# Modelling Confidence in Railway Safety Case

Rui Wang, Jérémie Guiochet, Gilles Motet, Walter Schön

▶ **To cite this version:**

Rui Wang, Jérémie Guiochet, Gilles Motet, Walter Schön. Modelling Confidence in Railway Safety Case. Safety Science, 2018, 110 (part B), pp.286-299. 10.1016/j.ssci.2017.11.012 . hal-01661045

**HAL Id: hal-01661045**
**https://hal.science/hal-01661045**

Submitted on 11 Dec 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Modelling Confidence in Railway Safety Case

Rui Wang[a,*], Jérémie Guiochet[a,*], Gilles Motet[a,*], Walter Schön[b,*]

[a] *LAAS-CNRS, Université de Toulouse, CNRS, INSA,UPS, Toulouse, France*
[b] *Sorbonne Universités, Université de Technologie de Compiègne, Heudiasyc UMR CNRS 7253 CS 60319 Compiègne Cedex France*

**Abstract**

Railway standard EN50129 clarifies the safety acceptance conditions of safety-related electronic systems for signalling. It requires using a structured argumentation, named Safety Case, to present the fulfilment of these conditions. As guidance for building the Safety Case, this standard provides the structure of high-level safety objectives and the recommendations of development techniques according to different Safety Integrity Levels (SIL). Nevertheless, the rationale connecting these techniques to the high-level safety objectives is not explicit. The proposed techniques stem from experts belief in the effectiveness and efficiency of these techniques to achieve the underlying safety objectives. So, how should one formalize and assess this belief? And as a result how much confidence can we have in the safety of railway systems when these standards are used? To deal with these questions, the paper successively addresses two aspects: 1) making explicit the safety assurance rationale by modelling the Safety Case with GSN (Goal Structuring Notation) according to EN5012x standards; 2) proposing a quantitative framework based on Dempster-Shafer theory to formalize and assessing the confidence in the Safety Case. A survey amongst safety experts is carried out to estimate the confidence parameters. With these results, an application guidance of this framework is provided based on the Wheel Slide Protection (WSP) system.

*Keywords:*
Safety case, Safety argumentation, Railway signalling system, Confidence assessment, EN50129, Dempster-Shafer theory

*Corresponding authors: [a] {name.surname}@laas.fr, [b] wschon@utc.fr

## 1. Introduction

The railway standard EN50129 (2003) clarifies the safety acceptance conditions of safety-related electronic systems for railway. It requires using a structured argumentation, named Safety Case to present the fulfilment of these conditions. In order to guide the development of the Safety Case, this standard provides high-level safety objectives as well as recommendations for development techniques. These techniques are formulated according to the various Safety Integrity Levels (SIL) required. Thus, it can be speculated that the experts involved in the development of the standard have varying degrees of belief in the effectiveness of the proposed techniques or technique combinations in order to achieve the safety objectives.

Nevertheless, the rationale of how the required techniques contribute to the achievement of the high-level safety objectives is not explicit in this standard. Little explanation is given about the effectiveness of each technique. The Safety Case documenting the safety evidence and arguments is then implicit. This raises the difficulties concerning the system safety assessment based on safety argumentation. We may have doubts about the extent to which the techniques used would lead to adequate confidence in the system safety. Moreover, the checklist of proposed techniques does not favour adaptation to innovative ones.

These issues motivated our research work, which was developed together with safety experts. These topics also strongly concern the railway safety authorities, the railway system designers, and the developers of the standards. To deal with these issues, our paper addresses two aspects: modelling the safety argumentation required by the railway standards and assessing the confidence in the system safety based on the safety argumentation. The first aspect aims to qualitatively render explicit the implicit rationale of safety assurance expressed in the standards. The Safety Case models based on EN50129 are made with graphically notated arguments called GSN (Goal Structuring Notation). To deal with the second aspect, we propose a confidence assessment framework. The confidence parameters are quantitatively defined and measured based on Dempster-Shafer theory. We have carried out a case study to demonstrate the estimation process of the confidence parameters with the help of a survey amongst safety experts. An application guidance of this framework is provided based on the Wheel Slide Protection (WSP) system.

The rest of this paper is organized as follows. In Section 2, we provide the background information on the railway safety standards for signalling systems, the GSN model and the Dempster-Shafer theory. In Section 3, we develop a fragment of the Safety Case based on the railway safety standards. Section 4 introduces the confidence assessment framework. At first, we present an overview of this framework; and then we develop the technical details, such as the confidence parameter formal-

ization, confidence aggregation rules and expert judgement extraction. A survey for the confidence parameters estimation is presented. Section 5 describes how to apply this framework with a simplified example of WSP system. Finally, we summarize the contributions of our approach and highlight future work.

## 2. Background

In this section, we introduce three prerequisites: a general introduction of EN50129, the modelling approach for safety argumentation (GSN) and the Dempster-Shafer theory.

### *2.1. The railway safety standards for signalling systems*

The EN5012x series standards provide a guideline for ensuring the functional safety of safety-related electronic systems for railway signalling applications. On a higher-level of safety regulation aiming to European nations, the European Railway Agency (ERA) proposes the concept and the guideline of Common Safety Method (CSM) (ERA, 2015) for risk evaluation and assessment. This regulation presents a more generic risk assessment process in compliance with the EN5012x standards to harmonised design targets for railway technical systems.

EN5012x standards are derived from the generic 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*. No specific format is required to describe a safety case.

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 conditions: 1) evidence of quality management, 2) evidence of safety management, 3) evidence of functional and technical safety.

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 4 severity classes (Insignificant, Marginal, Critical, Catastrophic). The SIL4 is the highest critical safety integrity level. The recommended safety assurance techniques are differentiated according to these SILs. In Section 3, we further discuss the technical safety arguments proposed by EN50129.

### *2.2. Safety Argumentation modelling with GSN*

Structuring an argument to convince regulation bodies is a main challenge for critical system developers. Many approaches, such as safety case (Kelly and Weaver, 2004; Bishop and Bloomfield, 1998), assurance case (Bloomfield et al., 2006), trust case (Cyra and Gorski, 2007), and dependability case (Bloomfield et al., 2007), provide concepts and notations for taking up this challenge. The safety case, a popular form of safety argumentation, can be defined as *"a clear, comprehensive and defensible argument that a system is acceptably safe to operate in a particular context"* (Kelly, 1998). It is used to communicate the developers' rationale for implementing the development or the choice of techniques. In some safety-critical domains, the safety case is explicitly required in the safety standards, such as the ISO26262 for automotive and EN50129 for railway. A main difference between the definitions of a safety case is the use of "acceptably safe" as the objective, where in standards "compliance with safety requirements" is considered. Both are safety objectives. The statement "acceptably safe" is supported by the statement "compliance with safety requirements", 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.

A graphic argumentation notation of safety case, called Goal Structure Notation (GSN) has been developed by Kelly (1998). An example of GSN is given in Fig. 1, which is derived from the Hazard Avoidance Pattern (Kelly and McDermid, 1997). GSN enables to structure and represent the objectives to be achieved: *goal* (e.g. G1); the goal-supporting evidence: *solutions* (e.g. Sn1); *strategy* for braking down a goal to sub-goals, (e.g. S1), *context* (e.g. C1), etc. Additionally, a notation of *Assumption* is also used in this paper (see *A1* in Fig. 3).

### *2.3. The Dempster-Shafer Theory*

The Dempster-Shafer (D-S) Theory or Belief Function Theory was developed by Arthur Dempster and Glenn Shafer successively (Shafer, 1976). This theory offers a powerful tool to model the human belief in evidence from different sources, and an explicit modelling of epistemic uncertainties. It possesses a higher mathematical generality than other uncertainty theories, such as probabilistic approaches, possibility

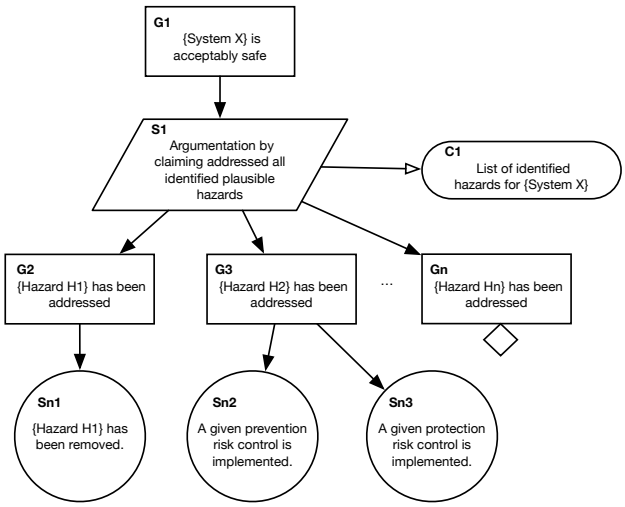



Figure 1: GSN example adapted from Hazard Avoidance Pattern (Kelly and McDermid, 1997)

theory, fuzzy set, etc. (Dubois and Prade, 2009). As presented later, we propose using the D-S Theory as it allows uncertainty, imprecision or ignorance, i.e., "we know that we don't know", to be explicitly expressed. We recall here two basic concepts to better understand this paper.

