An Evidential Reasoning Approach for Assessing Confidence in Safety Evidence

Sunil Nair[1], Neil Walkinshaw[2], Tim Kelly[3], and Jose Luis de la Vara[1]

[1]Certus Centre for Software V&V
Simula Research Laboratory, Norway
{sunil,jdelavara}@simula.no

[2]Department of Computer Science,
University of Leicester, United Kingdom
n.walkinshaw@mcs.le.ac.uk

[3]Department of Computer Science,
University of York, United Kingdom
tim.kelly@york.ac.uk

***Abstract* — Safety cases present the arguments and evidence that can be used to justify the acceptable safety of a system. Many secondary factors such as the tools and technique used to create the evidence, and the experience of the evidence creator, can affect the assessor's confidence in the evidence cited by a safety case. One means of reasoning about the confidence established in the evidence is to present an explicit confidence argument that corroborates the reason for having confidence on the evidence. In this paper, we propose a novel approach to automatically construct these confidence arguments through asking assessors to provide individual judgements concerning the trustworthiness of evidence and the appropriateness of its use in supporting the case. These judgements can be supported by further evidence, simply asserted, or expressed with stated uncertainty. The proposed approach enables these judgements to be presented within the context of an overall argument of confidence, and a quantified aggregate of the overall confidence to be derived. The approach is based on *Evidential Reasoning* – a decision-theoretical technique for reasoning about uncertainty and evidence. Our approach enables assessors to clearly present complex reasoning concerning evidence whilst making any doubt or uncertainty explicit. The proposed approach is supported by a prototype tool (EviCA) and is evaluated using the Technology Acceptance Model.**

Keywords: safety case; safety evidence; confidence argument; uncertainity; evidential reasoning; expert judgement

I. INTRODUCTION

Goal-based system safety standards such as DS 00-56 (MoD 2004a) often require the construction and provision of a safety case. A safety case is defined as "*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"* [1]. A structured safety argument explains how the available body of evidence supports the overall claim of acceptable safety.

Inevitably, both argument and supporting evidence are typically imperfect. Often it is left to the human assessor to decide if the presented evidence is sufficient to support the safety claims made in the case. A survey on the state of the practice of evidence management suggests that expert judgement is the most commonly used technique to assess safety evidence [31]. Determining the type or amount of evidence required to satisfy a claim can be difficult. Both the developer and the assessor may be uncertain about attributes of the evidence that is provided. The higher the uncertainty on the sufficiency of the evidence to support a claim, the lower the confidence in the overall safety case provided.

There are a number of secondary factors that may influence the assessor's confidence in the evidence provided. Our previous work towards understanding how safety experts assess safety evidence shows that the assessment process varies substantially from expert to expert, and that important safety assessments are frequently based on subjective evaluations [4]. The results indicated that experienced experts often initially form overall opinions (beliefs) regarding the completeness and appropriateness of the evidence in a given scenario. On further questioning, we identified that these subjective beliefs were based on many factors related to the process of the evidence creation, the techniques used, the people involved and certain characteristics of the evidence itself. The study allowed us to identify a set of such generic secondary factors related to the evidence under assessment that frequently influenced the expert's decision on the acceptance of the evidence.

In current practice, such reasons for establishing confidence in the evidence remain implicit in the assessment process. Moreover, the uncertainties associated to the different reasons for having confidence and thereby on the appropriateness and trustworthiness of the evidence is also implicit. For example, software testing evidence may not be sufficient to support a claim about system safety due to a number of reasons like the use of a faulty test oracle or the tester inadvertently testing a different version of the system. The knowledge gaps that prohibit perfect (100%) confidence in a safety argument can be described as 'assurance deficits'. In order to gain complete confidence in the evidence, these uncertainties must be identified and managed explicitly.

One approach of reasoning about the confidence established in the safety evidence is to build an explicit secondary confidence argument [2]. The role of the confidence argument is to explicitly detail the various reasons for having confidence in the evidence. There may be uncertainties associated with aspects of the evidence provided. The use of a secondary confidence argument may be beneficial for both the developer and the assessor. For system developers, making explicit the different factors that provide confidence in the evidence can help in producing stronger safety cases. For assessors, an explicit confidence argument can help focus on aspects that are weakly supported.

In this paper, we propose a novel approach to automatically construct confidence arguments for the evidence cited in the primary argument and quantify confidence using Evidential Reasoning (ER) [3]. The theory behind ER is concerned with the

challenge of taking expert´s subjective beliefs and combining them to form an aggregate, so that all of the individual confidence factors related to the evidence are taken into account. This paper makes the following specific contributions:

- We build upon our previous work [4] and analysing various evidence-type specific assessment checklists to propose a confidence argument pattern that explicitly details the various reasons for having confidence in a particular atomic piece of evidence.
- We present a technique, based upon ER [3], by which the low-level confidence information (on individual factors) can be propagated to a macroscopic safety claim, encompassing the appropriateness and trustworthiness of the evidence. This (1) explicitly details the reason for having confidence in the evidence, (2) captures any uncertainties associated with the evidence assessment, and (3) presents the confidence at each level both quantitatively and visually.
- We present an implementation of the approach named *EviCA (**Evi**dence **C**onfidence **A**ssessor*), an Eclipse-plugin that allows users to build primary arguments and assess safety evidence associated to the claim. The tool automatically builds and presents a confidence argument structure for the evidence and quantifies the uncertainties and confidence value.
- We provide a user-evaluation of the proposed approach and its tool support, using the Technology Acceptance Model [5].

The remainder of the paper is organized as follows. Section II presents the background. Section III describes the proposed approach including the confidence argument. Section IV presents the tool support and its features. Section V presents the evaluation method and its results. Section VI presents related work and finally, Section VII concludes the paper and presents future directions. Appendix A presents the questionnaire used to evaluate the technology and the tool support.

II. BACKGROUND

This section provides a brief overview of the relevant background information required to understand our proposed approach.

A. Safety Cases

The definition of a safety case provided by (MoD 2004a) and presented in the previous section highlights that both argument and evidence are crucial elements of the safety case that must go hand-in-hand. An argument without supporting evidence is unfounded, and therefore unconvincing. Evidence without argument is unexplained – i.e. it can be unclear whether (or how) safety claims have been substantiated. An explicit argument is required in order to communicate the relationship between the evidence and safety objectives. The Goal Structuring Notation (GSN) [25] - a graphical argumentation notation - explicitly represents the individual elements of any safety argument (requirements, claims, evidence and context) and (perhaps more significantly) the relationships that exist between these elements (i.e. how individual requirements are supported by specific claims, how claims are supported by evidence and the assumed context that is defined for the argument). When the elements of the GSN are linked together in a network they are described as a '*goal structure*'. The principal purpose of any goal structure is to show how goals (claims about the system) are successively broken down into sub-goals until a point is reached where claims can be supported by direct reference to available evidence (solutions). As part of this decomposition using the GSN, it is also possible to make clear the argument strategies adopted (e.g. adopting a quantitative or qualitative approach), the rationale for the approach and the context in which goals are stated (e.g. the system scope or the assumed operational role).

Hawkins et al. [2] proposes that the arguments within safety cases can be usefully defined in terms of two separate but interrelated arguments:

- A *safety* (or *technical risk*) argument that documents the arguments and evidence used to establish direct claims of system safety
- An accompanying *confidence* argument that justifies the sufficiency of confidence in this safety argument.

The technical risk argument must decompose the overall claim of acceptable safety into arguments that justify the acceptability of the risk posed by identified system hazards. For each hazard, the argument states what 'adequately' addressed means for that hazard and then identifies the evidence supporting the conclusion. This structure explains the purpose of each piece of evidence. A confidence argument records the justification for confidence in a safety argument. There will be uncertainties associated with aspects of the safety argument or supporting evidence. The role of the confidence argument is to explicitly address both the positive factors that establish confidence and these uncertainties. These uncertainties can be termed *assurance deficits*. This paper focuses on the creation and use of confidence arguments concerning the *evidence* that is cited in any technical risk argument. Each time evidence is referenced as a solution (i.e. evidence) in the technical risk argument, an assertion is being made that the evidence being put forward is sufficient to support the claim. The assurance of the solution depends upon the confidence that the evidence is appropriate to support the claim, and the evidence is trustworthy.

