\to A} > 0$ and $w_{B+C\to A} < 0$. Therefore, four arguments appear to be complementary, because none of the estimates for the redundant arguments is compliant with the constraints. For Arg2, the argument type is different from our expectation. It means that the experts believe the *formal verification* and *test* are both essential; and they do not have redundancy in terms of the safety justification. The value of contributing weights reflect the implicit judgements of the respondents on the sub-goals and their relationship. For Arg1-Arg3, the sub-goal B is deemed more important to varying extent; however, the weight of B for Arg4 is lower than the weight of C, which does make sense regarding the initial set of sub-goals: minor/critical hazards have been addressed. Among these arguments, the estimated degree of complementarity of Arg3 is the highest. This implies that *"test process"* and *"test results"* are considered as the closest example to a complementary argument.

As discussed in Section 3.4, we assume that the experts under investigation have preferences for the features of a given argument (types and weights). Compared with the estimation in the previous section, the least square method is more appropriate

for the parameter estimation. This is because the w_B and w_C in Table 4.3 are calculated based on the mean values (dec and $conf$) of expert data; and the least square method minimises the residual errors between all answers and the estimated w_B and w_C. It means that these results are closer to the consensus of the expert judgements. Nevertheless, the variance among the experts judgements and the residual error should not be ignored. Therefore, in the next subsection, we are going to discuss how to increase the accuracy of the estimated values and to further validate our proposed framework in more general ways.

4.5.3 Discussion

During this case study, especially in the process of the parameter estimation, we have got some valuable experience to share. It mainly relates to increasing the accuracy of parameter estimation (Section 4.5.3.1), further work for more general validation of the proposed framework (Section 4.5.3.2) and some consideration for confidence extraction from safety evidence (Section 4.5.3.3).

4.5.3.1 Better inputs for parameter estimation

According to our experience obtained in the case study, not all the initial inputs of the assessment model are efficient for the parameter determination. In the graphical analysis, the theoretical data are concentrated in a limited area (see Figure 4.10 Q3), which makes it difficult compare the expert opinions and the theoretical data. Here, some tips for choosing the pre-determined judgements are provided. To do so, we divide the evaluation matrix into 3 zones (see Figure 4.11). Indicated by the Jøsang triangle, these 3 zones correspond to ① - *"belief"*, ② - *"uncertainty"* and ③ - *"disbelief"*. It is important to provide conflicting opinions on sub-goals of an argument to increase the information conveyed in the expert judgements. Thus, we have more possibility to deduce what the experts think about the different contributions of sub-goals and their potential preference for the argument types. In Figure 4.11, the conflicts mean that the opinions on sub-goals are positioned in different areas.

For both argument fragments in Figure 4.12, if the initial opinions on sub-goals B and C are positioned in one zone, the expert probably wants to provide the judgements of the top goal in the same zone. Thus, little information will be deduced. For instance, if the initial opinions for B and C are ($opposable$, $very\ high\ confidence$) (Zone ③), the expert judgements would be the same with the inputs or closer to the lower left corner of the evaluation matrix.

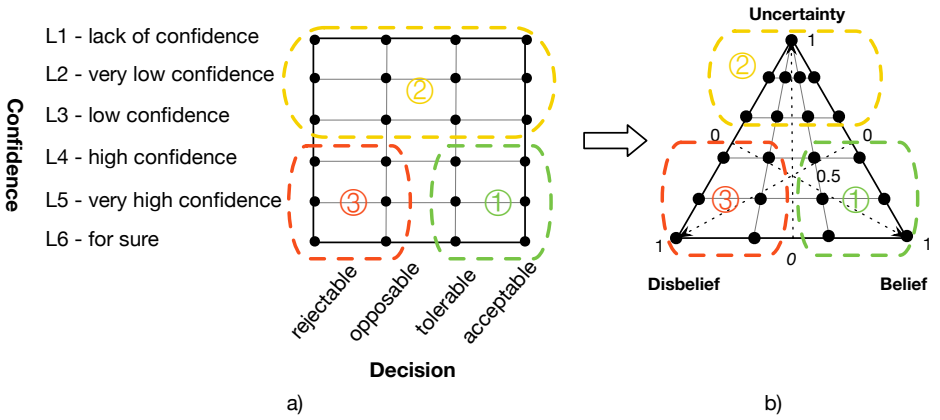


Figure 4.11: Division of the evaluation matrix

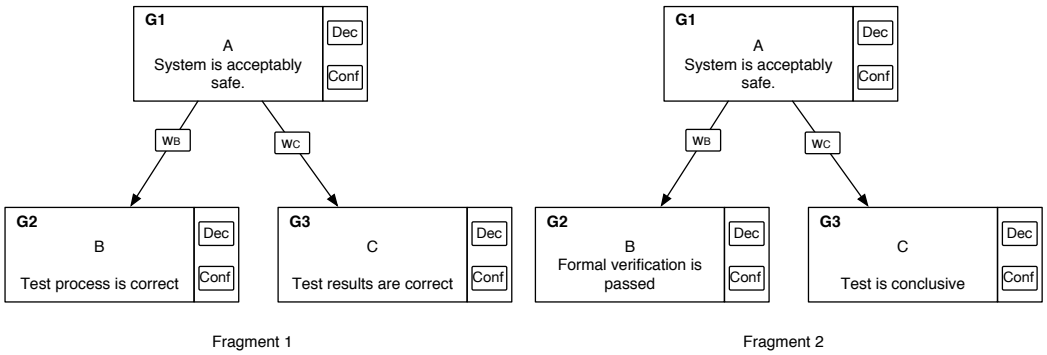


Figure 4.12: Two simple argument fragments

However, if the opinion of B is (*opposable*, *very high confidence*) (Zone ③), and opinion of C is (*acceptable*, *very high confidence*) (Zone ①). In this situation, the decision-making process for top goal A requires more consideration of the argument types and the contributing weight of each sub-goal. Thus, more information about the argument parameters might be expressed in the expert data.

Besides, it would be better to provide more sets of inputs for one argument, because the parameter estimation based on statistical method works well for the response to the case with many sets of inputs. This will also help to extract the respondent's judgement accurately.

4.5.3.2 Further work to validate the approach

Modelling safety arguments and its assessment are strongly based on expertise in the relating domain. Our proposed approach aims to help to build parametric safety argument patterns and make the safety assessment work more effective and

efficient. Nevertheless, this novel confidence assessment framework needs extensive validation. The capture of expert judgements and result analysis are implemented with the simplified safety argument fragments with two nodes. Thus, the further validation should be implemented for the general argument structure. The general aggregation rules have been already introduced in Table 2.9. Thus, it is possible to launch a case study based on a safety case of one real critical system.

Besides, once a parametrerised safety case model is built, we can also compare the results from the assessor and this assessment model. This can be regarded as another way to assure the validity of the proposed assessment framework. Indeed, this kind of case study will cost a lot of extra resources. However, these reusable arguments or argument patterns are of great importance to increase the efficiency of similar systems.

4.5.3.3 Considerations for confidence extraction from safety evidence

In Chapter 3, we introduce an approach to extract expert judgements for the acceptance of one goal and the confidence in this decision $(Dec, Conf)$. This approach is similar to several scientific paper reviewing systems, which is very practical to capture and quantify human opinions. Nevertheless, at the same time, the propagation model based on D-S theory is based on quantitative aggregation rules. The decoding and encoding processes for the semi-quantitative method will inevitably bring in uncertainty due to the loss of information. Therefore, it would be better that the trustworthiness of the evidence can be directly quantified. Some types of evidence are illustrated by the numerical key indicators, for instance, test coverage, the number of defects found, the percentage of traceability, unclosed hazards, etc. If these evidence information can be transferred to the trustworthiness of the related sub-goal, the final estimation of the top goal will be more accurate.

4.6 Guidance on the application of the framework

In this section, we present the procedure of applying the proposed framework to an example: the Wheel Slide Protection (WSP) system. This simplified example only aims to run through the assessment process illustrated in Figure 3.1 for a real railway subsystem. Most results of real systems are actually confidential, but this example has been developed with safety experts in the railway domain.