Let $X$ be a variable taking values in a finite set $\Omega$ representing a *frame of discernment*. $\Omega$ is composed of all the possible situations of interest. In this paper, we consider only "binary" frames of discernment. For instance, if $X$ represents the state of a bulb, then $\Omega$=$\{on, off\}$. The *mass function* $m^\Omega$ is the mapping of the power set of $\Omega$ on the closed interval [0,1] that is, $2^\Omega \rightarrow [0,1]$. Assume that $P$ is a subset of $\Omega$, $P \subseteq \Omega$. Thus, $P$ is an element of $2^\Omega$, that is, $P \in 2^\Omega$. The mass function of $P$ must satisfy:

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

The mass $m^\Omega(P)$ reflects the degree of belief committed to the hypothesis that the truth lies in $P$. For instance, we can have the following assignment of belief: $m(\{on\}) = 0.5$, $m(\{off\}) = 0.3$, $m(\{on, off\}) = 0.2$. Note that $m(\{on, off\})$ 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".

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

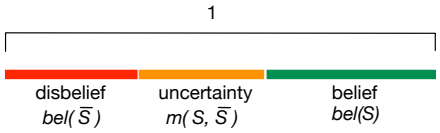



Figure 2: The measures of truth of statement S with D-S theory

$$bel(P) = \sum_{M \subseteq P, M \neq \emptyset} m^{\Omega}(M) \quad \forall P \subseteq \Omega \tag{2}$$

For instance, $bel(\{on\}) = m(\{on\}) = 0.5$. The belief function bel(P) for all $P \subseteq \Omega$ can be interpreted as *the degree of justified specific support* given to P (Smets, 1992).

Let's consider, for instance, a statement $S$ "{System X} is acceptably safe". Similarly, the frame of discernment $\Omega_S$ for the truth of $S$ is also binary: $\{S, \overline{S}\}$ or {True, False}. An opinion of the truth of this statement can be explicitly assessed with 3 measures represented in Fig. 2. These measures are *belief* ($bel_S = m(\{S\})$), *disbelief* ($disb_S = m(\{\overline{S}\})$), and the *uncertainty* ($uncer_S = m(\{S, \overline{S}\})$). This leads to $m(\{S\}) + m(\{\overline{S}\}) + m(\{S, \overline{S}\}) = 1$ (*belief + disbelief + uncertainty = 1*). Hence, we define the measures of the truth of statement S as follows:

$$\begin{cases} bel_S &= bel(\{S\}) = m(\{S\}) \\ disb_S &= bel(\{\overline{S}\}) = m(\{\overline{S}\}) \\ uncer_S &= m(\{S, \overline{S}\}) = 1 - m(\{S\}) - m(\{\overline{S}\}) \end{cases} \tag{3}$$

where $bel_S$, $disb_S$, $uncer_S \in [0, 1]$. $bel_S$, $disb_S$ and $uncer_S$ represent the degree of our belief in, disbelief in or doubt about the adequate safety of {System X}.

## 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 GSN model. Then, a proposal for the necessary but missing inference between them is given based on the analysis of standards and engineering experience.

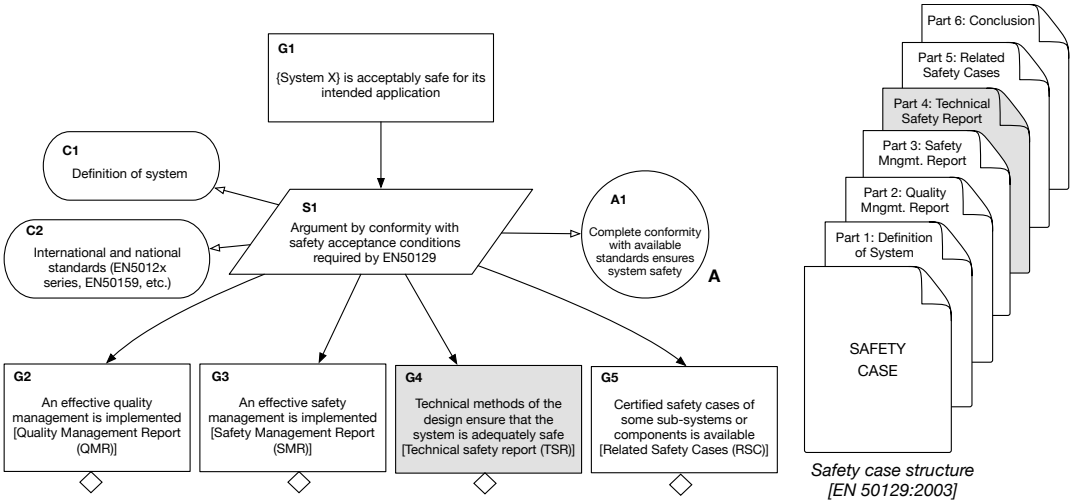



Figure 3: The highest level model of Safety Case

### *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 Fig. 3 and Fig. 4). 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 Fig. 3, 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 contextual information (*C1*) for the entire safety argument. Another context (*C2*) is the existing international and 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 formu-

lating the safety case. Goal G4 is taken as an example to represent the third layer of sub-goals. In Fig. 4, 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 Fig. 4, 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 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.

### *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 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 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 safety may decrease. Therefore, the use or non-use of a technique mentioned in the standards is considered

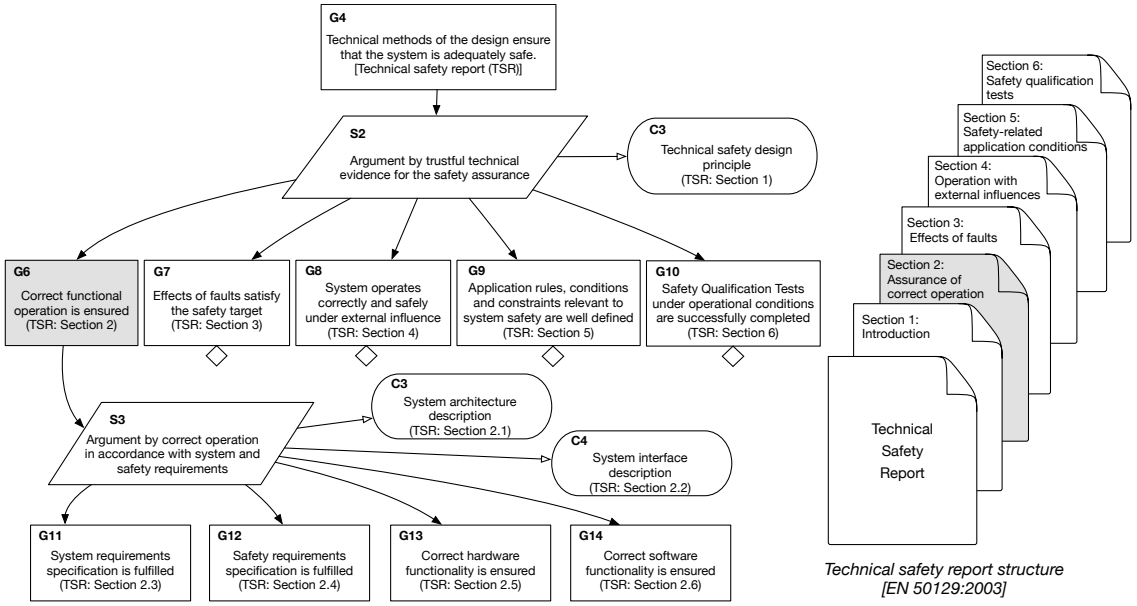



Figure 4: The main structure of the Technical Safety Report

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.

### *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 Fig. 5, the inference gap is presented with the dashed schema between the *High-level Safety*

Table 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 | |

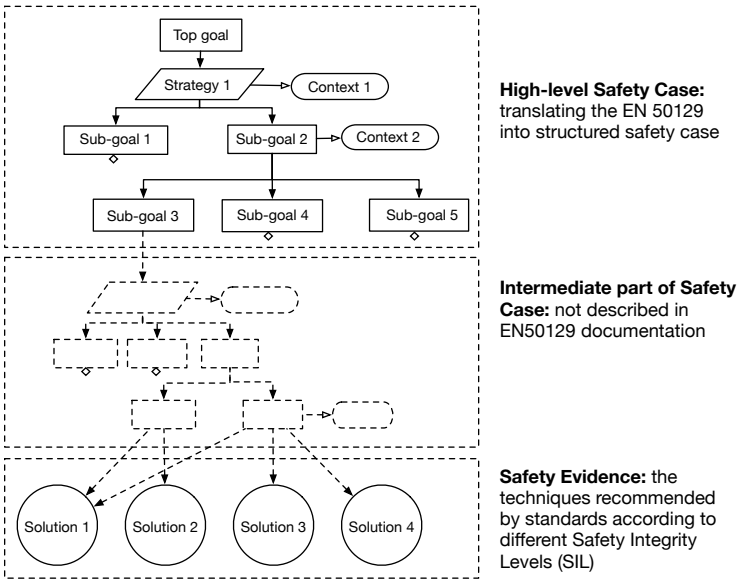



Figure 5: Inference gap between high-level goals and safety evidence required by the EN50129

*Case* and *Solutions* (also called evidence).