GSN can be used to record both the technical risk argument and the accompanying confidence argument. Common structures in safety case arguments can be reused through their documentation as 'Safety Case Patterns'. Annex A to Part 1 of the GSN Standard [25] describes extensions to GSN that enable the description of reusable, generic, argument structures.

B. *Evidential Reasoning*

As mentioned previously, the assessment of complex systems usually has to be decomposed in to various sub-claims of safety. The supplier might not always have *sufficient* evidence to produce arguments in which they can be 100% confident. In the worst case, they might not have any evidence at all (for example there might be no test-logs), rendering them incapable of drawing any justifiable conclusions about particular aspects of the system.

Evidential Reasoning (ER) [3] provides a technique to assimilate assessment based on various sub-claims into a single, coherent assessment. ER considers a hierarchy of `attributes' by which some system is to be assessed. Each attribute (e.g. personnel/team) can be subdivided into a set E of lower-level sub-attributes $\{e_1, \cdots, e_n\}$ (e.g. competence and domain knowledge). Each sub-attribute e_i can also be given a weight w_i representing the relative importance, such that $\sum_{i=1}^{n} w_i = 1$ (by default the weight is evenly distributed across all attributes as $1/n$).

A human expert assesses each of the lowest-level attributes by providing the subjective belief they have on the satisfaction of the attribute. The belief is provided as a distribution over a Likert-scale consisting of g grades $H =< H_1, \cdots, H_g >$ (i.e. if $g = 5$, H_1 and H_5 might correspond to "very poor" and "excellent" respectively). The distribution of probability is referred to as *'Belief Function'* – a term will be adopted throughout the paper. The fact that a belief-function is a distribution makes it possible to accommodate uncertainty. So, instead of stating categorically that their assessment (e.g. of the competency of a team) is "excellent", they might choose to suggest that they have a confidence of 50% for "excellent", and 50% for "good".

Formally, the expert´s confidence that a particular attribute e_i achieves a grade H_n is denoted $\beta_{n,i}$. For a given attribute, $\sum_{i=1}^{n} \beta_{n,i} \leq 1$. Thus, an expert´s complete assessment of attribute e_i (encompassing all possible grades) can be expressed as the distribution:

$$S(e_i) = \{(H_n, \beta_{n,i}), n = 1, \cdots, g\}$$

A key feature of ER is that, alongside uncertainty, it is also possible to capture complete ignorance on the part of the assessor. If they, for example, do not know anything about the development team, they are not forced to provide any assessment at all. The beliefs in $\beta_{n,i}$ do not have to sum up to 1 for a given attribute (as would be expected with conventional Bayesian probabilities). The sum of beliefs $\sum_{i=1}^{n} \beta_{n,i}$ can be interpreted as their overall confidence of the assessment, where a sum of 1 amounts to total confidence, and a sum of 0 amounts to total ignorance (no confidence).

Given a hierarchy of attributes, where the lowest-level attributes are associated with distributions corresponding to the assessments as presented above, ER presents a technique to assimilate them. Distributions of `beliefs' are propagated up from lower-level nodes to higher-level nodes, and are combined with the distributions from their sibling nodes to produce a representative macroscopic assessment.

Crucially, this process of propagation from basic attributes to aggregated result y obeys certain desirable axioms that ensure the following [3]:

1. y must not be assessed to a grade H_n if none of its basic attributes is assessed to a grade H_n.
2. y should be precisely assessed to a grade H_n if all of its basic attributes are precisely assessed to a grade H_n.
3. If all of the basic attributes are completely assessed to a subset of evaluation grades then y should be assessed to the same subset of grades.
4. If an assessment of any basic attribute is incomplete, then the assessment for y should also be incomplete to a certain degree.

III. PROPOSED APPROACH

The approach proposed in this paper automatically builds a confidence argument that explicates and justifies the confidence in the cited evidence by explicitly detailing the various confidence factors associated to the evidence. To explain how the approach works we consider a running example in the following section.

Let us consider a primary argument that claims (PG1), *"All Hazards related to System X have been identified and recorded"*. The claim is supported by the citing the evidence (AS1) *hazard log* as an asserted solution. When evidence is cited in a safety case it is typically asserted that the evidence presented is sufficient to support a claim. The truth of this assertion depends on the appropriateness and trustworthiness of this evidence (in this case the hazard log). A secondary confidence argument provides the explicit justification for having sufficient confidence in the evidence. To indicate the assertion in the safety argument that the confidence argument is associated with, the confidence argument is tied to an *Assurance Claim Point* (ACP). An ACP is indicated in GSN with a named black rectangle on the relevant link. A confidence argument is developed for each ACP. Figure 1.a shows the primary argument depicted in GSN with ACP1 between the goal and the solution.

There are many secondary factors that contribute to the confidence in the evidence and any uncertainty that might exist in demonstrating sufficient confidence must be identified and acknowledged. The confidence argument relies on various factors

that were involved in the creation of the evidence (e.g., the process/techniques used, the personnel/teams involved, the tools used, etc.) and the role of the evidence in the particular argument. Through systematic analysis of the practice of evidence assessment (via surveys and interviews with safety experts), we identified a number of such factors that influence confidence in the evidence [4].

We exploit this knowledge in our approach to propose a confidence argument pattern that enables the abstract notion of overall confidence to be broken down into two constituent properties namely trustworthiness and appropriateness. We then further break the factors into sub-factors and allow quantifying the relative weights that each factor plays in providing the overall confidence. Figure 1.b shows a skeleton of the confidence argument that details some of the factors (above dashed line) that may be considered to demonstrate sufficient confidence in the hazard log (the original confidence argument structure is explained in the next section). As noted in Figure 1, trustworthiness of the hazard log is broken into the process employed to create the hazard log, the personnel who carried out the activity and the tools used to create the log. The personnel factor is further broken into his/her competence and the guarantee of independence (e.g., the personnel who verified the log was independent of the creator of the log). Similarly, the appropriateness of the hazard log to the claim PG1 is decomposed into the intent of the hazard log and the asserted role it plays in the argument. The relative (hypothetical) weight that each factor plays is denoted on the edge.

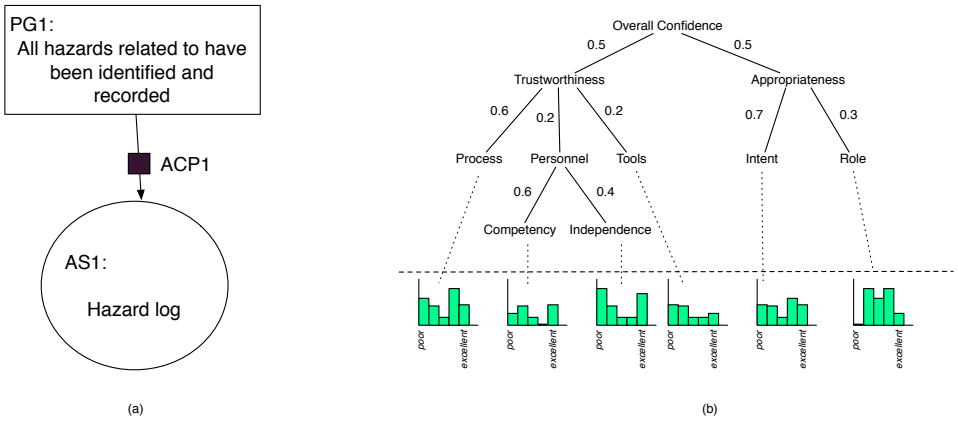


Figure 1.a. ACP relating to asserted solution (hazard log); 1.b. Confidence factors associated to the hazard log (above dashed line), manual subjective assessment (below dashed line).