As presented and studied in several works [Pugi et al., 2006; Allotta et al., 2013], the WSP system is used to detect the wheel sliding during braking and to

prevent it by applying periodic braking releases. This system includes an Electronic Control Unit (ECU) with the embedded software, speed sensors (tachometers) and pneumatic valves. The ECU receives the angular axle speed and braking torque from sensors. Then, it calculates the linear velocity and acceleration in order to detect the sliding state. If the state is *sliding*, ECU sends the command to implement the periodic braking release, that is, to return the torque of pneumatic braking to 0.

As the WSP system can modify the braking torque, its functions are highly safety-related. For instance, while in the degraded adhesion conditions (e.g., leaves or snow on the tracks), a failure of the ECU software may untimely release the periodic braking. Then, the longer braking distance resulted from a macro sliding would lead to the Signal Passed At Danger (SPAD) or even collision. Thus, one safety requirement should be *"SR1: untimely activation of Periodic Braking Release function should be avoided"*. Within this context, the confidence assessment framework is applied in the following steps (shown Figure 3.1):

Step 1): Safety case modelling. In order to justify that the embedded software in ECU is free from faults leading to this failure, we need to verify and validate the correctness of the software against the safety requirements. Assuming that the formal verification and functional testing are sufficient for the justification, the safety evidence and arguments are illustrated in the GSN model in Figure 4.13. The top goal *WSP-G1* considers only one safety requirement SR1. It is ensured by the formal verification (*WSP-G2*) and functional testing (*WSP-G3*). The goal *WSP-G3* is broken down into *Testing procedure is correct* (*WSP-G4*) and *Testing results are correct* (*WSP-G5*).

Step 2): Estimation of weights and argument types. This safety argument fragment shall be divided into two parts in order to consider the weights and argument types. These two parts correspond to the Argument 2 & 3 in Figure 4.7.

- Argument 2: *WSP-G2* and *WSP-G3* support *WSP-G1*
- Argument 3: *WSP-G4* and *WSP-G5* support *WSP-G3*

The argument types and weights are estimated in the previous section (shown in Table 4.5). Thus, we directly use the results and illustrate the parameters related to WSP safety argument in Table 4.6.

Step 3) & Step 4): Confidence assessment, aggregation and decision. As this example illustrates a guidance for using our approach rather than an industrial case study, a sensitivity analysis is implemented to present the impacts of the confidence in low-level arguments on the high-level argument.

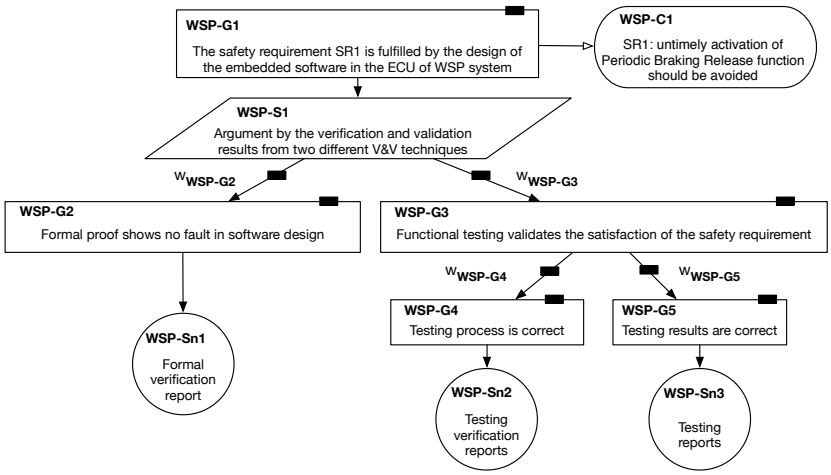


Figure 4.13: Safety argument fragment of WSP system

Table 4.6: Argument types and weights for safety argument of WSP system

Argument	Type	Weights	
Argument 2	Complementary	w_{WSP-G2} 0.5103	w_{WSP-G3} 0.2354
Argument 3	Complementary	w_{WSP-G4} 0.4138	w_{WSP-G5} 0.2808

We suggest performing a sensitivity analysis using a tornado graph as presented in Figure 4.14. It is a simple statistical tool, which shows the positive or negative influence of basic elements on a main function. In our case, we estimate the decision and the confidence in this decision $(Dec, Conf)$ of *WSP-G1*, with corresponding intervals for $(Dec, Conf)$ of *G2, G4, G5*. The basic values (X_i) and intervals $[X_{min}, X_{max}]$ for each parameter are shown in Table 4.7. The basic values (X_i) are given arbitrarily as a reference value. The intervals $[X_{min}, X_{max}]$ are deduced according to the limit values of $(Dec, Conf)$. The lower value of confidence is set at 0.1 to ensure the scale of $(Dec, Conf)$ to the great extent for a better presentation of the tornado graphs.

Table 4.7: Example of values and intervals for sensitivity analysis

	$Dec_{WSP-G2/G4/G5}$	$Conf_{WSP-G2/G4/G5}$
Basic value X_i	0.6667 (Tolerable)	0.6 (High confidence)
Intervals $[X_{min}, X_{max}]$	[0,1]	[0.0,1]

With the basic values above, $(Dec_{WSP-G1}, Conf_{WSP-G1}) = (0.5452, 0.5764)$. These two values are set as the positions of vertical axis in the tornado graphs. To determine the sensitivity to one sub-goal, for instance WSP-G2, we keep all the

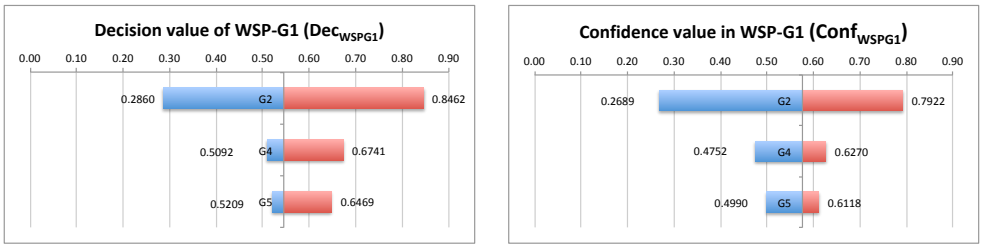


Figure 4.14: Tornado diagram of confidence assessment for WSP safety argument

values ($Dec_{WSP-G4/G5}, Conf_{WSP-G4/G5}$) as the basic values. Then, we calculate the values for WSP-G1 with ($Dec_{WSP-G2}, Conf_{WSP-G2}$)= (0,0.1) or (1,1). We obtain the values $Dec_{WSP-G1} = [0.2860, 0.8462]$ for decision value of WSP-G1 and $Conf_{WSP-G1} = [0.2689, 0.7922]$ for confidence value in this decision. The same approach is applied for other parameters; the results are presented in Figure 4.14.

Both graph shows that the sub-goal WSP-G2 has the most influence on the top goal either for positive or negative impact on the decision of acceptance of G1. A complete case of such an analysis is not presented here, but basic activities using this sensitivity analysis can be done, such as: identification of the weakness in the argument, analysis of additional new evidence and estimation of their impacts on the confidence of top goal, comparison between several arguments. This is actually out of the scope of this section, as it is an exploitation of the quantitative framework that we propose.

4.7 Conclusion

In this chapter, we carry out a case study mainly to implement the parameter estimation based on the confidence assessment framework. Firstly, we study the safety-related standard in railway domain and find out the links between the recommended techniques and high-level argument structure. Accordingly, a fragment of the general safety case for electronic railway system is established. Secondly, the application of the argument assessment framework is carried out via a survey amongst experts including safety engineers, safety assessors and researchers in the domain of dependability and railway safety. This application aims to estimate the parameters of assessment models, which is an essential step for applying our confidence assessment framework. 35 responses to the survey have been received. We compare the collected answers (expert data) with the theoretical data derived from the aggregation rules. The consensus among experts is discovered on the variation of the contributing weights of sub-goals to the top goal and also the argument types.

Then, we use a statistical approach to further the parameter estimation. Based on the method of nonlinear least square and statistical hypothesis test, more accurate parameters for the assessment model are obtained. This gives a first validation of the proposed confidence assessment framework. Some other approaches to increase the parameter accuracy and more general way to confirm the validity of the proposed framework are discussed. Finally, we provide guidance on the application of this confidence assessment framework with a simplified example of WSP system.