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 Fig. 6.

There is little guidance to justify the fulfilment of *G12* (see B.2.4 EN50129). 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 (Table E.9 in the standard annex). In this table, there are 11 recommended techniques and measures for the V&V process. Thus, the supporting *solutions* (Sn12.1-Sn12.11 in Fig. 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. 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 Fig. 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: *Argument by the traceability and satisfaction of all safety requirements (S12.1)* and *Argument by high confidence demonstrated by actual use (S12.2)* with the assumption that *Previous operational evidence is available (A12.1)*.

For a newly developed system, the high-level safety requirements are generally ensured through 1) a vertical view: all safety requirements are traced in each life-cycle phases (G12.1) and 2) a horizontal view: the conclusive validation of the satisfaction of safety requirements (G12.2). This strategy is actually from the "V-model" representation of the life-cycle introduced in EN50126. During the development and design phases, traceability amongst various requirement specifications, test cases, etc. is ensured (G12.4). Three techniques are related to this goal: *Sn12.1 Checklists* (e.g. checklist for the deliverable required in each life-cycle phase), *Sn12.2 Inspection of documentation* (e.g. coverage verification between high-level and low-level requirements) and *Sn12.3 Design review* (e.g. verification of conformity between specification and design implementation, code review). In the manufacturing, installation and maintenance processes, the design assumptions are required to be ensured (G12.5). The requirements, precautions and audit reports for corresponding processes constitute the safety evidence (Sn12.4 and Sn12.5). In the other branch, G12.2 is ensured by simulation (G12.6), functional testing (G12.7) and robustness testing (G12.8). The functional testing is required to be applied during development (G12.9) and under environmental conditions (G12.10). Test results (Sn12.7 and Sn12.9) and independence of test design (Sn12.8) can be the related evidence.

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, safety operational time 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 the guidance to build a reasonable safety case. The techniques requirements in Annex E 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 Fig. 6 only presents what should be done for one sub-goal (G12), but similar analyses are needed for other sub-goals (see Fig. 3 and 4).

## 4. Quantitative Safety Case Assessment

In this section, we present an extension of the assessment framework for structured safety arguments introduced by Wang et al. (2017). It allows 1) assessors to provide their opinions on the lowest level claims of a structured safety argument based on available evidence; and 2) to aggregate these opinions hierarchically until we obtain the confidence in the top claim of the argument.

### *4.1. Framework Overview*

The proposed assessment framework of the safety argument is summarized in Fig. 7. The assessment process contains the following four steps:

1) Building the Safety Case of a safety-related electronic railway system (developed in Section 3). The safety argument model follows the structure required by EN50129 and the guideline for the more specific arguments and safety evidence studied in the previous section. Then, 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. Necessary modifications of the safety arguments should be implemented.

2) Estimating the parameters affecting the confidence propagated upwards in the argumentation, i.e. the *contributing weights* of sub-goals and the *degree of complementary/redundant* among sub-goal statements depending on argument types. The estimation of these parameters requires data from experts. The parameter estimating strategy is introduced in Section 4.4.

After the first two steps, we obtain a complete safety argumentation 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 other specific railway systems.

3) Assessing confidence in sub-goals based on available safety evidence. The confidence assessment starts from the lowest level of sub-goals. A scaled evaluation table is used to extract the experts' judgement of a sub-goal based on the supporting evidence presented in Section 4.3. This judgement is assessed based on two values: the *Decision* (*dec*) on the statement of the goal and the *Confidence in this decision* (*conf*). They are then transformed into a 3-tuple *(bel,uncer,disb)* representing *Belief*, *Uncertainty* and *Disbelief* in this goal, as shown in Fig. 7, with an uncertainty triangle named Jøsang triangle (Jøsang, 2001). This transformation from the experts' judgement to *(bel,uncer,disb)* refers to a related work (Cyra and Gorski, 2011).

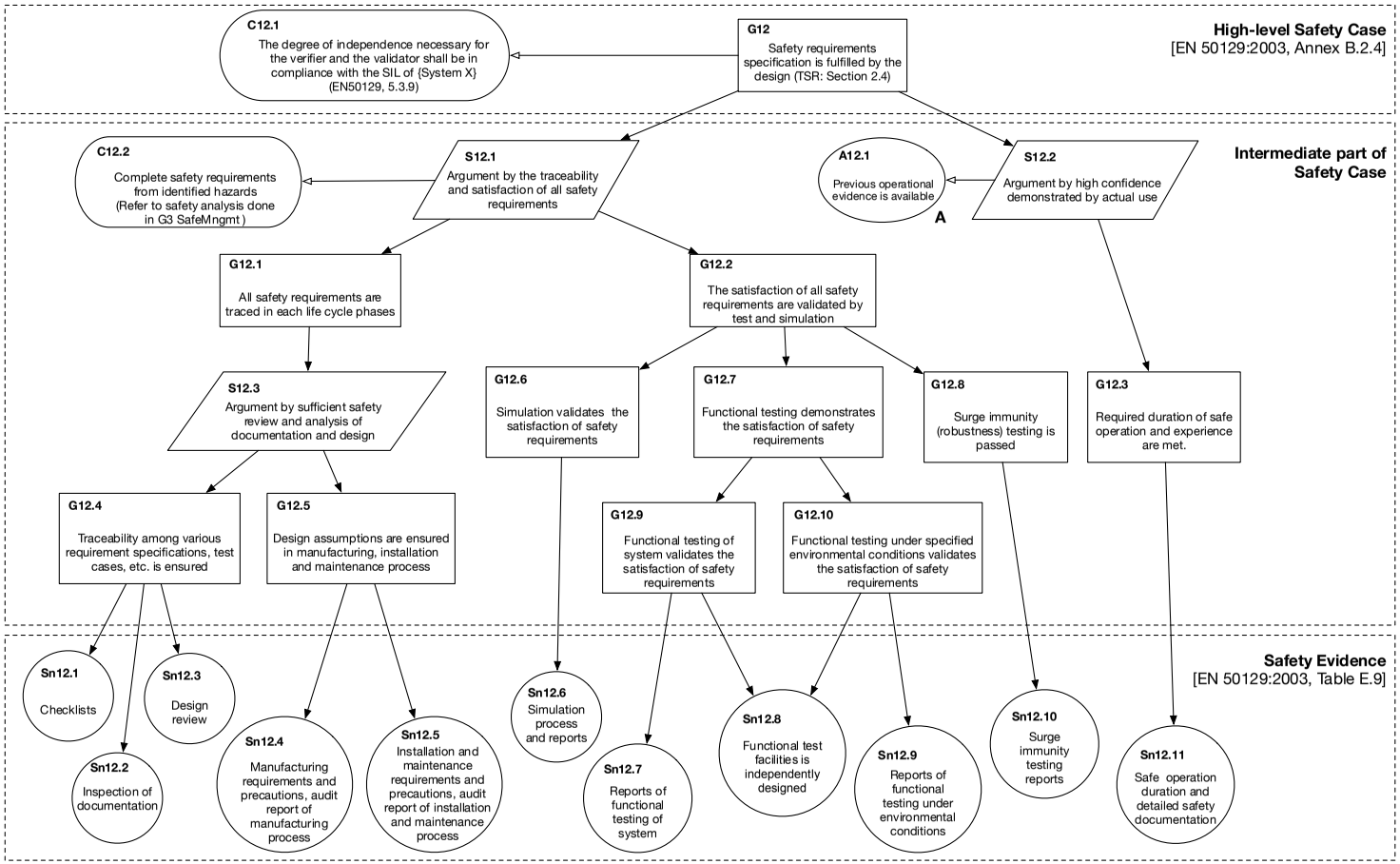



Figure 6: GSN presenting the rationale between the goal G12 and relating safety evidence