This break-down into individual confidence-factors (detailed in Section III.A) enables us to gauge the assessor's subjective impression with respect to each factor, via a series of questions. One way of reasoning the confidence factors is through evidence-type specific checklists. A previous survey on the state of the practice of evidence management shows that evidence assessment is predominantly carried out with the help of checklists [31]. In addition, interviews with experts also indicate that checklist based assessments are common practice in industry [4]. Our approach allows the use of type-specific evidence assessment checklists as a basis for collecting belief functions that represent subjective assessments for the lowest level factors in the confidence argument skeleton (shown in Figure 1.b, below dashed line).

The subsequent process of confidence quantification is discussed in Section III.B. The ER approach incorporates uncertainties stated as part of the answer to a particular question i.e., if there is a certain amount of uncertainty or ignorance about a particular question relating to a factor, it can still be included in the form of belief functions. The ER algorithm combines the lower-level belief functions for each question related to the factor, and propagates the belief functions up to the parent factor. This automatic propagation happens till the top most parent factor is reached yielding a general belief function. The end result is a confidence argument structure that graphically depicts the various confidence factors associated to the evidence with a quantified confidence value based on lower-level subjective assessments at each claim level.

A. Confidence Argument Pattern

The confidence argument pattern is represented using the GSN pattern extensions [23]. Details about GSN pattern extensions can be found in [24][25]. To build the confidence argument pattern we followed four steps. The first step consisted of determining the criteria that should be considered for confidence assessment. We identified such criteria from the results of the interview study presented in [4], in which domain experts indicated the information that made them gain confidence in safety evidence. The criteria are defined later in this section. As a second step, we specified an initial structure (based on the results of step 1) for the argument pattern structured according to two main criteria for assessing confidence in safety evidence: *trustworthiness and appropriateness*. Each criterion was broken down into subsequent specific factors that affected their confidence. All of the authors reviewed the confidence argument pattern and discussed its structure and content. This resulted in an initial version of the pattern. To improve the pattern´s coverage of factors influencing assessor's

confidence, we analysed 16 checklists from the aerospace, avionics, railway and defence domains. Most of the checklists (13 out of 16) are in the public domain[†]. As a result, we identified three factors (*Bound Qualification, Scope for Document format and Expected Structure of the Evidence Type*) that had not been included in the initial pattern. Finally, we modified the argument pattern based on the outcome from the validation step. Again, all the authors reviewed and discussed the pattern.

We acknowledge that the completeness of the confidence argument pattern for all types of evidence cannot be guaranteed. We built the pattern based on a small subset of checklists that are used in practice and a relatively narrow range of domains. Nonetheless, our evaluation (Section V) suggests that the presented structure covers all major areas of concern relating to evidence confidence assessment. To accommodate situations where the criteria are incomplete, our approach and tool support allows users to add new factors and criteria to the existing proposed pattern. We provide means and guidance to do this (described in Section V).

Figure 2 shows our proposed confidence argument pattern expressed in GSN. The structure is broken into seven different parts for explanation purposes. Relating to the running example, the top most goal (G1) of the pattern describes that there is sufficient *overall confidence* in the evidence used as *asserted solution* (hazard log). Confidence can be defined using GSN context element (Con1) as a measure of the belief that the evidence cited for a particular claim is trusted for its integrity and it is appropriate for its intended purposes. However, assessment of a qualitative statement such "*sufficient confidence"* with regards to the evidence assertion is complex. Confidence is a quantifiable entity that relates to the probable truth of a claim. In our approach we encode the quantified confidence value calculated through the ER approach with a more assessable qualitative tag. Assuming that confidence can be quantified from a scale of 0-100%, we use the following scale to rate confidence: *0-20% Very Low Confidence, 20-40% Low Confidence, 40-60% Medium Confidence, 60-80% High Confidence*, and *80 -100% Very High Confidence.* The *overall confidence* claim is achieved by decomposing it into *trustworthiness* (G2) of the evidence and *appropriateness* (G3) of the evidence.

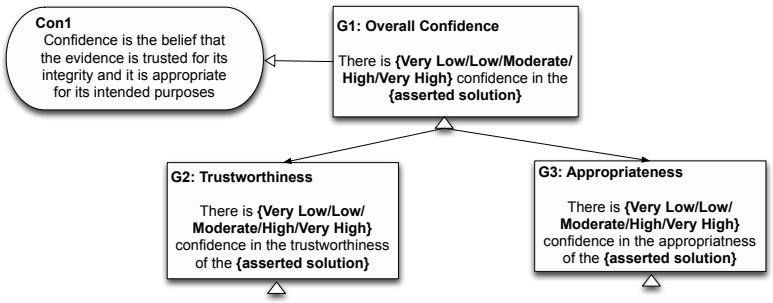


Figure 2. Overall confidence argument pattern for the asserted solution

The trustworthiness of the asserted solution (hazard log) often relates to freedom from flaw. In the legal domain, integrity of evidence is used to refer to its soundness or quality. In systems engineering domain, trustworthiness of the evidence often related to the processes used to generate the evidence [33]. In the pattern, we define trustworthiness (Con2) as the property of the evidence to provide trust or belief that evidence can be assured to be as specified. We decompose *trustworthiness* of the *asserted solution* as shown in Figure 3 into *Personnel, Process/Techniques, Tool Integrity, Content Compliance,* and *Evidence Past*. The pattern also allows users to define their own trustworthiness factors as denoted by the G8.

Each of above factors are further decomposed into sub-factors as follows:

- *Personnel* (G4) – Arguments regarding the personnel or the team(s) involved in the creation or verification of the asserted solution. As a strategy (S2), arguments can be made over each personnel factor that influences the overall confidence. There is a set of pre-defined factors (derived from previous study) in the pattern. In addition, as mentioned earlier, the pattern allows the addition of user-defined personnel factors (Con3).

[†] Public checklists have been collected and shared at
https://drive.google.com/a/simula.no/folderview?id=0B42RvDI04vjnbzgtM0RXRFJqZWs&usp=drive_web

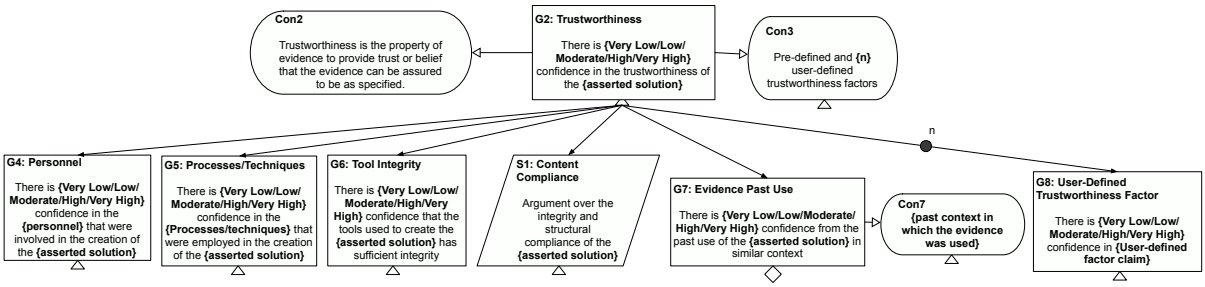


Figure 3. Trustworthiness argument pattern for the asserted solution

We make use of the multiplicity element in GSN to denote different user-defined personnel factors. Figure 4 shows the personnel argument pattern. For each personnel or team involved in the creation or verification of the asserted solution, we further decompose the arguments as:

- *Past Knowledge* (G12) – the confidence that the personnel or team(s) involved have any past knowledge about the creation or verification of the asserted solution in similar context. The similar context of the past knowledge (e.g., project details) should be explicitly shown in the argument structure by using the context Con14.
- C*ompetency* (G13) – the confidence that the personnel or team(s) involved in the creation of verification have the required skills, training and competency to produce the asserted solution. Any further evidence supporting this, e.g., developer resumes can be provided if available.
- *Independence* (G14) – the confidence that the different personnel or team(s) involved in the creation of verification of the asserted solution is independent of each other. This is to ensure that the person creating the hazard log is not the same as the one verifying the log.
- *Domain Experience* (G15) – the confidence that the personnel or team(s) involved in the creation or verification of the asserted solution has the required domain experience.
- *User-Defined Personnel Factors* (G16) – user-defined factors related to personnel that are considered to have an effect on the overall confidence.