However, the argument examples used in the case study are simple argument patterns with two premises. For a complicated safety case in railway domain, considerable argument branches are involved. The aggregation rules for an n-node argument have been already introduced in Table 2.9. Further validation is needed for a general argument with the hierarchical structure and multiple sub-goals. Concerning the parameter estimation, the experience gained in the case study is summarised. We realise that the choices of the input values of *Dec, Conf* have an impact on the results. More studies are essential to estimate this impact, which asks for more expert judgements of various arguments. Moreover, in our application, we consider temporarily that $v = 1$. Thus, how to determine the value of v is still an open issue.

Conclusion

In this thesis, we studied the issue on the justification of the safety assurance for critical systems via evidence-based approaches. In particular, we focused on the quantitative assessment of the confidence in safety arguments.

We conclude the work of the thesis in this chapter by summarising the different aspects addressed by each chapter, reminding the main contributions, discussing the lessons learned and open issues, and our perspectives for future research.

Summary

We studied the approaches to justify the confidence in the safety assurance. Evidence-based structured arguments are recommended or even required by the industrial standards. Nevertheless, an enormous amount of technical and managerial evidence and arguments leave plenty of uncertainties hampering the effective decision-making. Our work was motivated by this argument evaluation issue. To solve it, we decide to measure these uncertainties quantitatively and facilitate the argument assessment process.

In the first chapter, we explored two scientific fields connected with our research issue. It aimed to get to know the relating knowledge 1) the safety argument, notably including its relationship with various functional safety standards and the structuring approaches; and 2) multiple uncertainty theories with the focus on belief function theory. Afterwards, we critically studied the existing work on confidence assessment (via uncertainty theories) for the safety argument. As a novel research area, we noted that there were still limitations related to the aspects, such as confidence factors definition, aggregation rules, expert judgement extraction and parameter estimation. We conclude that our contribution should address these open issues.

In the second chapter, we identify the factors influencing the confidence in an argument; then they are formally defined as *trustworthiness:* $trust = (bel, uncer, disb)$ and *appropriateness:* $appr = (w_i, w_{TYPE}, v)$ based on the D-S theory. Then, these definitions are further specified according to different argument structures (simple argument and multi-node argument) and different inference types (complementary and redundant). Corresponding confidence aggregation rules are developed, and they are finally generalised into the aggregation rules of n-node arguments. Taking

the double-node argument as an example, we carry out a sensitivity analysis of the aggregation rules. The analysis results reveal the impacts of different parameters.

In the third chapter, we proposed a 4-step confidence assessment framework for the safety case of a critical system. The four steps of this framework are 1) building the structured safety case, 2) estimating the parameters, 3) assessing confidence in sub-goals, and 4) confidence aggregation and decision-making. This work was based on the quantitative model of the confidence assessment for the safety argument proposed in the previous chapter. We integrated a judgement extraction method to obtain the *trustworthiness* of a sub-goal from the *decision* on this sub-goal and the *confidence* in the decision of an expert. We also elaborated the principles to estimate the *appropriateness* parameters of sub-goals. Another sensitivity analysis was carried out to identify the behaviours of the new integrated framework. Furthermore, we opened up an additional discussion about the treatment of the contextual elements in GSN.

In the fourth chapter, we implemented a case study mainly to realise the parameter estimation by applying the confidence assessment framework; other subjects around this proposed framework were also explored. Firstly, we studied the safety-related standard in railway domain and found out the links between the recommended techniques and high-level argument structure. Accordingly, a fragment of the general safety case for railway electronic system was established. Secondly, the application of the argument assessment framework was carried out via a survey amongst experts including safety engineers, safety assessors and researchers in the domain of dependability. This application aimed to estimate the parameters of assessment models, which was an essential step for applying our confidence assessment framework. 35 responses to the survey had been received. We analysed the collected data in the graphical and the statistical methods. We obtained the estimated parameters for the assessment model. This gave a first validation of the proposed confidence assessment framework. Some other approaches to increase the parameter accuracy and more general way to confirm the validity of the proposed framework were discussed. Finally, we provided guidance on the application of this confidence assessment framework, illustrated with a simplified example of the WSP system.

Main contributions

We summarise the principal contributions of our research work. These contributions, categorised by different subjects, are presented in this following list:

- Study on the safety argumentation in the safety related standards;

  According to our study, the evidence-based argument is the most common way to present the justification of system dependability attributes, such as safety. Various standards (ISO26262 [2011] for automotive, EN50129 [2003] for railway, ISO/IEC15026-1 [2013] for software assurance, etc.) provide the definitions of safety arguments, usually named safety cases (see Section 1.2.2). Notably, we study the standard EN50129 [2003] (see Section 4.2) and made explicit the rationale by building the goal-based safety arguments for railway signalling system (see Section 4.3).

- Formal definition of confidence and new aggregations rules based on D-S theory

  We formally define two confidence measures in an argument: *trustworthiness* and *appropriateness* of premises. In the definition of the *appropriateness*, two argument types are considered: *complementary* and *redundant* arguments. These two general types can be extended to five sub-types, which cover all the categorization of argument types in related works. Based on these definitions, we develop the aggregation rules to propagate the confidence consistently using the D-S theory. The aggregation rules are extended for an n-node argument. This mathematical model can be also regarded as a general approach to merge uncertain data from difference sources.

- Proposition of an integrated confidence assessment framework for the safety case of a critical system;

  We put forward a four-step integrated framework based on our proposed mathematical model for confidence propagation. It is an application of the model to assess the confidence in the safety case. To improve the assessment process, we integrate the necessary steps, such as the preparation of safety arguments, parameter estimation and expert judgement extraction. It makes the model more practical for a real engineering application. An assessor only needs to provide his/her *decision* and the *confidence* in the decision for the lowest level of sub-goals of the argument based on the available evidence. Then, these opinions will automatically be combined to generate the *decision* and the *confidence* in the decision of the top goal.

Open issues/limitations

This research work belongs to a relatively novel subject combining the quantitative uncertainty assessment with subjective argument reasoning. The proposed assessment framework overcomes the limitations identified, regarding the definition and aggregation of the confidence measures, subjective expert judgement extraction, parameter estimation, etc. Nonetheless, several open issues remain to be further studied.

- Further validation

  As we mentioned in Chapter 4, this framework is not yet an off-the-shelf approach to assess the safety of a system. More means for robust validation are needed. The capture of expert judgements and result analysis are carried out with the simplified safety argument fragments with two nodes. Thus, the further validation should be implemented for a general argument with the hierarchical structure and multiple sub-goals.

  Besides, once a parametric safety case model is built, we can also compare the results from a human assessor and this assessment model. This can be regarded as another way to assure the validity of the proposed assessment framework.

- Issues related to judgement extraction

  We introduce a way transform the *trustworthiness* of a goal into *decision* and *confidence* in the decision. The scales are linearly distributed as an initial proposal. For an application of a real safety case, it is essential to calibrate this transformation from numerical values to the ordinary *decision* and *confidence* scales. A possible approach of the calibration is proposed by Cyra and Gorski [2011] based on expert opinions.

- Criteria for decision-making

  A common doubt for quantitative approaches is that no decision can be drawn when only considering the final value of the confidence (e.g., what to decide when confidence is 0.8, or 0.9?). In fact, for the classical probability theory, we also have to confront such issue. Ledinot et al. [2016] point that there are probabilities available for software failure in standards, as the software assurance levels (e.g., SIL, ASIL, DAL, etc.) may correspond to probability objectives. The confidence estimated based on our approaches is similar

to these probabilistic requirements for software. The determination of the confidence criteria supporting the decision-making is an open issue.

- Applicability regarding cost/benefit

  We provide a method to estimate the parameter such as argument types and weights. The determination of these parameters requires expert opinions, and it has to be done for all “gates". It could be of high cost. Nevertheless, to generate some reusable patterns of the parametric arguments might be a choice, which, in turn, increases the efficiency of assessment work for similar systems.