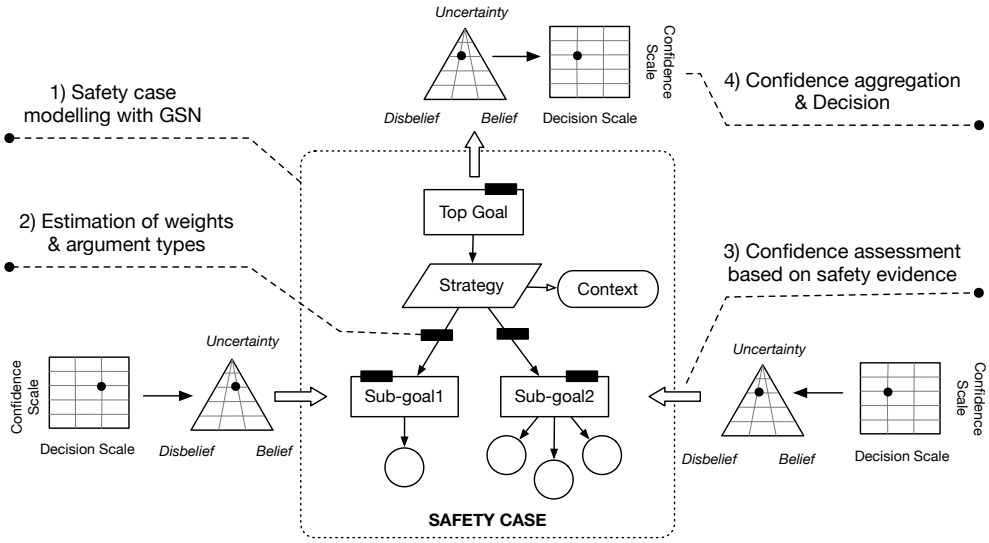



Figure 7: Overview of confidence assessment framework for the safety argument

4) Aggregating confidence and calculating the decision on the *Top Goal*. Confidence aggregation is realized through the combination of the estimations *(bel,uncer,disb)* of lower-level goals into the higher-level goals (see Section 4.2). This step is based on the confidence assessment method derived from the Dempster-Shafer theory (Wang et al., 2016b). As shown in Fig. 7, *(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 judgement on the *Top Goal*, i.e. the *decision* and the *confidence in this decision (dec, conf)* of the *Top Goal*.

### *4.2. Argument Types and Confidence Assessment*

Like most structured arguments, a GSN argument has a tree structure, which is composed of a top goal and branches of sub-goals. As described by Hawkins et al. (2011), the assessment of confidence in the top goal may be based on the assessment of the *trustworthiness* of each sub-goal and the *appropriateness* of the sub-goals with regard to the top goal. Thus, we propose the assessment parameters shown in Fig. 8 with an example of a simple argument: goal A is supported by two sub-goals B and C. These parameters are:

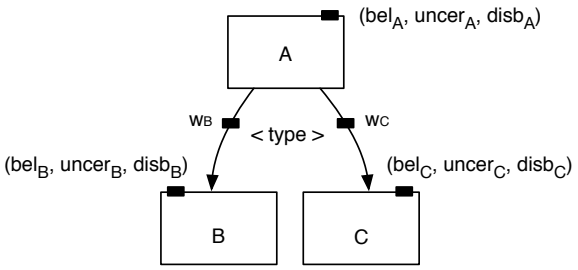



Figure 8: A two-node argument annotated with assessment parameters

- The *trustworthiness* of the sub-goals B ($bel_B, disb_B, uncer_B$) and C ($bel_C, disb_C$, $uncer_C$);
- The *appropriateness* of the sub-goals B and C, including the contributing weights, i.e. $w_{B\to A}$, $w_{C\to A}$ (simplified as $w_B$, $w_C$) and the argument types of inference ($< type >$). Here, the $< type >$ could be equal to $< Single >$ (in cases where a single argument is used) or $< Complementary >$ or $< Redundant >$ (in cases where both arguments are used).

D-S theory enables to formalize these measurements and aggregate their values to obtain a combined result for the *trustworthiness* of top goal A. We briefly introduce below the formalized concepts and the final aggregation rules. The development of the calculus has been presented by Wang et al. (2016b).

*4.2.1. Trustworthiness of the Sub-goals*

As it has been introduced in Section 2.3, masses can be assigned to a statement of a safety case as the belief, disbelief and uncertainty ($bel, disb, uncer$): this 3-tuple will represent the trustworthiness of the statement. According to Eq. 3, the assessment parameters for sub-goals B ($bel_B, disb_B, uncer_B$) are:

$$Trusworthiness_B : \begin{cases} bel_B &= bel(B) = m(B) \\ disb_B &= bel(\overline{B}) = m(\overline{B}) \\ uncer_B &= m(B, \overline{B}) = 1 - m(B) - m(\overline{B}) \end{cases} \tag{4}$$

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

For instance, goal B is "Testing is conclusive". $bel_B$, $disb_B$ and $uncer_B$ represent the degree of our belief in, disbelief in or doubt about the conclusiveness of the functional testing. It depends, for instance, on the completeness of the test sequence and the correctness of the test results.

*4.2.2. Appropriateness of the Sub-goals*

This subsection presents how sub-goals can be combined to support a top goal. Here, only arguments with 1 or 2 sub-goals (as in Fig. 8) are considered, but the following definitions and aggregation rules could be extended to the arguments with more sub-goals.

We consider that sub-goals B and C may contribute to the top goal A in four different ways. In our approach, we formally model these four ways with the weight of each sub-goal and the types of the inference. The formalization of these parameters are based on D-S theory. We propose to assign masses to different combinations of inferences to express the ways of confidence contributions. Here, subsets of $\{(\overline{B},\overline{A}),(B,\overline{A}),(\overline{B},A),(B,A)\}$ are used to represent the possible inferences between A and B: e.g. $\{(\overline{B},\overline{A})\}$ stands for "when B is false, A is false". Additionally, subsets of $\{(\overline{B},\overline{C},\overline{A}),(B,\overline{C},\overline{A}),(\overline{B},C,\overline{A}),(\overline{B},\overline{C},A),(B,C,\overline{A}),(\overline{B},C,A),(B,\overline{C},A)(B,C,A)\}$ are used to represent the possible inferences among A, B and C: e.g. $\{(\overline{B},\overline{C},\overline{A})\}$ stands for "when B and C are false, A is false". Therefore, the four contributing ways are:

- Pure B alone: A exclusively depends on B.

$$w_B = m(\{(\overline{B},\overline{A}),(B,A)\}) = 1 \tag{5}$$

  $w_B$: the degree that A depends on B.

- Pure C alone: A exclusively depends on C

$$w_C = m(\{(\overline{C},\overline{A}),(C,A)\}) = 1 \tag{6}$$

  $w_C$: the degree that A depends on C.

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

$$w_{B\times C\to A} = m(\{(\overline{B},\overline{C},\overline{A}),(\overline{B},C,\overline{A}),(B,\overline{C},\overline{A}),(B,C,A)\}) = 1 \tag{7}$$

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

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

$$w_{B+C\to A} = m(\{(\overline{B},\overline{C},\overline{A}),(\overline{B},C,A),(B,\overline{C},A),(B,C,A)\}) = 1 \tag{8}$$

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

Referring to some existing works of Govier (2013), Ayoub et al. (2013), Cyra and Gorski (2011) and Guiochet et al. (2015), most arguments are not always *"pure AND"* nor *"pure OR"*. Therefore, some of them can be regarded as "mixed" arguments. We proposed five main argument types including two "pure" and three "mixed" argument cases:

- Fully complementary argument (FC-Arg) corresponds to the *"Pure AND gate"* case (Eq. 7). In this case, $w_{B\times C\to A} = 1$, representing the full complementarity between sub-goals.

- Partial complementary argument (PC-Arg) is the combination of *"Pure B alone"* (Eq. 5), *"Pure C alone"* (Eq. 6) and *"Pure AND gate"* (Eq. 7). $w_{B\times C\to A} \in (0, 1)$, representing the degree of the complementarity between sub-goals.

- Fully redundant argument (FR-Arg) corresponds to the *"Pure OR gate"* case (Eq. 8). In this case, $w_{B+C\to A} = 1$, representing the full redundancy between sub-goals.

- Partial redundant argument (PR-Arg) is the combination of *"Pure B alone"* (Eq. 5), *"Pure C alone"* (Eq. 6) and *"Pure OR gate"* (Eq. 8). $w_{B+C\to A} \in (0, 1)$, representing the degree of the redundancy between sub-goals.

- Disparate argument (D-Arg) is the combination of *"Pure B alone"* (Eq. 5) and *"Pure C alone"* (Eq. 6). It can be seen as the limit case of partial redundant with $w_{B\times C\to A} = 0$ or the limit case of partial complementary with $w_{B+C\to A} = 0$.

As most arguments are considered as "mixed" arguments, here are two examples for the partial complementary/redundant arguments. The argument: *"B: High-level requirements coverage is achieved"* and *"C: Low-level requirements coverage is achieved"* support *"A: System is acceptably safe"* can be regarded as *partial complementary argument*. The argument: *"B: Formal verification is passed"* and *"C: Test is conclusive"* support *"A: System is acceptably safe"* is believed as *partial redundant argument*.

In those mixed argument cases, the $w_B$, $w_C$, $w_{B\times C\to A}$ and $w_{B+C\to A}$ are no more equal to 1. According to the definition of mass function, the constraints $w_B + w_C + w_{B\times C\to A} = 1$ (for partial complementary) or $w_B + w_C + w_{B+C\to A} = 1$ (for partial redundant) should be satisfied. Thus, we have $w_{B\times C\to A} = 1 - w_B - w_C$ or $w_{B+C\to A} = 1 - w_B - w_C$.