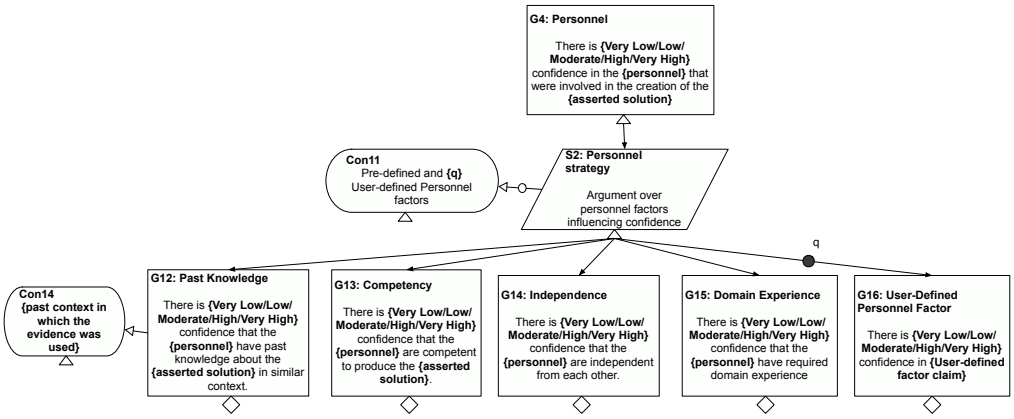


Figure 4. Personnel argument pattern for the asserted solution

- *Processes/Techniques* (G5) – Arguments regarding the different processes or techniques employed to create or verify the asserted solution. Similar to the *personnel* (G4), arguments can be made over each pre-defined or user-defined processes factors (S3). Figure 5 shows the processes/techniques argument pattern. For each process or technique, we further decompose the arguments as:
  - *Past Use* (G17) – the confidence based on the past experience of using the same process or technique to create or verify the asserted solution.
  - *Definition* (G18) – the confidence that the process or technique has been clearly defined and the definitions have been followed.

- *Peer Review* (G19) – the confidence that the various outcomes of applying the process or technique have been peer reviewed (e.g., the hazard log has been peer reviewed at least once).
- *User-Defined Processes/Techniques Factor* (G20) – user-defined factors related to the process or technique that are considered to have an effect on the overall confidence.

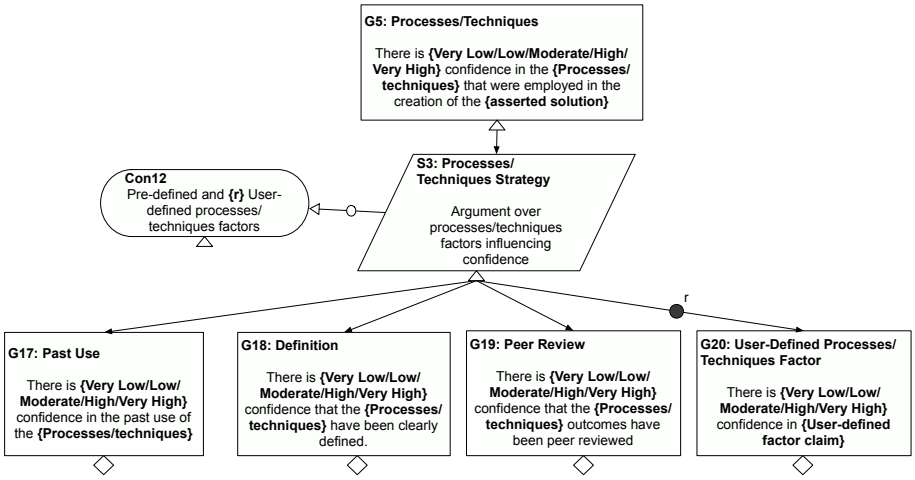


Figure 5. Processes/Techniques argument pattern for the asserted solution

- *Tool Integrity* (G6) – Arguments regarding the integrity of the different tools used to create or verify the asserted solution. Arguments can be made over pre -defined or user-defined factors (S4). Figure 6 shows the tool integrity argument pattern. The integrity of the tools used can be decomposed as:
  - *Bound Qualification* (G21) – the confidence that the specific usage of the different tools used for creation or verification of the evidence is within the constraints of its qualification, e.g., the tool used to record the hazard log was used in the context of the appropriate process or that the tool was configured appropriately.
  - Standard Qualification (G22) – the confidence that the different tools were qualified in accordance to the safety standard used, e.g., DO-178C requires tools to be qualified when used as part of the software assurance process. The safety standard complied with should be made explicit in context Con15.
  - *User-Defined Tool Factor* (G23) - user-defined factors related to the tools that demonstrates sufficient integrity in its usage.

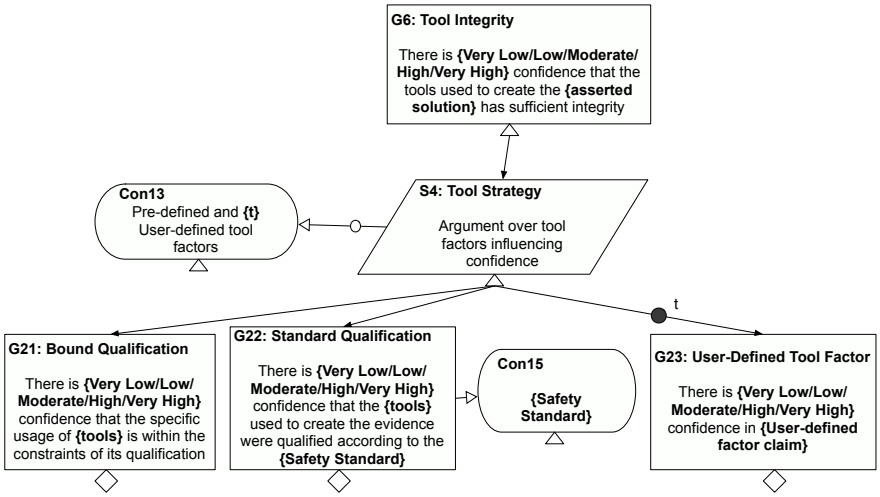


Figure 6. Tool Integrity argument pattern for the asserted solution

- *Structural Compliance* (S1) – Arguments regarding the structural integrity and structural compliance of the asserted solution. Figure 7 shows the content compliance argument pattern. The strategy is decomposed into two sub-goals:
    - *Scope* (G24) – the confidence that the asserted solution has been scoped and defined according to the required document format to demonstrate compliance with a specific safety standard, e.g. are all terms and acronyms defined in the hazard log. The safety standard conformed to must be made explicit (Con16).
    - *Expected structure* (G25) – the confidence that the asserted solution conforms to the expected structure for the particular evidence type, e.g., all identified hazards in the log must be presented in the order of their severity. The type of the evidence must be made explicit (Con 17).
    - *User-Defined Content Compliance Factor* (G26) - user-defined factors related to the compliance of content required for the asserted solution.

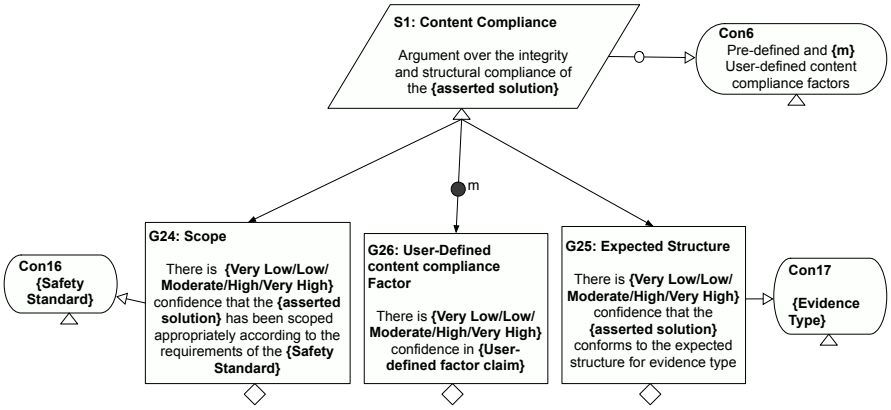


Figure 7. Content Compliance argument pattern for the asserted solution

- *Evidence of Past Use* (G7) – The confidence obtained from the past use of the asserted solution in a similar context. The similar context (e.g. the associated claim) should be made explicit here via the context Con7.
- *User-Defined Trustworthiness Factor* (G8) - user-defined factors other than the ones already defined in the pattern that relate to the trustworthiness of the asserted solution. Required claim description must be added to the pattern and further decomposition to be carried out of required.