Perspectives

Short-term perspectives

To enrich our work, there are possible improvements and validations of the proposed approach. In the short term, we consider:

- Further validation of the proposed confidence assessment framework. We may employ other statistical methods to validate the parameter estimation results. In addition, more considerations should be given to the confidence aggregation model in the following aspects:
  - Estimation and interpretation of the discounting factor v;
  - Conflicting issue when combining several expert opinions on one sub-goal
- Application to a complete safety argument. Available argument patterns proposed by the safety case community are good candidates to start this further case study.
- Explore new methods to determine the trustworthiness of goals. The expert judgement extraction involving semi-quantitative values will inevitably bring in uncertainties due to the loss of information. Therefore, it would be better that the trustworthiness of the objective evidence can be directly quantified. Some evidence is illustrated by the numerical key indicators, for instance, test coverage, the number of defects found, the percentage of traceability, unclosed hazards, etc. If the evidence information can be transferred to the trustworthiness of the related sub-goal, it would avoid, to a great extent, bringing in new uncertainties to the final estimation of the top goal.

Long-term perspectives

The work presented in this manuscript also opens several research paths. Hence, in the long term, we consider:

- Introduce counter arguments into the confidence assessment framework. Our work is based on GSN argument structuring approach. In GSN model, there is no notation representing the counter arguments. However, this element is quite often used in argumentation, for example the rebuttal in Toulmin's model. Thus, the evaluation of counter arguments and the corresponding confidence aggregation model need further developed.

- Application of the proposed confidence assessment framework on a safety case for a real safety critical system. This is definitely a systematic engineering. It is necessary to the development of an automatic tool to realise the evaluation. In addition, the SACM metamodel would be preferred for the automation of argument assessment.

- Consider a new vision for certification. Currently, we must demonstrate that a "system is acceptable safe" by calculating probabilities (such as $10^{-9}/h$ for ASIL D in automotive, $10^{-9}/h$ for SIL4 in railway, $10^{-5}/h$ for DAL C in aeronautics, etc.) for the hardware. When it turns to software, the process-based means (applying the "best practices") are used to ensure the software dependability. What this thesis brings is that the definition of the acceptability of the risk does not concern the system but the engineering practices through the quantitative assessment of the "Safety Arguments". This could lead to a new vision for certification of the process (confidence in these practices described and evaluated by safety cases) instead of a certification of the product (e.g., a plane).

Appendix A

Questionnaire for argument assessment research

Rigorous argument plays an important role in communicating the attributes of systems among stakeholders. It is recommended or even required by standards for safety critical systems (for instance, ISO 26262 for automotive, EN 50129 for railway, etc.). This questionnaire aims to help obtaining a better understanding of such argument, thanks to your expertise in this domain. [1]

Please make sure to answer the questions in sequence. Questionnaire results will be used anonymously. The conclusion will be sent back to you. This questionnaire takes approximately 20 minutes to be completed.

Please return the scanned copy to *rwang@laas.fr*
by close of play *31 January 2017*

Name _____

Position _____

Company/institution _____

[1] Online version of this questionnaire is available: `https://goo.gl/forms/usaJNq34ObliP9263`

How to play

In order to assess an argument, an assessor needs to evaluate all the elements in the argument, i.e. statement, evidence, etc. In Figure A.1. a), a statement is proposed: *"High-level requirements coverage is achieved"*. We adopt an evaluation matrix to assess this statement by two criteria: the *decision* on the statement and the *confidence* in this decision (this confidence is mainly based on the amount and quality of available evidence). In Figure A.1. c), there are 4 levels for Decision Scale and 6 levels for Confidence Scale. The solid dot represents the opinion of an assessor on this statement. The right one indicate that the assessor *accepts* this statement *with very high confidence*. The decision *"acceptable"* indicates that the assessor believes the high-level requirements were actually covered. Moreover, the *"very high confidence"* comes from abundance of evidence of functional testing, for instance.

In Figure A.1. b), another statement is provided: *"Low-level requirements coverage is achieved"*. The decision on it is *opposable* as the assessor might consider that the conditions for establishing this statement are not met; therefore, he/she weakly rejects the statement. Moreover, the *very high confidence* in this decision might be due to results of structural testing, such as a low coverage rate.

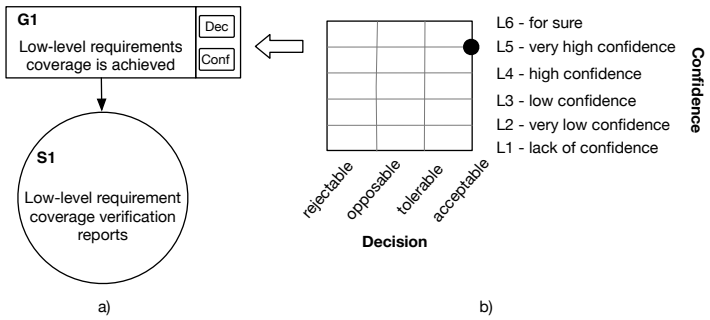


Figure A.1: The evaluation matrix for a statement

The questions in this questionnaire are relevant to arguments composed of statements. We will provide the initial opinions for the sub-statements and let you make a decision on the top statement. For example, in Figure A.2, one assessor has given the *opposable* opinion on Statement B (weak reject) *with very high confidence* and the *acceptable* opinion on Statement C *with very high confidence* based on available evidence. Then, the opinion on Statement A needs your decision.

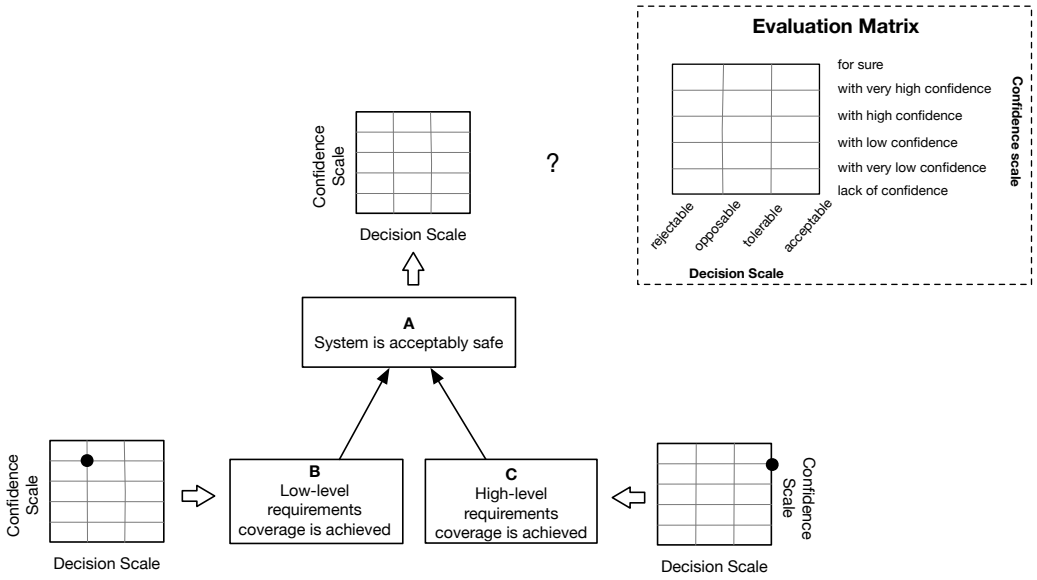


Figure A.2: An argument example

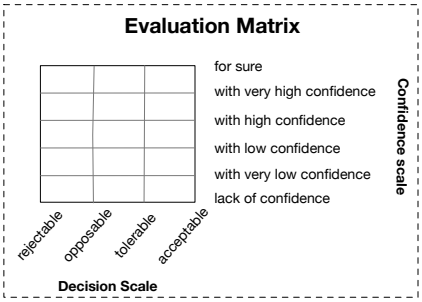
Would you conclude that *The system is acceptably safe* from the opinions of the two sub-statements? Therefore the decision on the top statement is *acceptable*? Or would you consider that its safety can be questioned? Therefore, the top statement is opposable. Then, what is your confidence in your decision? For sure? Very high? Etc.