Table 2: Trustworthiness aggregation rules for two argument types

| Types | Aggregation rules | |
|---|---|---|
| PC-Arg | $bel_A = bel_B w_B + bel_C w_C + bel_B bel_C(1 - w_B - w_C)$<br>$disb_A = disb_B w_B + disb_C w_C + (disb_B + disb_C - disb_B disb_C)(1 - w_B - w_C)$<br>$uncer_A = 1 - bel_A - disb_A$ | (9) |
| PR-Arg | $bel_A = bel_B w_B + bel_C w_C + (bel_B + bel_C - bel_B bel_C)(1 - w_B - w_C)$<br>$disb_A = disb_B w_B + disb_C w_C + disb_B disb_C(1 - w_B - w_C)$<br>$uncer_A = 1 - bel_A - disb_A$ | (10) |

*4.2.3. Trustworthiness propagation*

All confidence parameters of sub-goals have been defined. The *trustworthiness* of top goal A is obtained based on the combination of the *trustworthiness* (Eq. 4) and *appropriateness* of sub-goals (Eq. 5,6 and 7 or 5,6 and 8) based on the Dempster Rule. The issue of conflicting combinations is avoided because of the special definition of masses. The aggregation rules for the *trustworthiness* of A ($bel_A$, $disb_A$, $uncer_A$) are listed in Table 2 for the partial complementary and partial redundant arguments, respectively. Due to the limited size of this paper, the parameter formalization and the development process of the aggregation rules are not presented. For more details and a general assessment model for N-node arguments, please refer to (Wang et al., 2016b).

It's worth noting that the aggregation rules for the three special arguments *fully complementary argument*, *fully redundant argument* and *disparate argument* are relatively intuitive and consistent with corresponding cases:

- Fully complementary argument (FC-Arg) ($w_{B\times C\to A} = 1$). According to Eq. 9, the belief in goal A conforms to the expression of *AND logic*: $bel_A = bel_B bel_C$.
- Fully redundant argument (FR-Arg) ($w_{B+C\to A} = 1$). According to Eq. 10, the belief in goal A conforms to the expression of *OR logic*: $bel_A = bel_B + bel_C - bel_B bel_C$.
- Disparate argument (D-Arg): When $w_{B\times C\to A} = 0$ AND $w_{B+C\to A} = 0$ (i.e. $1 - w_B - w_C = 0$), the aggregation rules of the complementary and redundant arguments become the same formula: $bel_A = bel_B w_B + bel_C w_C$.

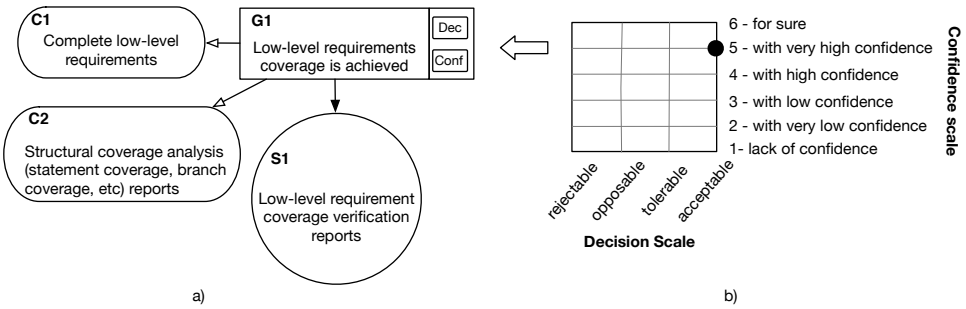



Figure 9: An evaluation matrix for safety argument

### *4.3. Expert Judgement Extraction*

While assessing an argument, an assessor has to evaluate all the elements of this argument, i.e. statement, evidence, context, etc. In Fig. 9 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 based on the contexts C1: *"Complete low-level requirements"* and C2: *"Structural coverage analysis (statement coverage, branch coverage, etc) reports"*. 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 Fig. 9 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"*. We assume that, in both scales, the levels are evenly and linearly distributed. A solid dot represents the evaluation of this goal by an assessor. Here, the assessor 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 further assess the higher-level goals, we need to aggregate the assessor's evaluation of sub-goals. As mentioned in the Section 4.1, the evaluation of the experts *(dec, conf)* will be transformed into *belief*, *uncertainty* and *disbelief*. In fact, this step is used to formalize the evaluation as a mass function in order to take advantage of the D-S Theory to combine uncertain information. This uncertainty theory offers a powerful tool to explicitly model and process information with uncertainty. We adopt the definition of *decision* and *confidence in the decision* of any claim A based on belief functions proposed in (Cyra and Gorski, 2011) to fit the input of the aggregation rules (refer to Table 2). The modified definition is presented in Eq. 11 and 12.

$$conf_A = bel_A + disb_A \tag{11}$$

$$\begin{cases} dec_A = bel_A/(bel_A + disb_A), & bel_A + disb_A \neq 0 \\ dec_A = 0, & bel_A + disb_A = 0 \end{cases} \tag{12}$$

Due to the constrain condition of Eq. (1), we can deduce that $conf_A$, $dec_A \in [0, 1]$. Furthermore, the inverse functions from $(dec_A, conf_A)$ to $(bel_A, uncer_A, disb_A)$ are given in the Eq. 13.

$$\begin{cases} bel_A = conf_A * dec_A \\ disb_A = conf_A * (1 - dec_A) \\ uncer_A = 1 - bel_A - disb_A \end{cases} \tag{13}$$

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.

In brief, our framework can be regarded as a function $f$:

$$(dec_A, conf_A) = f[(dec_B, conf_B), (dec_C, conf_C), w_B, w_C, < types >] \tag{14}$$

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 (Wang et al., 2016b).

### *4.4. Parameter Estimation*

According to the framework proposed in Section 4.1, it is necessary to determine argument types and weights of sub-goals in order to complete the argumentation model for assessment. We implemented a survey amongst safety experts for estimating the values of these parameters.

#### *4.4.1. Implementation of the Survey*

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 instinctive, that is, actual, opinions of respondents. As our objective is to demonstrate the

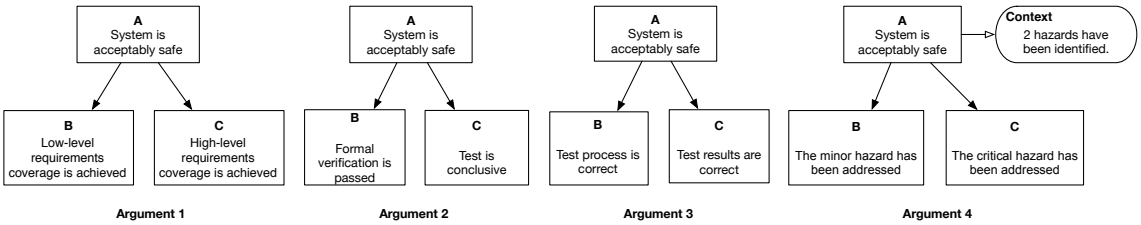



Figure 10: Argument fragments questioned in the survey

effect of our method and not to exploit the full railway standards, four simple and general safety argument patterns are used in the questionnaire. They are presented in Fig. 10. These four arguments have the same form with an identical top goal A and two sub-goals B and C in accordance with the argument in Fig. 8. Two pairs of initial inputs expressing the trustworthiness of sub-goals B and C are provided for each argument in the evaluation matrix (see Fig. 11). The respondents are asked to make the decision on the acceptance of goal A, and the confidence in this decision, e.g. Questions for 1.a) (on the right of Fig. 11). The parameter values are then deduced from these responses as introduced in Section 4.4.2.

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 is given in 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". An online version of this questionnaire is available (Wang, 2017).

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

*4.4.2. Analysis Results*

The case study aims to estimate the weights and argument types of sub-goals implicitly considered by the experts. The collected data (*expert data*) are compared with the data calculated based on the assessment formulas introduced in Section 4.2 (*theoretical data*).