The appropriateness of the evidence (hazard log) relates to satisfaction of the claim (all hazards were identified). The appropriateness of the evidence can be defined as the property of the evidence to sufficiently satisfy the claim it was cited for. The type of evidence that is most appropriate can only be determined based on the nature of the claim and argument that the evidence is intended to support. Hence, we decompose the *appropriateness* of the *asserted solution* (Figure 8) as:

- *Asserted role* (G9) – the confidence that the evidence *type* of the asserted solution is capable of providing the asserted role in the argument [26]. For example, the role of the hazard log is to identify hazardous functional failures that may occur in system X. The type of the evidence and the asserted role has to be made explicit using contexts Con 8 and Con 9 respectively.
- *Intent* (G10) – the confidence that the *specific* evidence satisfies the asserted role. For example, the intent of the hazard log to demonstrate that the system X does not contain errors that could manifest as hazards. The asserted role of the evidence has to be made explicit (Con10)
- *User-Defined Appropriateness Factor* (G11) – user-defined factors other than the ones already defined in the pattern that relate to the appropriateness of the asserted solution. Required claim description must be added to the pattern and further decomposition to be carried out of required.

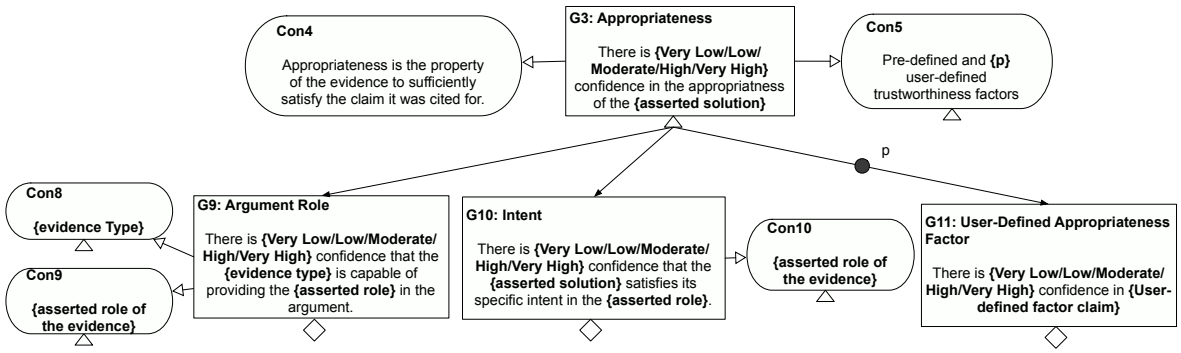


Figure 8. Appropriateness argument pattern for the asserted solution

Each of the lower level goals can be further decomposed if required. However in the current approach, we provide question prompts for each of the leaf claims. The process of collecting the individual belief functions for each of the lower-level claims and the confidence quantification is discussed in the following subsection.

B. Confidence Quantification with ER

The ER algorithm [3] provides a means to combine the individual lower-level belief functions for each factor into a coherent overall assessment. It uses depth-first traversal of a defined hierarchical structure. The following description will show how an example set of lower-level attribute assessments is combined using ER. Details regarding the algorithm and the processing can be found in [3][27].

Let us consider the example of the hazard log. In order to gain confidence in the evidence source, more specifically the *trustworthiness* of the *hazard log*, lets consider two factors related to the *Personnel/Teams* from the above argument pattern: *Independence* (I) and *Competence* (C). The *Independence* factor assesses the extent to which there is sufficient level of independence among the personnel/teams that were involved in the creation of the hazard log. The *Competence* factor assesses the skill and proficiency of the personnel/teams who created the hazard log. Each of these two factors is further broken into two leaf-node attributes representing the two questions (Q). These lowest-level attributes form the basis for obtaining the assessor's atomic assessments that are to be aggregated to a high-level assessment of hazard log. For this example, let's consider some questions identified from the Railway Safety Commission checklist [32] that are used in practice to assess hazard logs.

Independence:	$q1.$	If independence is required, is the person doing the verification different than the one responsible for developing the hazard log?
	$q2.$	Is the manager to whom the team reports identified so that it can be confirmed that the requirements on independence are met?
Competence:	$q3.$	Are there records of attendance / participation in hazard identification workshops / exercises of the personnel, that include the name, organisation and role?
	$q4.$	Is there information on the competency of the personnel?

For each question $q \in Q$, the assessment consists of two parts:

1. The assessor's agreement a on the Likert scale $1 \geq a \geq 5$ where (for these questions) 1 represents *Definitely not, 2* represents *No, 3* represents *Maybe, 4* represents *Yes* and 5 represents *Absolutely.*
2. The assessor's confidence $0 \geq c \geq 1$, where 0 represents no confidence at all (total ignorance), and 1 represents total confidence.

These answers are then used to construct a distribution $S(q) = \{(H_{n=a}, c), (H_{n\neq a}, 0), n = 1, \cdots, 5\}$. For example, the assessor might answer the questions (in the order listed above) as follows: (4, 0.8), (5, 1), (3, 0.5), (4, 0.8). In other words, they *agree* with a certainty of 80% that the person doing the verification is different from the one responsible for developing the documents, they *strongly agree* with a certainty of 100% that the manager to whom the team reports is identified, etc. The resulting distributions then look as follows: [0,0,0,0.8,0], [0,0,0,0,1], [0,0,0.5,0,0], [0,0,0,0.8,0].

These distributions then form the lowest-level attributes for the ER algorithm to aggregate. To begin with, the algorithm combines the two leaf-node attributes for *Independence* ([0,0,0,0.8,0], [0,0,0,0,1]), setting the weights for each attribute to 0.5 (1/[number of attributes]). This results in the aggregate distribution for *Independence* [0,0,0,0.364,0.545] (the *doubt* is quantified explicitly as $1 - (0.364 + 0.545) = 0.091$. Similarly, combining the answers for *Competence* ([0,0,0.5,0,0], and [0,0,0,0.8,0]) yields [0,0,0.231,0.462,0], with an explicit doubt of 0.308. Now, to compute the final assessment of the *trustworthiness* of the hazard log, ER combines the two scales computed for *Independence* and *Competence* again

([0,0,0,0.364,0.545] and [0,0,0.231,0.462,0]), to produce a final assessment of [0,0,0.099,0.452,0.281], with an explicit doubt of 0.168 (16.8%).

This final distribution reflects the two confidence values:

(1) The assessor's confidence in evidence – To represent the assessor's confidence in the trustworthiness of the hazard log, we treat the final distribution of the assessment as a five point Likert-scale: *Very Low, Low, Medium, High* and *Very High*. From the above final distribution, we obtain 9.9% of the mass is attributed to *Medium* confidence, 45.2% can be attributed to *High* confidence, and 28.1% can be attributed to *Very High* confidence. To best represent the assessor's confidence, in our approach we use the median of the distribution. In the above distribution, the median is 0.099, which represents *Medium* in the Likert-scale.

(2) The assessor's confidence in their assessment– To represent the assessor's confidence in their assessment of the *trustworthiness*, we quantify confidence from a scale of 0-100%, with intervals: *0-20% Very Low, 20-40% Low, 40-60% Medium, 60-80% High*, and *80 -100% Very High*. In the above distribution the assessor has 9.9% confidence on the assessment of the *trustworthiness* of the hazard log, which represents *Very Low* confidence in the scale.