Argument 1/4

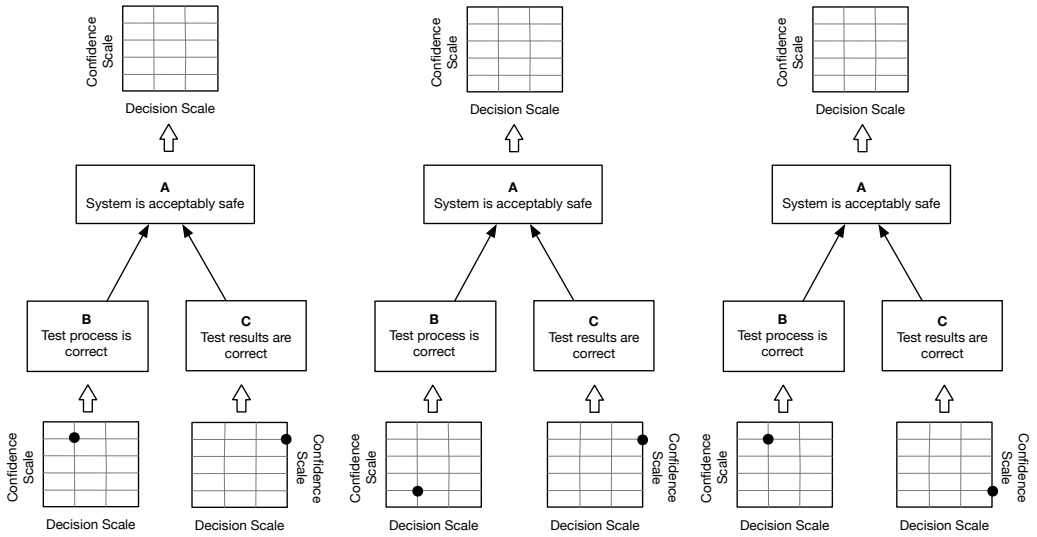
We assume that the system safety can be assured by conclusive testing. Therefore, the system safety can be deduced if “test process is correct" and “test results are correct". Please read the argument below and the initial opinions for Sub-statements B and C. Then, answer the following questions.

- How much do you understand the subject of this argument?

  ☐ To a Great Extent
  ☐ Somewhat
  ☐ Very Little
  ☐ Not at All



- Please give your opinions on Statement A in the three cases below by placing a dot in each blank table over A.



* The opinion on the B implies that based on abundant evidence (“with very high confidence"), assessor weakly rejects (“opposable") that the test procedure is correct; and the opinion on C means the assessor strongly accepts that test results are correct based on abundant evidence (“with very high confidence").

** The opinion on B remains the same with first case*; however, the assessor accepts statement C (“acceptable") based on not enough evidence (“with very low confidence").

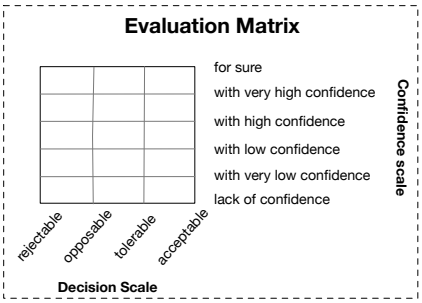
*** Inversely, the “opposable" decision on B ‘test process is correct" is made based on not sufficient evidence (“with very low confidence"); and the opinion on C remains the same with first case *.

Argument 2/4

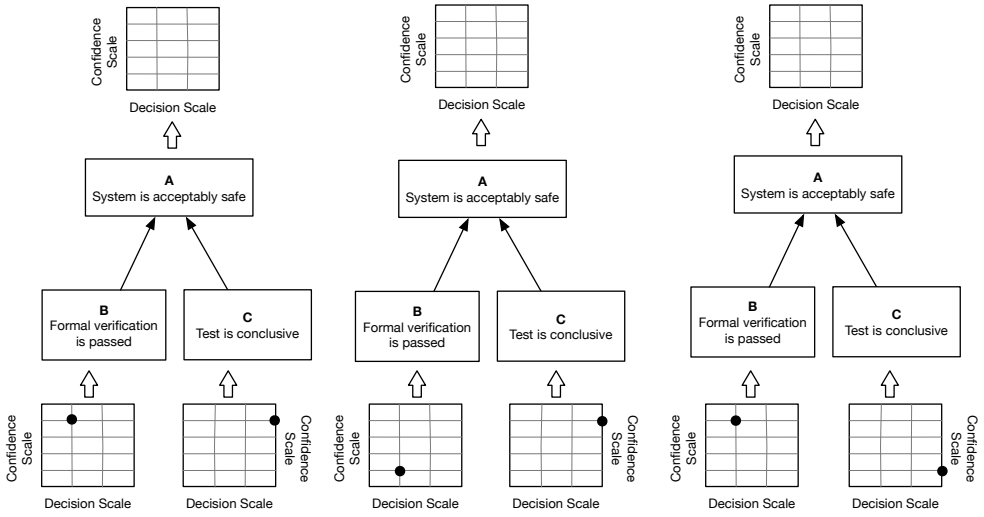
In this part, we assume that safety can be assured by two techniques: formal verification and testing. Therefore, the system safety depends on the results of formal verification and testing. Please read the argument below and the initial opinions for Sub-statements B and C. Then, answer the following questions.

- How much do you understand the subject of this argument?

  ☐ To a Great Extent
  ☐ Somewhat
  ☐ Very Little
  ☐ Not at All



- Please give your opinions on Statement A in the three cases below by placing a dot in each blank table over A.

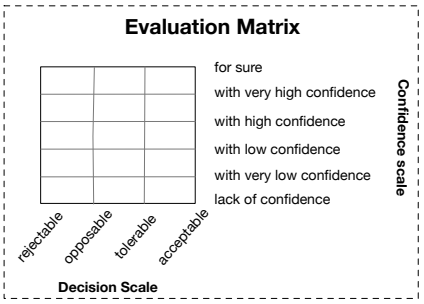


Argument 3/4

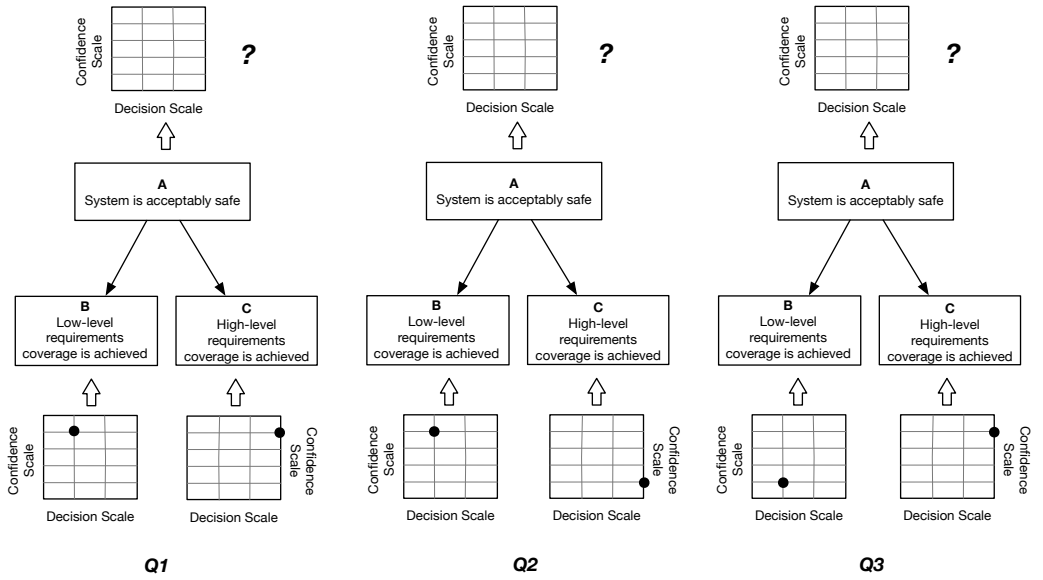
Developing and fulfilling system and safety requirements is a way to ensure the system safety. Therefore, system safety depends on the “low-level requirement coverage" and “high-level requirement coverage". Please read the argument below and the initial opinions for Sub-statements B and C. Then, answer the following questions.

- How much do you understand the subject of this argument?

  ☐ To a Great Extent
  ☐ Somewhat
  ☐ Very Little
  ☐ Not at All



- Please give your opinions on Statement A in the three cases below by placing a dot in each blank table over A.