In Fig. 12, the theoretical data of complementary and redundant arguments are shown as a cloud of grey dots derived from all possible combinations of the values of $w_B$ and $w_C$. The triangles in this figure will be explained later. Figures a) and b)

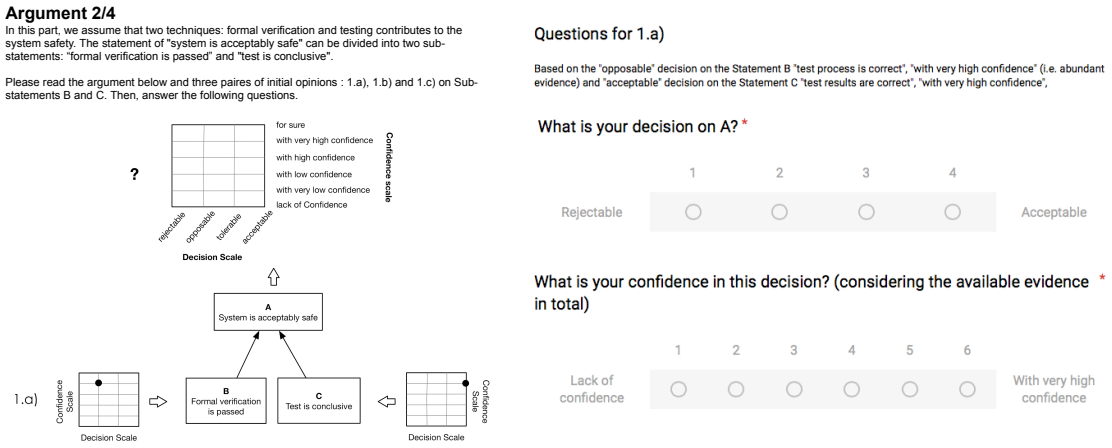
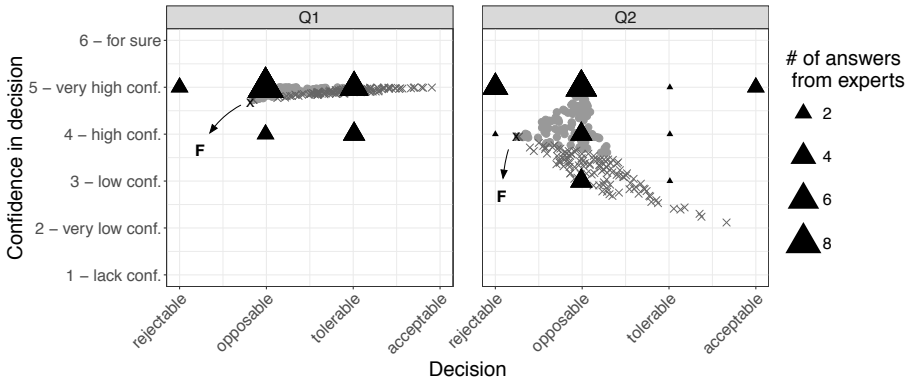



Figure 11: Screenshot of a fragment of the online questionnaire

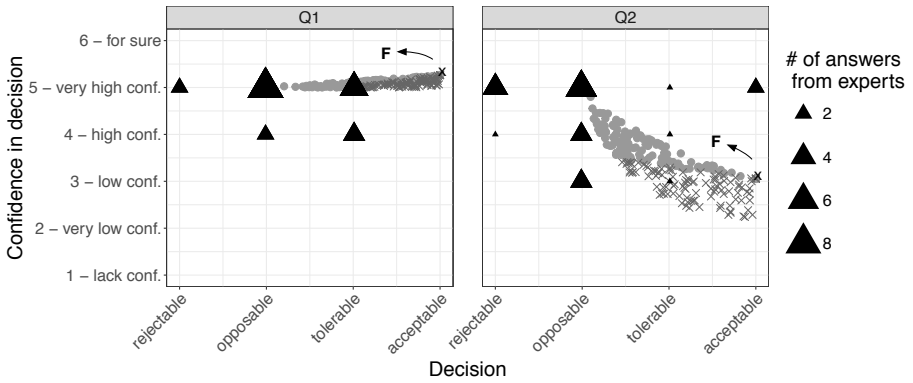
correspond to the two inputs of B and C, denoted as questions Q1 and Q2. According to the proposed assessment approach, we calculate the values of $(dec_A, conf_A)$ from inputs $(dec_B, conf_B)$ and $(dec_C, conf_C)$. The values are then plotted in the evaluation matrix. The solid dots represent the 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 case of complementary argument: *fully complementary argument*.

Concerning the expert data, the results collected from the questionnaire are processed first in order to remove the outliers. Then, they are depicted with triangles in the evaluation matrix (Fig. 12) and compared with the theoretical data. In this figure, the data of question Q1 and Q2 of the Argument 1 are presented. The size of the triangle indicates the number of respondents giving the same opinion.

The values to be estimated are deduced by comparing the distributions of theoretical data and expert data by some statistics means. We discover that experts have potential preference over the argument types, because different tendencies for decision-making, towards *rejectable* or *acceptable* for different arguments with the same inputs, are revealed. Moreover, the behaviours of two aggregation rules also have opposite inclinations. The complementary rule trends to produce "negative" results as the grey cloud locates towards *rejectable* (Fig. 12 a); whereas the redundant rule is subject to more "positive" results, as the grey cloud locates towards *acceptable* (Fig. 12 b). Hence, the experts' answers situating exactly in the clouds of complementary argument and the ones more "negative" than the clouds (at the left of the cloud) are considered for the preference for the complementary argument. Similarly,




a. Responses from experts vs. Calculated results based on Complementary Argument




b. Responses from experts vs. Calculated results based on Redundant Argument

Figure 12: Experts responses of Argument 1 and theoretical data

the experts' answers situating exactly in the clouds of redundant argument and the ones more "positive" than the clouds (at the right of the cloud) are considered for the preference for the redundant argument. The *type preference* is calculated as a rate over all the total number of the answers. The analysis results are summed up in the Table 3.

Moreover, the weights of B and C are derived by solving the Eq. 9 13 or Eq. 10 13 based on the mean values of experts *decision* ($dec_A$) and *confidence in this decision* ($conf_A$). 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 cloud of the corresponding argument type.

The deduction of the argument type is based on the weight values and the type preference. As introduced in Section 4.2, $1 - w_B - w_C$ (i.e. $w_{B \times C \to A}$ or $w_{B+C \to A}$) represents the degree of the complementarity or redundancy of an argument. Especially,

Table 3: Calculation results for the parameter estimation

| Arg. | Mean of $dec_A$ | Mean of $conf_A$ | Complementary | | | Redundant | | | Expected arg. types | Validated arg. types |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | $w_B$ | $w_C$ | Type preference | $w_B$ | $w_C$ | Type preference | | |
| Arg1 | 0.36 | 0.68 | 0.8 | 0.2 | 81.8% | 0.8 | 0.2 | 77.3% | C-Arg. | D-Arg. |
| Arg2 | 0.41 | 0.66 | - | - | 82.6% | 0.7 | 0.2 | 82.6% | R-Arg. | R-Arg. |
| Arg3 | 0.33 | 0.68 | 0.7 | 0.1 | 83.3% | - | - | 70.8% | C-Arg. | C-Arg. |
| Arg4 | 0.36 | 0.49 | 0.3 | 0.4 | 88.0% | - | - | 64.0% | D-Arg. | C-Arg. |

when $1 - w_B - w_C = 0$, the argument is a disparate argument. Thus, Argument 1 is deduced as a disparate argument (a special case for both complementary and redundant arguments). Comparing the "expected argument types" with "validated argument types", Argument 4 is considered as "complementary argument" rather than the "disparate argument" based on expert responses.

The confidence assessment and decision-making are believed to be subjective. However, the obtained answers from experts are more concentrated 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 of different techniques or sub-goals to the top goal and also the way that the sub-goals support the top goal. The results derived from 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 estimated the parameters of the 4 argument examples including argument types and the disjoint contributing weights.

## 5. 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 Fig. 7 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 Electronic Control Units (ECUs) with embedded software, 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.

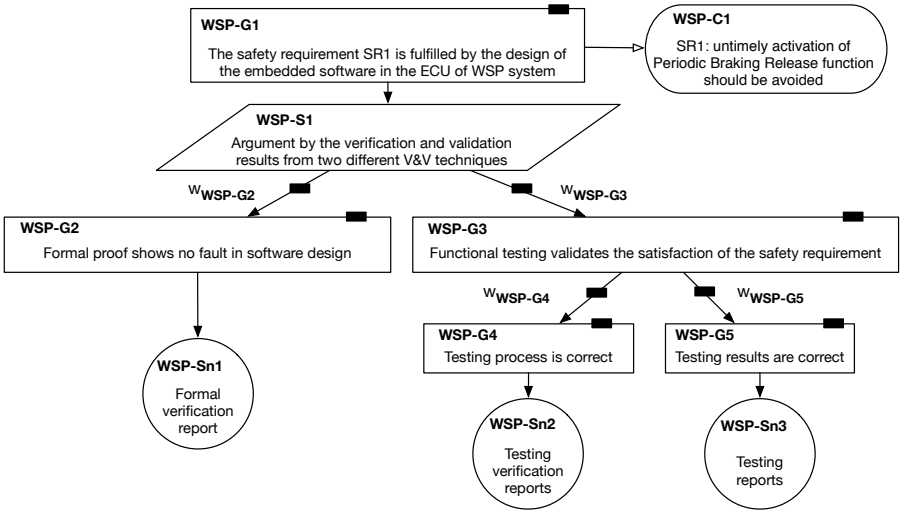



Figure 13: Safety argument fragment of WSP system

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 shall be 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 cause untimely activation of periodic braking release. 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 Fig. 7):

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. Assume that the formal verification and functional testing are sufficient for the justification. Therefore, the safety evidence and arguments are illustrated in the GSN model in Fig. 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 Fig. 10.