To summarise, the final distribution indicates that (1) the assessor has *Medium* confidence in the *trustworthiness* of the hazard log, and (2) the assessor has *Very Low* confidence in the assessment of the *trustworthiness* of the hazard log.

IV. TOOL SUPPORT

In this section, we briefly describe the prototype tool named EviCA (**Evi**dence **C**onfidence **A**ssessor), developed to support the proposed framework. Specifically, EviCA allows users to: (1) create and edit safety arguments using GSN, (2) question the various reasons for having confidence in the used in primary argument, (3) automatically build confidence arguments based on a predefined GSN pattern that is customisable, and (4) calculate the confidence and the uncertainty at each level of the argument automatically.

EviCA is written in Java programming language as a plug-in to the Eclipse IDE. It uses some utilities of the underlying Eclipse framework, notably the Graphical Editing Framework (GEF). We use Microsoft Excel as one of the means to import checklists for reasoning lowest-level factors. We also use Graphviz**,** an open source graph visualization software to visualize the individual belief functions the user provides and build a model of the confidence argument summarizing the belief functions. Figure 4 shows the technology stack used for EviCA

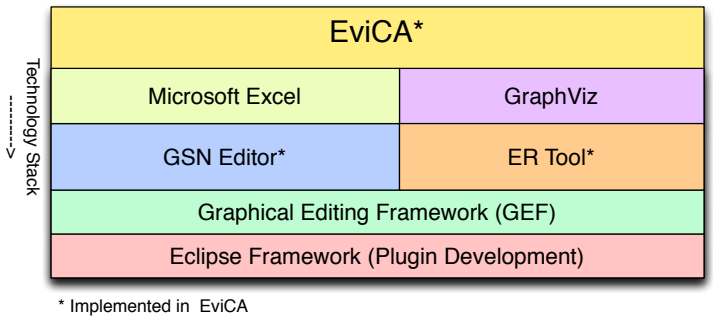


Figure 9. Technology stack of EviCA tool

A. Creating and Editing Safety Arguments

Figure 10 shows a screenshot of a sample safety argument fragment described in GSN. The pallet to the right of the screen provides users with the various GSN elements (Goals, solutions, strategies, context, etc.) that they need to create a goal structured safety case. The properties of a selected item can be accessed at the bottom of the screen. The node description can be either edited in the properties window or can be edited directly in the canvas. All edits in the elements are reflected in real-time. The nodes can be selected, resized, moved or deleted as required. The pane in the left of the window is a project explorer that displays the different projects and their associated safety case diagrams. The GSN editor developed as part of EviCA is the first of its type that allows users to create and manipulate confidence arguments. Users can click and drag *Assertion Claim Points* (ACP), between goals and solutions. An ACP is indicated by a black rectangle on the relevant link. Fig. 10 shows ACPs named ACP36 and ACP 37.

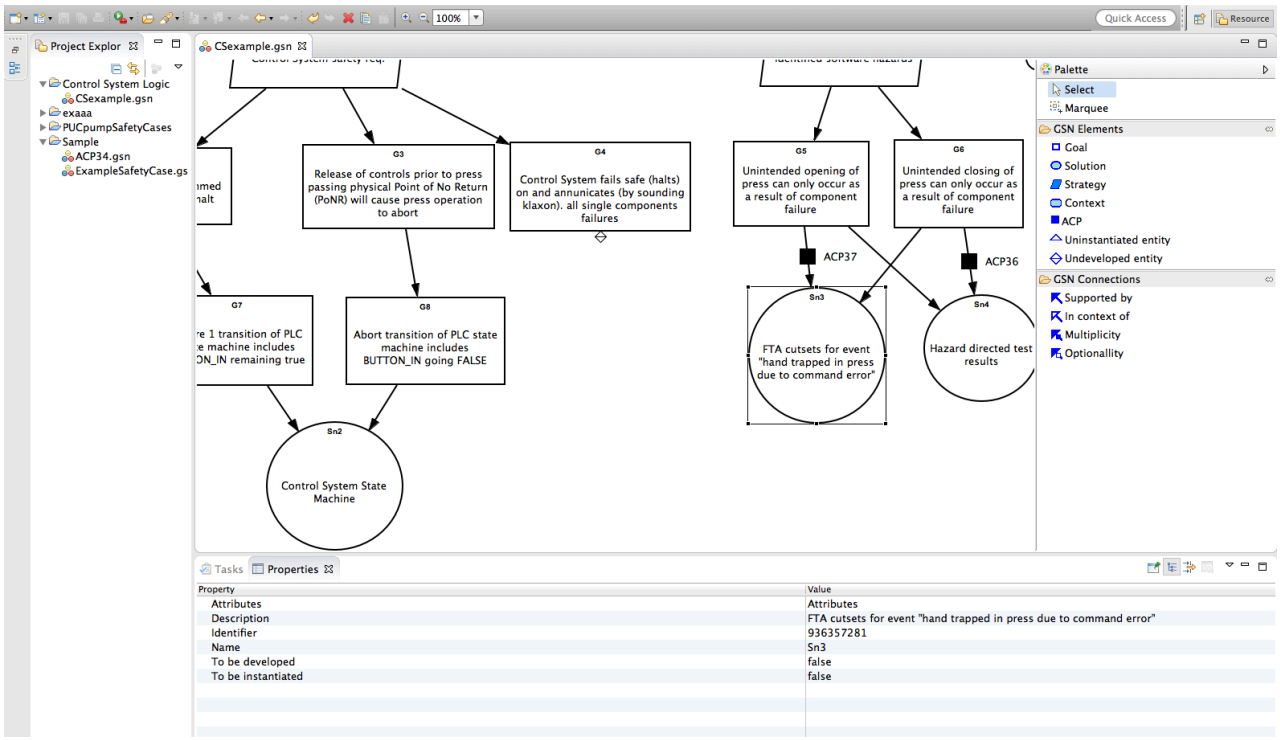


Figure 10. Screenshot of EviCA's GSN editor with sample GSN safety case fragment

B. *Confidence Argument Generator*

Double clicking on any ACP created between a goal and an asserted solution opens the *Confidence Argument Generator* wizard. This wizard is the means to create the confidence argument based on the various confidence factors. It initially has a predefined tree structure with a set of pre-defined confidence factors. These factors were identified through systematic examination of evidence assessment practices (Section IV.A). Right clicking on any parent factor allows users to edit the tree structure. Users can edit, add, delete and move factors if required. When adding a new child to a parent, the user can specify whether it is a goal or a strategy to correspond in the confidence argument pattern. The user can enter element description at this stage. All goal element descriptions need to have a mandatory placeholder **"{0}"** that will be replaced by the qualitative confidence tag (Section IV.A). Each element in the tree has a weight function. The user can define the weights for any factor in the tree. The sum of weights of any factor must not exceed 1. By default, the weights are equally split among all the children depending on the number of children. Figure 11 shows a screenshot of the edit window for factor *Past Knowledge*. Since Past Knowledge factor is a child (along with three other factors) of *Personnel/team*, the weight is equally split to all four factors (0.25) by default.