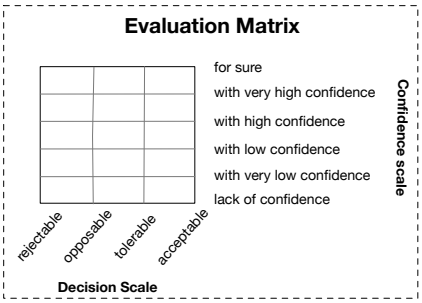


Argument 4/4

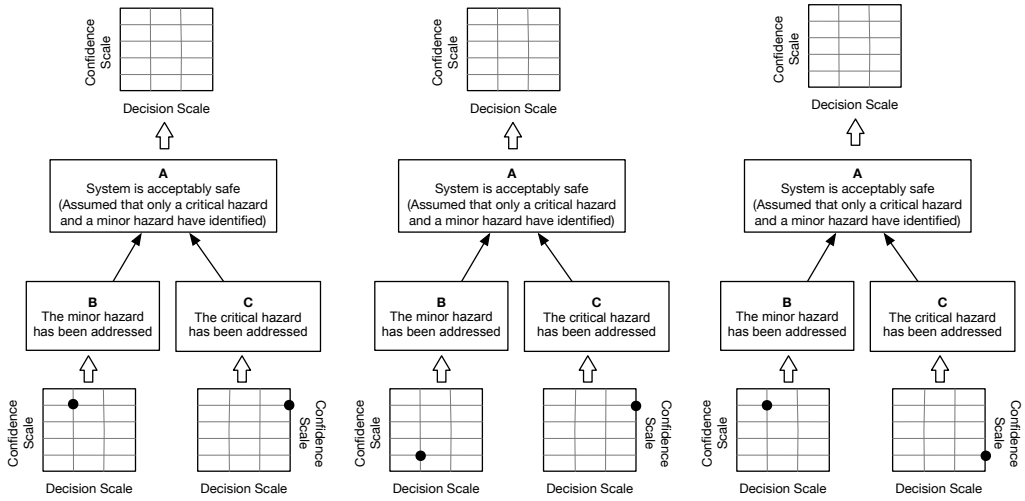
We assumed that only two hazards have been identified in one system. Concerning the risk, one is minor and the other is critical. Please read the argument below and the initial opinions for Sub-statements B and C. Then, answer the following questions.

- How much do you understand the subject of this argument?

  ☐ To a Great Extent
  ☐ Somewhat
  ☐ Very Little
  ☐ Not at All



- Please give your opinions on Statement A in the three cases below by placing a dot in each blank table over A.



Open question

Do the questions above show that based on the same given opinions on the sub-statements, you've given different results for different arguments? If so, in your opinion, why?

Thank you for your answers!

Bibliography

[Adams and Levine 1975] Adams, Ernest W. ; Levine, Howard P.: On the uncertainties transmitted from premises to conclusions in deductive inferences. In: *Synthese* 30 (1975), Nr. 3, S. 429–460 (Cited in page 22.)

[Allotta et al. 2013] Allotta, B ; Conti, R ; Meli, E ; Pugi, L ; Ridolfi, A: Development of a HIL railway roller rig model for the traction and braking testing activities under degraded adhesion conditions. In: *International Journal of Non-Linear Mechanics* 57 (2013), S. 50–64 (Cited in page 108.)

[Aughenbaugh and Paredis 2006] Aughenbaugh, Jason M. ; Paredis, Christiaan J.: The value of using imprecise probabilities in engineering design. In: *Journal of Mechanical Design* 128 (2006), Nr. 4, S. 969–979 (Cited in pages 23 and 24.)

[Aven 2010] Aven, Terje: On the need for restricting the probabilistic analysis in risk assessments to variability. In: *Risk analysis* 30 (2010), Nr. 3, S. 354–360 (Cited in page 20.)

[Avižienis et al. 2004] Avižienis, Algirdas ; Laprie, Jean-Claude ; Randell, Brian ; Landwehr, Carl: Basic concepts and taxonomy of dependable and secure computing. In: *Dependable and Secure Computing, IEEE Transactions on* 1 (2004), Nr. 1, S. 11–33 (Cited in page 38.)

[Ayoub et al. 2013] Ayoub, Anaheed ; Chang, Jian ; Sokolsky, Oleg ; Lee, Insup: Assessing the overall sufficiency of safety arguments. In: *21st Safety-Critical Systems Symposium (SSS'13)*, 2013, S. 127–144 (Cited in pages 33, 34, 35, 47, and 48.)

[Ayoub et al. 2012] Ayoub, Anaheed ; Kim, BaekGyu ; Lee, Insup ; Sokolsky, Oleg: A systematic approach to justifying sufficient confidence in software safety arguments. In: *Computer Safety, Reliability, and Security*. Springer, 2012, S. 305–316 (Cited in pages 33 and 38.)

[BE 2001] BE, Eurocontrol B.: *Eurocontrol safety regulatory requirement 4 - risk assessment and mitigation in ATM*. 2001. – ESARR4, SRC (Cited in pages 9 and 10.)

[BE 2006] BE, Eurocontrol B.: *Safety Case Development Manual*. 2006. – DAP/SSH/091, Edition 2.2 (Cited in page 9.)

[Bezdek et al. 1984] Bezdek, James C. ; Ehrlich, Robert ; Full, William: FCM: The fuzzy c-means clustering algorithm. In: *Computers & Geosciences* 10 (1984), Nr. 2-3, S. 191–203 (Cited in page 26.)

[Bishop and Bloomfield 1998] Bishop, Peter ; Bloomfield, Robin: A methodology for safety case development. In: *Industrial Perspectives of Safety-Critical Systems*. Springer, 1998, S. 194–203 (Cited in pages 6 and 38.)

[Blockley 2013] Blockley, David: Analysing uncertainties: Towards comparing Bayesian and interval probabilities. In: *Mechanical Systems and Signal Processing* 37 (2013), Nr. 1, S. 30–42 (Cited in pages 19 and 20.)

[Bloomfield et al. 2007] Bloomfield, Robin ; Littlewood, Bev ; Wright, David: Confidence: its role in dependability cases for risk assessment. In: *Dependable Systems and Networks, 2007. DSN'07. 37th Annual IEEE/IFIP International Conference on* IEEE (Veranst.), 2007, S. 338–346 (Cited in pages 6 and 38.)

[Bloomfield et al. 2006] Bloomfield, Robin E. ; Guerra, Sofia ; Miller, Ann ; Masera, Marcelo ; Weinstock, Charles B.: International working group on assurance cases (for security). In: *Security & Privacy, IEEE* 4 (2006), Nr. 3, S. 66–68 (Cited in pages 6 and 38.)

[Chang et al. 2002] Chang, Sheng-Yung ; Lin, Cheng-Ren ; Chang, Chuei-Tin: A fuzzy diagnosis approach using dynamic fault trees. In: *Chemical Engineering Science* 57 (2002), Nr. 15, S. 2971–2985 (Cited in page 26.)

[Coletti and Scozzafava 2002] Coletti, Giulianella ; Scozzafava, Romano: Trends in Logic, Kluwer Academic Publishers, 2002 (Cited in page 22.)

[Cyra and Gorski 2007] Cyra, Lukasz ; Gorski, Janusz: Supporting compliance with security standards by trust case templates. In: *Dependability of Computer Systems, 2007. DepCoS-RELCOMEX'07. 2nd International Conference on* IEEE (Veranst.), 2007, S. 91–98 (Cited in pages 6 and 38.)

[Cyra and Gorski 2011] Cyra, Lukasz ; Gorski, Janusz: Support for argument structures review and assessment. In: *Reliability Engineering & System Safety* 96 (2011), Nr. 1, S. 26–37 (Cited in pages 33, 34, 35, 38, 47, 48, 53, 72, 73, 75, and 116.)

[De Finetti 1974] De Finetti, Bruno: *Theory of Probability.* 1974 (Cited in page 22.)

[Dempster 1966] Dempster, Arthur P.: New methods for reasoning towards posterior distributions based on sample data. In: *The Annals of Mathematical Statistics* (1966), S. 355–374 (Cited in page 27.)

[Dempster 1967] Dempster, Arthur P.: Upper and lower probabilities induced by a multivalued mapping. In: *The annals of mathematical statistics* (1967), S. 325–339 (Cited in pages 27 and 31.)

[Denney et al. 2011] Denney, Ewen ; Pai, Ganesh ; Habli, Ibrahim: Towards measurement of confidence in safety cases. In: *Empirical Software Engineering and Measurement (ESEM), 2011 International Symposium on* IEEE (Veranst.), 2011, S. 380–383 (Cited in pages 33 and 38.)

[Denœux 1999] Denœux, Thierry: Reasoning with imprecise belief structures. In: *International Journal of Approximate Reasoning* 20 (1999), Nr. 1, S. 79–111 (Cited in page 20.)