- 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 3). Thus, we directly use the results and illustrate the parameters related to WSP safety argument in Table 4.

Table 4: Argument types and weights for safety argument of WSP system

| Argument | Type | Weights | |
|---|---|---|---|
| Argument 2 | Redundant | $w_{WSP-G2}$ 0.7 | $w_{WSP-G3}$ 0.2 |
| Argument 3 | Complementary | $w_{WSP-G4}$ 0.7 | $w_{WSP-G5}$ 0.1 |

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.

We suggest performing a sensitivity analysis using a tornado graph as presented in Fig. 14. It is a simple statistical tool, which shows the positive or negative influence of basic elements on a 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 $x_i \in [X_{min}, X_{max}]$, 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 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 5. 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 remained at 0.6 to ensure the scale of $(Dec, Conf)$ to the great extent for a better presentation of the tornado graphs.

Table 5: Example of values and intervals for sensitivity analysis

| | $Dec_{WSP-G2/G4/G5}$ | $Conf_{WSP-G2/G4/G5}$ |
|---|---|---|
| Basic value $X_i$ | 0.67 (Tolerable) | 0.6 (High confidence) |
| Intervals $[X_{min}, X_{max}]$ | [0,1] | [0.6,1] |

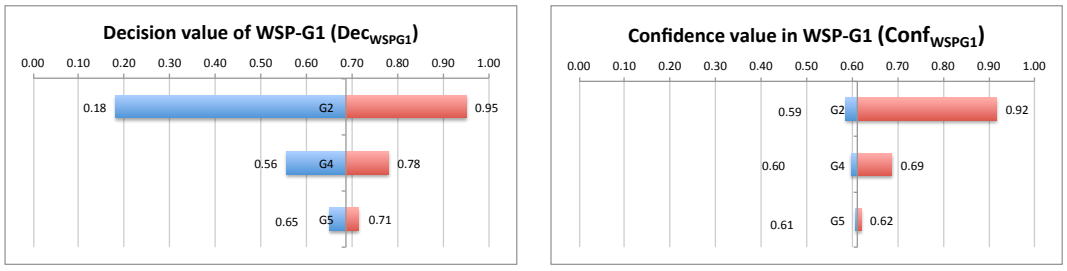



Figure 14: Tornado diagram of confidence assessment for WSP safety argument

With the basic values above, $(Dec_{WSP-G1}, Conf_{WSP-G1})$= (0.69,0.60). 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 values $(Dec_{WSP-G4/G5}, Conf_{WSP-G4/G5})$ as the basic values. Then, we calculate the values for WSP-G1 with $(Dec_{WSP-G4/G5}, Conf_{WSP-G4/G5})$= (0,0.6) or (1,1). We obtain the values $Dec_{WSP-G1}$= [0.18,0.95] for decision value of WSP-G1 and $Conf_{WSP-G1}$= [0.59, 0.92] for confidence value in this decision. The same approach is applied for other parameters; the results are presented in Fig. 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 the paper, as it is an exploitation of the quantitative framework that we propose.

## 6. Related work

Most safety standards are organized around checklists of best practices associated with Safety Integrity Levels. Nevertheless, they do not convey how these practices lead to high-level safety properties. Such a statement is also done by Knight and Rowanhill (2016), who introduced the concept of "rationalized standard" to make explicit this rationale using GSN. Even if their objective is similar to ours, no practical contribution of the proposed approach is presented. Moreover, nothing is introduced about confidence in the argumentation.

Several researches dedicated to application domains have been published on this issue. For instance, in the aeronautics domain, Holloway (2013) presents several works addressing the issue of making explicit the rationale in the previous version

of the software standard DO178-B:1992 and then in the current version DO178-C: 2012. He also develops GSN models (extended in Holloway (2015)) to make more explicit the rationale used for justifying that a required level of integrity has been obtained. However, it is difficult in this contribution to identify a mapping between this model and the DO178C document elements. Moreover, nothing is proposed for assessing the confidence in a high-level safety objective. This confidence assessment of a DO178-C compliant GSN is addressed in Wang et al. (2016a) but only simple high-level examples are provided without any validation.

A similar approach has been done in the automotive domain, for the ISO 26262 (2011) standard. Several research papers have been published making explicit the rationale of the standard using GSN. For instance, Birch et al. (2013) present several GSN diagrams to make explicit the link between high-level safety requirements (safety goals) and low level evidences. In this work, the authors do not address the confidence estimation, but they draw several important conclusions about using GSN in certification argumentation. One is that using GSN is not always the most effective mean of expression. Sometimes, tables can be used. Rushby (2015) does the same analysis in the context of the DO178, and he proposed a textual notation to annotate goals and evidences (sub-goals are referenced as premises). In the context of certification, GSN may be not the most convenient way to express argument. Nevertheless, this is an open issue not explained in this paper.

Compared with other application domains, to our knowledge, very few works have been done for safety standards in the railway domain. Gallina et al. (2016) used GSN in the context of railway certification, but only to justify the use of a development process and not for explaining the rationale of the standard itself. Müller et al. (2010) expresses the issue of using GSN to make explicit the rationale, but very few details are presented even in the associated project FP7-INESS (2008-2011).

Most of the previous works mainly address the implicit rationale in the standards, and use GSN to make them explicit. They are preliminary works, and even if some rationales are really comparable (close to GSN diagrams), no generic GSN patterns can be extracted for now. This is the reason why we propose our own interpretation using GSN for the railway domain. Moreover, the previous papers do not address the issue of the assessment of the confidence, whereas it is the underlying concept in every safety case analysis.

This issue has been widely addressed but not in the context of standards. First approaches are focusing on qualitative assessment of the confidence in a safety case. Ayoub et al. (2012) and Hawkins et al. (2011) focus on the identification of "defeaters" of an argument, and the construction of an additional argument dedicated to confidence. Considering that safety cases might become more and more complex

and oversize, it is not possible to qualitatively analyse the confidence in a complete safety case without having some quantitative facilities. Hence, even if it was not the original objective of the inventors of the safety case notation GSN, it is now admitted that confidence should be addressed quantitatively. Most of the approaches are based on probabilistic theory to quantify the confidence in the top goal of a GSN. For instance, we can cite Denney et al. (2011) working with Bayesian Network, and Cyra and Gorski (2011) using belief function theory, or a mix by Guiochet et al. (2015). As presented in the review and analysis paper of Graydon and Holloway (2017), many other approaches are studied for quantitative assessment of safety argument confidence. The authors of this paper conclude that whereas quantitative approaches for confidence assessment are of high interest, no method is currently fully applicable mainly due to the absence of methodology to estimate parameters for the confidence propagation among different layers of goals. We believe that our approach with the help of the experts' judgement extraction activity will contribute to reduce this limitation.

The quantitative assessment framework proposed in this paper could be compared to studies like the paper of Nair et al. (2015), but only for the expert judgement extraction. They also provide confidence propagation calculation rules based on belief theory. However, the judgement extraction is not based on the same parameters as ours, and in particular they do not address the decision judgement regarding the elements of a GSN, which is a fundamental point in our framework. A common pitfall of 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?).

A quite similar approach is presented by Ayoub et al. (2013). They introduce four argument types and formulas to combine confidence based on belief theory for calculation. They do not use the term confidence, but each goal is assessed with a belief, disbelief and uncertainty estimation for the statement. Some argument types are comparable with our proposal (e.g., their "Alternative" is near our "Redundant"), but they do not provide any justification of the combining formulas. Moreover, no intuitive interpretation of the parameters of the aggregation rules is provided. As in the previous work, the results do not provide any justification for a decision regarding the acceptability of the safety case.

The closest work can be found in the publication of Cyra and Gorski (2011). As already mentioned, we reuse their decision and confidence scales and the transformation rules into belief theory parameters (belief, disbelief, and uncertainty). Compared with our approach, they did not use GSN for safety case modelling as we do; and we also developed our own argument types and associated formulas. Indeed, they extended the work from Govier (2013) to propose 6 types of arguments. We found

them too complex for an intuitive identification in a real safety case. Moreover, according to each of these types, their parameters are difficult to determine and interpret. Our objective is really to provide an efficient and pragmatic approach for analysts; thus we only propose 2 types of argument, with a dedicated approach for the weight parameter identification based on an expert judgement extraction.