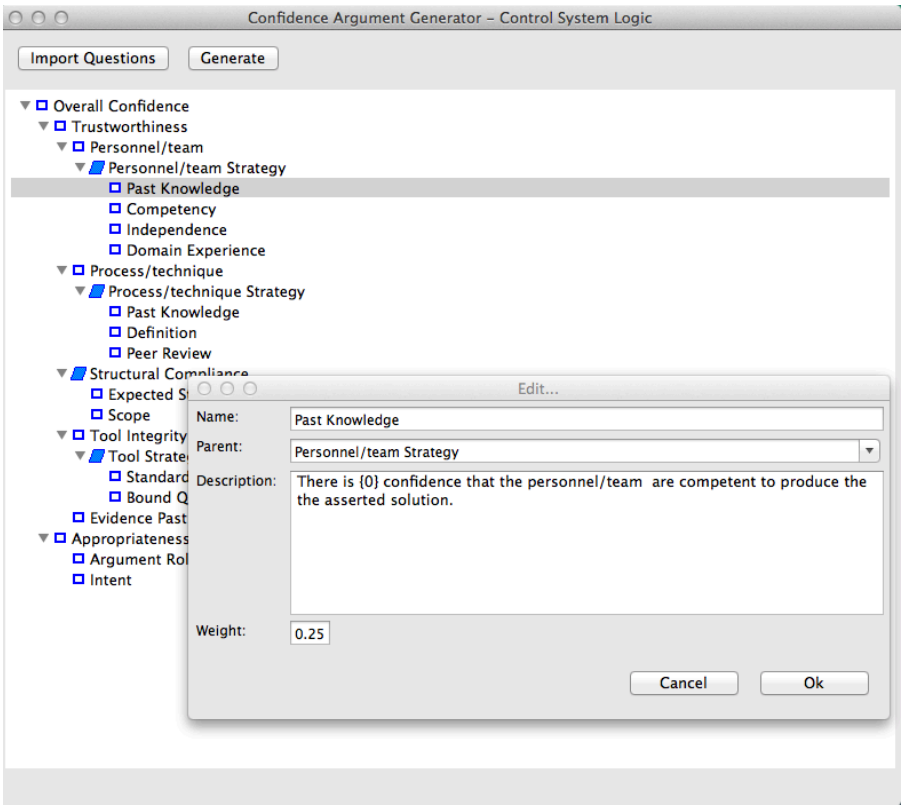


Figure 11. Screenshot of confidence argument generator window and edit window for past knowledge factor

C. *Collecting Belief Functions*

The tool allows users to add individual questions to each of confidence factor in order to factor them into a safety case. There are two ways to add questions. Users can manually add one question at a time by right clicking the lowest child. This will bring up the *Add Question* window (Figure 12). The user can define the weight of each question the same way as the parent. The second way to add questions is with the help of *Import Questions* button. EviCA allows users to import a set of Microsoft Excel-based checklist questions. In an Excel-based checklist, the rows represent the different checklist items and the columns represent the confidence factors. A checklist item can be mapped to a confidence factor by marking an X in Excel. EviCA then automatically imports the marked questions into their appropriate factors in the *Confidence Argument Generator* dialog. This way EviCA allows users to import large sets of questions at once with ease.

To obtain the belief function for each question, the user has to input two values: (1) an agreement grade and (2) a confidence value. By default, we use a five point Likert scale for the grade values ranging from *Definitely not (0)* to *Absolutely (5)* for all questions. However, the user can change the default scale by selecting from a list of pre-defined scales using the *Import Grade* button. The confidence on the answer can be entered using the slider that ranges from 0% to 100% confidence. The user can also provide additional evidence information (if required) as part of their answer for a particular question. The *Attach Files* button allows users to attach any additional evidence that might favour their rationale for choosing a particular answer. For example, when answering a question regarding the competency of the developer, additional evidence such as the CV of the developer can be provided. In addition, the user can justify their rationale behind a particular answer by providing further comments. Unanswered questions in the *Confidence Argument Generator* window are denoted by a ✖ on the left side, while answered questions are denoted by a ✓ (Figure 12).

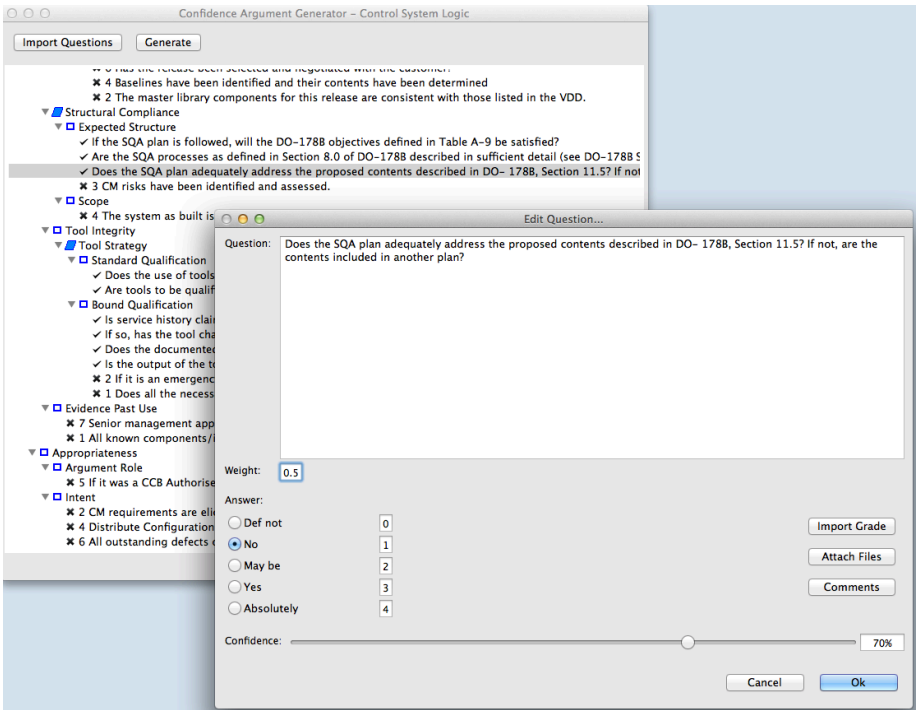


Figure 12. Screenshot of confidence argument generator window and add question dialog

Once all the questions are answered, the user can click the *Generate* button to automatically build the confidence argument for the asserted solution with the confidence in the evidence and the assessment visually presented. Figure 13 shows the fragment of the confidence argument structure built by EviCA. The structure is editable by the user.

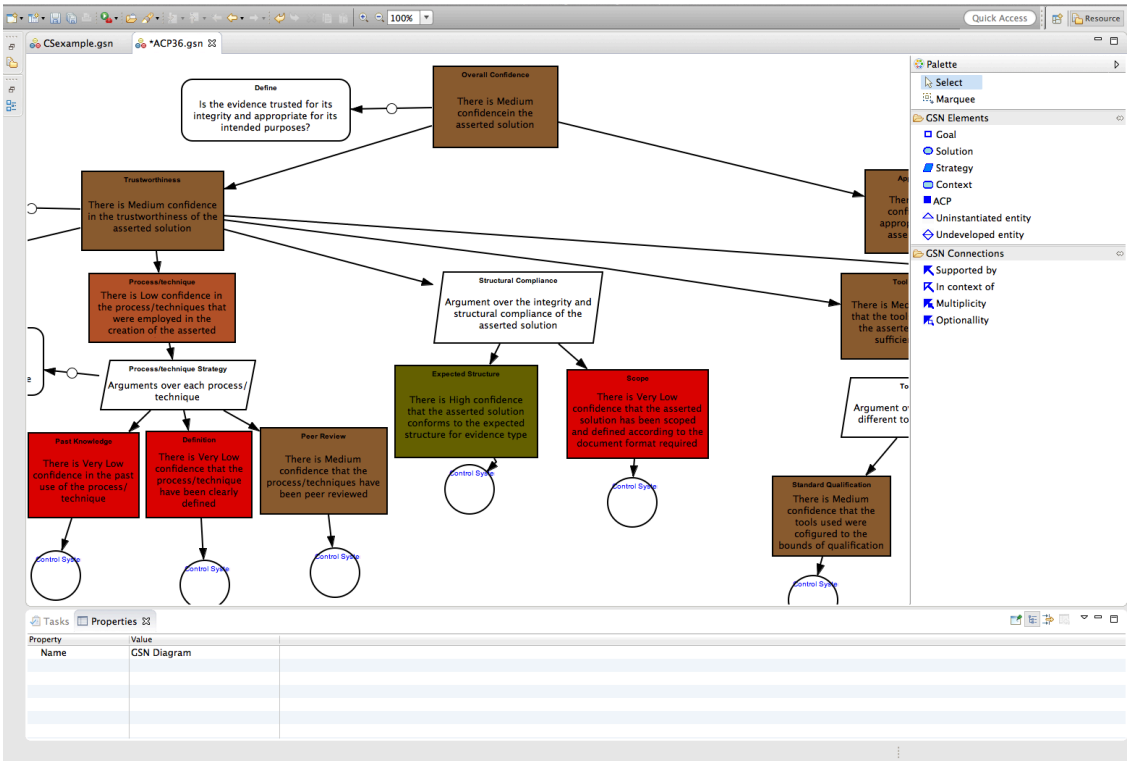
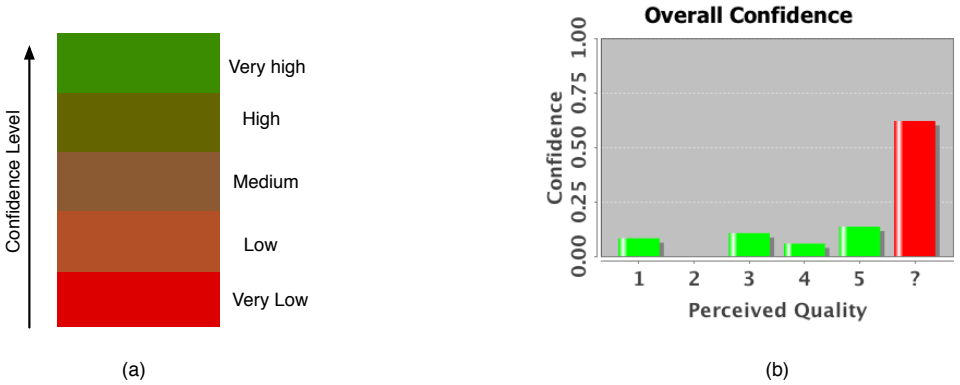