[Denœux 2011] Denœux, Thierry: Maximum likelihood estimation from fuzzy data using the EM algorithm. In: *Fuzzy sets and systems* 183 (2011), Nr. 1, S. 72–91 (Cited in page 26.)

[Denoeux and Masson 2004] Denoeux, Thierry ; Masson, M-H: Principal component analysis of fuzzy data using autoassociative neural networks. In: *IEEE Transactions on Fuzzy Systems* 12 (2004), Nr. 3, S. 336–349 (Cited in page 26.)

[DO-178C/ED-12C 2011] DO-178C/ED-12C: *Software Considerations in Airborne Systems and Equipment Certification.* 2011. – RTCA/EUROCAE (Cited in page 9.)

[DO-278A/ED-109A 2011] DO-278A/ED-109A: *Software Integrity Assurance Considerations for Communication, Navigation, Surveillance and Air*

Traffic Management (CNS/ATM) Systems,. 2011. – RTCA/EUROCAE (Cited in page 9.)

[Duan et al. 2014] Duan, Lian ; Rayadurgam, Sanjai ; Heimdahl, Mats P. ; Ayoub, Anaheed ; Sokolsky, Oleg ; Lee, Insup: Reasoning about confidence and uncertainty in assurance cases: A survey. In: *Software Engineering in Health Care*. Springer, 2014, S. 64–80 (Cited in page 33.)

[Dubois 2010] Dubois, Didier: Representation, propagation, and decision issues in risk analysis under incomplete probabilistic information. In: *Risk analysis* 30 (2010), Nr. 3, S. 361–368 (Cited in page 20.)

[Dubois 2011] Dubois, Didier: Uncertainty Theories, Degrees of Truth and Epistemic States. In: *ICAART (1)*, 2011, S. 13–14 (Cited in pages 19 and 20.)

[Dubois and Prade 2001] Dubois, Didier ; Prade, Henri: Possibility Theory, Probability Theory and Multiple-Valued Logics: A Clarification. In: *Annals of Mathematics & Artificial Intelligence* 32 (2001), Nr. 1-4, S. 35–66 (Cited in pages 27 and 31.)

[Dubois and Prade 2009] Dubois, Didier ; Prade, Henri: Formal representations of uncertainty. In: *Decision-Making Process: Concepts and Methods* (2009), S. 85–156 (Cited in pages 21, 30, and 31.)

[Dubois and Prade 1988] Dubois, Didier ; Prade, Henry: *Theory of possibility an approach to computerized processing of uncertainty*. 1988 (Cited in pages 26 and 31.)

[EN50126 1999] EN50126: *Railway applications - The specification and demonstration of Reliability, Availability, Maintainability and Safety (RAMS)*. 1999. – CENELEC, European Committee for Electrotechnical Standardization (Cited in pages 8 and 88.)

[EN50128 2011] EN50128: *Railway Applications - Software for railway control and protection systems*. 2011. – CENELEC, European Committee for Electrotechnical Standardization (Cited in pages 8, 88, and 94.)

[EN50129 2003] EN50129: *Railway Applications - Safety related electronic systems for signaling.* 2003. – CENELEC, European Committee for Electrotechnical Standardization (Cited in pages 8, 88, 93, 94, and 115.)

[EN50159 2010] EN50159: *Railway Applications - Safety related communication in transmission systems.* 2010. – CENELEC, European Committee for Electrotechnical Standardization (Cited in page 90.)

[ERA 2015] ERA, European Railway A.: *Regulation 2015/1136/EU on the Common safety method for risk evaluation and assessment.* 8 2015 (Cited in page 88.)

[Ferson et al. 2003] Ferson, Scott ; Kreinovich, Vladik ; Ginzburg, Lev ; Myers, Davis S. ; Sentz, Kari: Constructing probability boxes and Dempster-Shafer structures / Technical report, Sandia National Laboratories. 2003. – Forschungsbericht (Cited in page 23.)

[Fox and Weisberg 2011] Fox, John ; Weisberg, Sanford: *An R companion to applied regression.* Sage Publications, 2011 (Cited in page 103.)

[Gacôgne 1997] Gacôgne, Louis: *Eléments de logique floue.* Hermes, 1997 (Cited in pages 21 and 31.)

[Gallina et al. 2013] Gallina, Barbara ; Gallucci, Antonio ; Lundqvist, Kristina ; Nyberg, Mattias: VROOM & cC: a method to build safety cases for ISO 26262-compliant product lines. In: *SAFECOMP 2013-Workshop SASSUR (Next Generation of System Assurance Approaches for Safety-Critical Systems) of the 32nd International Conference on Computer Safety, Reliability and Security*, 2013 (Cited in page 7.)

[Govier 1991] Govier, Trudy: *A practical study of argument.* Wadsworth, Cengage Learning, 1991 (Cited in pages 12, 13, 14, 34, 41, 47, and 48.)

[Graydon and Holloway 2017] Graydon, Patrick J. ; Holloway, C. M.: An investigation of proposed techniques for quantifying confidence in assurance arguments. In: *Safety Science* 92 (2017), S. 53 – 65 (Cited in pages 34 and 38.)

[GSN Standard 2011] GSN Standard: *GSN COMMUNITY STANDARD VERSION 1.* 2011. – Origin Consulting (York) Limited (Cited in pages 16, 71, and 85.)

[Guiochet et al. 2015] Guiochet, Jérémie ; Do Hoang, Quynh A. ; Kaaniche, Mohamed: A Model for Safety Case Confidence Assessment. In: *Computer Safety, Reliability, and Security (SAFECOMP)*. Springer, 2015, S. 313–327 (Cited in pages 33, 38, and 47.)

[Guo 2003] Guo, Baofeng: Knowledge representation and uncertainty management: applying Bayesian belief networks to a safety assessment expert system. In: *Natural Language Processing and Knowledge Engineering, 2003. Proceedings. 2003 International Conference on* IEEE (Veranst.), 2003, S. 114–119 (Cited in page 33.)

[Hacking 1975] Hacking, Ian: *The Emergence of Probability: A Philosophical Study of Early Ideas about Probability, Induction and Statistical Inference*. Cambridge University Press, 1975 (Cited in page 19.)

[Hawkins et al. 2011] Hawkins, Richard ; Kelly, Tim ; Knight, John ; Graydon, Patrick: A new approach to creating clear safety arguments. In: *Advances in systems safety*. Springer, 2011, S. 3–23 (Cited in pages 32 and 34.)

[Hobbs and Lloyd 2012] Hobbs, Chris ; Lloyd, Martin: The application of bayesian belief networks to assurance case preparation. In: *Achieving Systems Safety*. Springer, 2012, S. 159–176 (Cited in page 33.)

[IEC61508 2010] IEC61508: *Functional safety of electrical/electronic/programmable electronic safety-related systems*. 2010. – International Electrotechnical Commission (IEC) (Cited in pages 7, 8, 88, and 89.)

[Inagaki 1991] Inagaki, Toshiyuki: Interdependence between safety-control policy and multiple-sensor schemes via Dempster-Shafer theory. In: *Reliability, IEEE Transactions on* 40 (1991), Nr. 2, S. 182–188 (Cited in page 28.)

[ISO26262 2011] ISO26262: *Software Considerations in Airborne Systems and Equipment Certification*. 2011. – International Organization for Standardization (ISO) (Cited in pages 7 and 115.)

[ISO/IEC15026-1 2013] ISO/IEC15026-1: *Systems and software engineering - Systems and software assurance - Part 1: Concepts and Vocabulary*.

2013. – International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC) (Cited in pages 8 and 115.)

[ISO/IEC15026-2 2011] ISO/IEC15026-2: *Systems and software engineering - Systems and software assurance - Part 2: Assurance Case*. 2011. – International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC) (Cited in pages 8 and 17.)

[Jøsang 2001] Jøsang, Audun: A logic for uncertain probabilities. In: *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 9 (2001), Nr. 03, S. 279–311 (Cited in pages 72 and 74.)

[Kelly 1998] Kelly, Tim: *Arguing Safety - A Systematic Approach to Safety Case Management*, Department of Computer Science, University of York, Dissertation, 1998 (Cited in pages 6, 11, 15, and 16.)