## 7. Conclusion

One main issue in the context of the safety assessment of railway critical systems is effectively measuring the confidence in the development process that is documented in the safety case. Firstly, we propose to make explicit the rationale of the railway standard EN50129 in terms of safety justification with the safety case. An intermediate part of the GSN safety case is built to fill the inference gap between the high-level argumentation structure and safety evidence required by this standard. Then, we propose a quantitative framework to assess the confidence in the structured safety cases. Two types of arguments are put forward in this paper: complementary and redundant arguments. A safety case model under this framework includes a reasonable structure of argumentation and parameters, i.e. the weights of sub-goals and the argument types. Once the quantified argument model is built, it only requires evaluating the trustworthiness of the lowest-level of the argument; then these values are aggregated to estimate the trustworthiness in the top goal. The quantitative aggregation rules were developed based on the D-S theory in the previous work Wang et al. (2016b). An intuitive evaluation matrix used in the framework is adopted from Cyra and Gorski (2011) for extracting expert opinions on arguments.

A preliminary case study is carried out via a survey amongst experts including safety engineers, safety assessors and researchers in the domain of dependability. This case study aims to estimate the parameters of arguments, which is an essential prerequisite for applying our confidence assessment framework. 35 responses of the survey have been received. We compare this expert data with the theoretical data derived from the aggregation rules. The consensus among experts on the variation of the contributing weights of different techniques or sub-goals to the top goal and also the way in which they support the top goal, is delivered. The disjoint contributing weights and argument types are deduced based on statistical analysis of the answers. The analysis results are mostly within our expectation. This gives a first validation of the proposed confidence assessment framework. With these results, we provide a 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. Referring to a complex safety case in railway domain,

considerable argument branches are involved. Nevertheless, it is still feasible to build the safety case model and estimate the parameters for common safety argument patterns, because these models are then reusable. The confidence framework can provide the efficient and effective assessment based on these quantified argument models. On the other hand, the massive documents of the safety case might also be a reason against the graphical representation of the safety arguments. The textual documentation with the goal-structured arguments can be an alternative for our confidence framework notation. Furthermore, a comprehensive sensitivity analysis for complex safety arguments, considering the influence of the structure and parameter changes, is our future work.

## References


Allotta, B., Conti, R., Meli, E., Pugi, L., Ridolfi, A., 2013. Development of a hil railway roller rig model for the traction and braking testing activities under degraded adhesion conditions. International Journal of Non-Linear Mechanics 57, 50–64.

Ayoub, A., Chang, J., Sokolsky, O., Lee, I., 2013. Assessing the overall sufficiency of safety arguments. In: 21st Safety-Critical Systems Symposium (SSS'13). pp. 127–144.

Ayoub, A., Kim, B., Lee, I., Sokolsky, O., 2012. A systematic approach to justifying sufficient confidence in software safety arguments. In: Computer Safety, Reliability, and Security. Springer, pp. 305–316.

Birch, J., Rivett, R., Habli, I., Bradshaw, B., Botham, J., Higham, D., Jesty, P., Monkhouse, H., Palin, R., 2013. Safety cases and their role in ISO 26262 functional safety assessment. In: Computer Safety, Reliability, and Security (SAFECOMP). Springer, pp. 154–165.

Bishop, P., Bloomfield, R., 1998. A methodology for safety case development. In: Industrial Perspectives of Safety-Critical Systems. Springer, pp. 194–203.

Bloomfield, R., Littlewood, B., Wright, D., 2007. 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, pp. 338–346.

Bloomfield, R. E., Guerra, S., Miller, A., Masera, M., Weinstock, C. B., 2006. International working group on assurance cases (for security). Security & Privacy, IEEE 4 (3), 66–68.

Cyra, L., Gorski, J., 2007. Supporting compliance with security standards by trust case templates. In: Dependability of Computer Systems, 2007. DepCoS-RELCOMEX'07. 2nd International Conference on. IEEE, pp. 91–98.

Cyra, L., Gorski, J., 2011. Support for argument structures review and assessment. Reliability Engineering & System Safety 96 (1), 26–37.

Denney, E., Pai, G., Habli, I., 2011. Towards measurement of confidence in safety cases. In: Empirical Software Engineering and Measurement (ESEM), 2011 International Symposium on. IEEE, pp. 380–383.

Dubois, D., Prade, H., 2009. Formal representations of uncertainty. Decision-Making Process: Concepts and Methods, 85–156.

EN50126, 1999. Railway applications - The specification and demonstration of Reliability, Availability, Maintainability and Safety (RAMS). CENELEC, European Committee for Electrotechnical Standardization.

EN50128, 2011. Railway applications - Software for railway control and protection systems. CENELEC, European Committee for Electrotechnical Standardization.

EN50129, 2003. Railway applications - Safety related electronic systems for signaling. CENELEC, European Committee for Electrotechnical Standardization.

EN50159, 2010. Railway applications - Safety related communication in transmission systems. CENELEC, European Committee for Electrotechnical Standardization.

ERA, E. R. A., 8 2015. Regulation 2015/1136/eu on the common safety method for risk evaluation and assessment.

FP7-INESS, 2008-2011. INtegrated European Signalling System. Project supported by the European Commission under the 7th Framework Programme, `www.iness.eu`, accessed: 2017-07-10.

Gallina, B., Gómez-Martínez, E., Earle, C. B., 2016. Deriving Safety Case Fragments for Assessing MBASafe's Compliance with EN 50128. Springer International Publishing, Cham, pp. 3–16.

Govier, T., 2013. A practical study of argument. Wadsworth, Cengage Learning.

Graydon, P. J., Holloway, C. M., 2017. An investigation of proposed techniques for quantifying confidence in assurance arguments. Safety Science 92, 53 – 65.

Guiochet, J., Do Hoang, Q. A., Kaaniche, M., 2015. A model for safety case confidence assessment. In: Computer Safety, Reliability, and Security (SAFECOMP). Springer, pp. 313–327.

Hawkins, R., Kelly, T., Knight, J., Graydon, P., 2011. A new approach to creating clear safety arguments. In: Advances in systems safety. Springer, pp. 3–23.

Holloway, C. M., 2013. Making the implicit explicit: Towards an assurance case for DO-178C. In: Proceedings of the 31st International System Safety Conference (ISSC).

Holloway, C. M., 2015. Explicate '78: Uncovering the Implicit Assurance Case in DO-178C. In: Safety-Critical Systems Symposium 2015 (SSS 2015), Bristol, United Kingdom.

IEC61508, 2010. Functional safety of electrical/electronic/programmable electronic safety-related systems. International Electrotechnical Commission (IEC).

ISO 26262, 2011. Software considerations in airborne systems and equipment certification. International Organization for Standardization (ISO).

Jøsang, A., 2001. A logic for uncertain probabilities. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 9 (03), 279–311.

Kelly, T., 1998. Arguing safety - a systematic approach to safety case management. Ph.D. thesis, Department of Computer Science, University of York.

Kelly, T., McDermid, J., 1997. Safety case construction and reuse using patterns. In: Computer Safety, Reliability, and Security (SAFECOMP). Springer, pp. 55–69.

Kelly, T., Weaver, R., 2004. The goal structuring notation–a safety argument notation. In: Proceedings of the Dependable Systems and Networks (DSN) workshop on assurance cases.

Knight, J. C., Rowanhill, J., 2016. The Indispensable Role of Rationale in Safety Standards. Springer International Publishing, Cham, pp. 39–50.

Müller, J. R., Zheng, W., Schnieder, E., 2010. The Improvement Of The Safety-case Process In Practice: From Problems And A Promising Approach To Highly Automated Safety Case Guidance. In: International Conference on System Design and Operation in Railways and other Transit Systems (COMPRAIL), Computers in Railways XII. Beijing, China.

Nair, S., Walkinshaw, N., Kelly, T., de la Vara, J. L., Nov 2015. An evidential reasoning approach for assessing confidence in safety evidence. In: 2015 IEEE 26th International Symposium on Software Reliability Engineering (ISSRE). pp. 541–552.

Pugi, L., Malvezzi, M., Tarasconi, A., Palazzolo, A., Cocci, G., Violani, M., 2006. Hil simulation of wsp systems on mi-6 test rig. Vehicle System Dynamics 44 (sup1), 843–852.

Rushby, J., 2015. The interpretation and evaluation of assurance cases. Tech. Rep. SRI-CSL-15-01, Computer Science Laboratory SRI International, Menlo Park CA 94025, USA.

Shafer, G., 1976. A mathematical theory of evidence. Vol. 1. Princeton university press Princeton.

Smets, P., 1992. 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., pp. 292–297.

Wang, R., 1 2017. Questionnaire for safety argument assessment research. Https://goo.gl/forms/V3vMnl59cTWA6Lws2.

Wang, R., Guiochet, J., Motet, G., 2016a. A framework for assessing safety argumentation confidence. In: International Workshop on Software Engineering for Resilient Systems. Springer, Gothenburg, Sweden, pp. 3–12.

Wang, R., Guiochet, J., Motet, G., 2017. Confidence assessment framework for safety arguments. In: SafeComp 2017. Accepted.

Wang, R., Guiochet, J., Motet, G., Schön, W., 2016b. DS theory for argument confidence assessment. In: International Conference on Belief Functions, (BELIEF2016), Prague, CZ. Springer, pp. 190–200.