[Kelly and McDermid 1997] Kelly, Tim ; McDermid, John: Safety case construction and reuse using patterns. In: *Computer Safety, Reliability, and Security (SAFECOMP)*. Springer, 1997, S. 55–69 (Cited in pages 15 and 17.)

[Kelly and Weaver 2004] Kelly, Tim ; Weaver, Rob: The goal structuring notation–a safety argument notation. In: *Proceedings of the Dependable Systems and Networks (DSN) workshop on assurance cases*, 2004 (Cited in pages 6, 32, 35, and 38.)

[Ledinot et al. 2016] Ledinot, Emmanuel ; Blanquart, Jean-Paul ; Gassino, Jean ; Ricque, Bertrand ; Baufreton, Philippe ; Boulanger, Jean-Louis ; Camus, Jean-Louis ; Comar, Cyrille ; Delseny, Hervé ; Quéré, Philippe: Perspectives on Probabilistic Assessment of Systems and Software. In: *8th European Congress on Embedded Real Time Software and Systems (ERTS 2016)*, 2016 (Cited in page 116.)

[Menon et al. 2009] Menon ; Catherine, Cather ; Hawkins, Richard ; McDermid, John: Defence standard 00-56 issue 4: Towards evidence-based safety standards. In: *Safety-Critical Systems: Problems, Process and Practice*. Springer, 2009, S. 223–243 (Cited in pages 32 and 33.)

[Mercier et al. 2005] Mercier, David ; Quost, Benjamin ; Denœux, Thierry: Contextual discounting of belief functions. In: *Symbolic and*

Quantitative Approaches to Reasoning with Uncertainty. Springer, 2005, S. 552–562 (Cited in page 29.)

[MISRA 2017] MISRA: *Guidelines for Automotive Safety Case Arguments*. 2017. – not issued yet (Cited in page 7.)

[MoD 1996] MoD: *Defence Standard 00-56 Issue 2: Safety Management Requirements for Defence Systems*. 1996 (Cited in page 9.)

[MoD 1997] MoD: *Defence Standard 00-55: Requirements for Safety Related Software in Defence Equipment*. 1997 (Cited in page 9.)

[MoD 1999] MoD: *JSP 318B - Regulation of the Airworthiness of Ministry of Defence Aircraft*. 1999 (Cited in page 9.)

[Nair et al. 2015a] Nair, S. ; Walkinshaw, N. ; Kelly, T. ; Vara, J. L. de la: An evidential reasoning approach for assessing confidence in safety evidence. In: *2015 IEEE 26th International Symposium on Software Reliability Engineering (ISSRE)*, Nov 2015, S. 541–552 (Cited in pages 33, 34, and 35.)

[Nair et al. 2015b] Nair, Sunil ; Vara, Jose L. de la ; Sabetzadeh, Mehrdad ; Falessi, Davide: Evidence management for compliance of critical systems with safety standards: A survey on the state of practice. In: *Information and Software Technology* 60 (2015), S. 1–15 (Cited in page 33.)

[OMG 2018] OMG: *Object Managment Group: Structured Assurance Case Metamodel - SACM, version 2.0*. 2018. – URL `https://www.omg.org/spec/SACM/2.0/PDF`. – Technical report (Cited in page 17.)

[Pal et al. 2005] Pal, N. R. ; Pal, K. ; Keller, J. M. ; Bezdek, J. C.: A possibilistic fuzzy c-means clustering algorithm. In: *IEEE Transactions on Fuzzy Systems* 13 (2005), Nr. 4, S. 517–530 (Cited in page 26.)

[Pugi et al. 2006] Pugi, L ; Malvezzi, M ; Tarasconi, A ; Palazzolo, A ; Cocci, G ; Violani, M: HIL simulation of WSP systems on MI-6 test rig. In: *Vehicle System Dynamics* 44 (2006), Nr. sup1, S. 843–852 (Cited in page 108.)

[Quost and Denoeux 2016] Quost, Benjamin ; Denoeux, Thierry: Clustering and classification of fuzzy data using the fuzzy EM algorithm. In: *Fuzzy Sets and Systems* 286 (2016), S. 134–156 (Cited in page 26.)

[Shackle 1961] Shackle, George Lennox S.: *Decision order and time in human affairs*. Cambridge University Press, 1961 (Cited in page 26.)

[Shafer 1976] Shafer, Glenn: *A mathematical theory of evidence*. Bd. 1. Princeton university press Princeton, 1976 (Cited in pages 19, 27, 31, and 54.)

[Smets 1992] Smets, Philippe: The nature of the unnormalized beliefs encountered in the transferable belief model. In: *Proceedings of the 8th international conference on Uncertainty in artificial intelligence* Morgan Kaufmann Publishers Inc. (Veranst.), 1992, S. 292–297 (Cited in page 28.)

[Smets 1993] Smets, Philippe: Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem. In: *International Journal of approximate reasoning* 9 (1993), Nr. 1, S. 1–35 (Cited in page 29.)

[Smets and Kennes 1994] Smets, Philippe ; Kennes, Robert: The transferable belief model. In: *Artificial intelligence* 66 (1994), Nr. 2, S. 191–234 (Cited in page 21.)

[Toulmin 1969] Toulmin, Stephen E.: *The uses of argument*. Cambridge University Press, 1969 (Cited in page 14.)

[Walley 1991] Walley, Peter: Statistical reasoning with imprecise probabilities. (1991) (Cited in page 31.)

[Walley 2000] Walley, Peter: Towards a unified theory of imprecise probability. In: *International Journal of Approximate Reasoning* 24 (2000), Nr. 2, S. 125–148 (Cited in page 23.)

[Wang et al. 2016a] Wang, Rui ; Guiochet, Jérémie ; Motet, Gilles: A Framework for Assessing Safety Argumentation Confidence. In: *International Workshop on Software Engineering for Resilient Systems* Springer (Veranst.), 2016, S. 3–12 (Cited in pages 3 and 38.)

[Wang et al. 2017a] Wang, Rui ; Guiochet, Jérémie ; Motet, Gilles: Confidence assessment framework for safety arguments. In: *International Conference on Computer Safety, Reliability, and Security (SafeComp)* Springer (Veranst.), 2017, S. 55–68 (Cited in pages 3 and 70.)

[Wang et al. 2016b] Wang, Rui ; Guiochet, Jérémie ; Motet, Gilles ; Schön, Walter: DS theory for argument confidence assessment. In: *International Conference on Belief Functions* Springer (Veranst.), 2016, S. 190–200 (Cited in pages 3 and 38.)

[Wang et al. 2017b] Wang, Rui ; Guiochet, Jérémie ; Motet, Gilles ; Schön, Walter: Modelling confidence in railway safety case. In: *Safety Science* (2017). – URL `http://www.sciencedirect.com/science/article/pii/S0925753517313164`. – ISSN 0925-7535 (Cited in pages 3 and 87.)

[Werro 2016] Werro, Nicolas: *Fuzzy classification of online customers*. Springer, 2016 (Cited in pages 23, 24, and 25.)

[Yager 1987] Yager, Ronald R.: On the Dempster-Shafer framework and new combination rules. In: *Information sciences* 41 (1987), Nr. 2, S. 93–137 (Cited in page 28.)

[Yuan et al. 2017] Yuan, Chunchun ; Wu, Ji ; Liu, Chao ; Yang, Haiyan: A Subjective Logic-Based Approach for Assessing Confidence in Assurance Case. In: *International Journal of Performability Engineering* 13 (2017), Nr. 6, S. 807 (Cited in page 34.)

[Zadeh 1965] Zadeh, LA: Fuzzy sets. In: *Information and Control* 8 (1965), S. 338–353 (Cited in page 23.)

[Zadeh 1973] Zadeh, Lotfi A.: Outline of a new approach to the analysis of complex systems decision processes. In: *IEEE Trans. Systems, Man, and Cybernetics* 3 (1973), Nr. 1, S. 28–44 (Cited in page 19.)

[Zadeh 1978] Zadeh, Lotfi A.: Fuzzy sets as a basis for a theory of possibility. In: *Fuzzy sets and systems* 1 (1978), Nr. 1, S. 3–28 (Cited in page 26.)

[Zimmermann 1996] Zimmermann, H-J: Fuzzy control. In: *Fuzzy Set Theory—and Its Applications*. Springer, 1996, S. 203–240 (Cited in page 26.)