Figure 13. Screenshot of final confidence argument fragment generated in EviCA

To present the assessor's confidence on the evidence, EviCA uses the placeholders **"{0}"** in the goal description to be automatically replaced with the calculated confidence value. Based on the lowest child´s belief masses, EviCA automatically propagates the confidence to the top-level goals using the ER algorithm. The weights of each branch are used in calculating the overall confidence. The answers to all the questions and the corresponding comments are summed and attached as solutions to corresponding parent goal. Any additional evidence provided in the question is also added as a solution to the corresponding parent goal.

To present the assessor's confidence in the assessment, EviCA uses a specific colour scheme for representing different levels of evidence. Figure 14.a shows the colour schema used by EviCA. Additionally, EviCA also provides individual bar charts of *belief function*s for each goal and solution node in the structure. Figure 14.b shows the distribution of confidence values as individual bar chart for the overall confidence in the evidence. As seen in the figure, the values 1-5 correspond to the assessor's confidence in the evidence (1 being *Very Low* and 5 being *Very High*). The height of a bar represents the assessor's confidence in their assessment. An additional bar (in red) denotes the uncertainty or doubt inherent in the *overall confidence*. This is computed from those belief functions where the total sum of beliefs did not add up to 1.



(a) (b)

Figure 14. a. Colour code used in EviCA to represent confidence in the assessment; 14.b. A bar chart presenting the individual belief function for overall confidence with explicit uncertainty (in red)

V. EVALUATION

User acceptance is a critical success factor for any technology adaptation and has been a major area of research especially within Information systems (IS) [28]. In order to evaluate the acceptance of our proposed approach and tool support, we adopted the Technology Acceptance Model (TAM) [5]. TAM aims to "Provide an explanation of the determinants of computer acceptance that is general, capable of explaining user behaviour across a broad range of end-user computing technologies and user populations, while at the same time being both parsimonious and theoretically justified" [5]. TAM focuses on three main facets of user acceptance:

- *Perceived Ease of Use:* "the degree to which a person believes that using a particular method would be free of effort"
- *Perceived Usefulness:* "a person's subjective probability that using a particular system would enhance his or her job performance"
- *Intention to Use:* "the extent to which a person intends to use a particular system"

In this paper, we evaluated the proposed approach and the tool support based on the above three elements of TAM. The study targeted safety experts, mainly practitioners, who are directly involved in safety case development and safety evidence assessment for critical systems. We developed a short presentation of the proposed approach and a tool demo video[‡] showing the main features of EviCA. This allows us to mitigate some threats related to internal validity, e.g., threats related to possible bias in the way the technology were presented to each participant. At the end of the presentation, the participants were asked to fill a short questionnaire with 12 questions. The questionnaire was divided into four parts. The first section contained a short description of the aim of the survey and five background questions. The second section consisted of three questions regarding the participants perceived usefulness of the approach. The third section consisted of three questions regarding the participants perceived ease of use of the approach and the tool support. And finally, the last section consisted of one question regarding the intention of use of the approach and any further comments regarding the proposed approach. We used a five-point Likert Scale ranging from *Strongly Agree* to *Strongly Disagree* to collect answers for all questions. Additionally, we also allowed participants to mention any comments to each question with the help of '*Others'* option. Other than section 1,

[‡]https://www.youtube.com/watch?v=f5d9kIuteqM

the remaining sections of the questionnaire were randomised. The entire questionnaire is shown in Appendix A. The presentation and the survey questionnaire were sent via personal email invitations and subsequent remainders to some practitioners and safety experts we knew. We also asked them to let other colleagues know about the survey. Additionally, the presentation and the survey were also posted on a social networking websites for people in professional occupations (http://www.linkedin.com).

A total of 9 participants provided their feedback on the approach. All had more than 2 years of experience in safety certification/assurance/assessment related activities. All also indicated that, as part of their work they assess safety evidence information and/or develop/assess safety case documents.

Relating to the perceived usefulness of the approach and the tool support, 67% (6 out of 9) of the participants indicated that they agree that the ability to express ignorance or doubt in the approach is useful for the assessment of safety evidence. A similar number of responses (67%) indicated that the range of safety evidence assessment factors covered in the approach is adequate. For the final question in this section, 67% mentioned that they agree that the use of the approach will lead to more accurate safety evidence assessments. The % of responses for each of the three questions in the perceived usefulness of the approach is shown in Figure 15, 16 and 17 respectively.

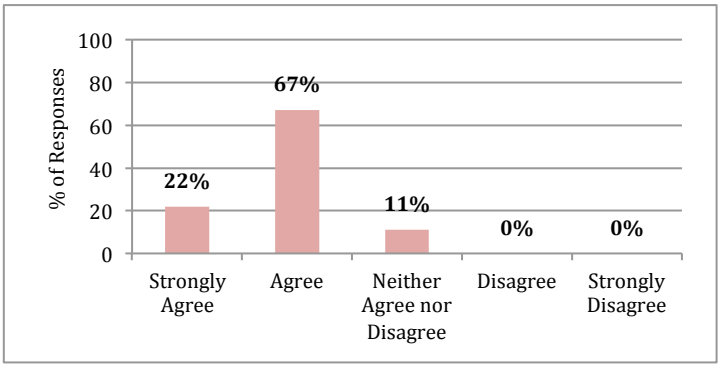


Figure 15. The ability to express ignorance or doubt in the approach is useful for the assessment of safety evidence.

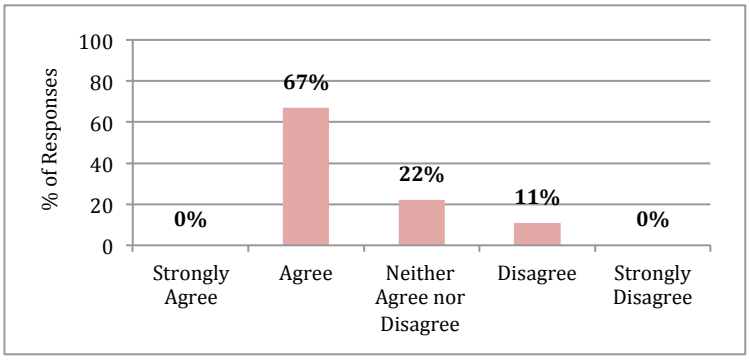


Figure 16. The range of safety evidence assessment factors covered in the approach is adequate.

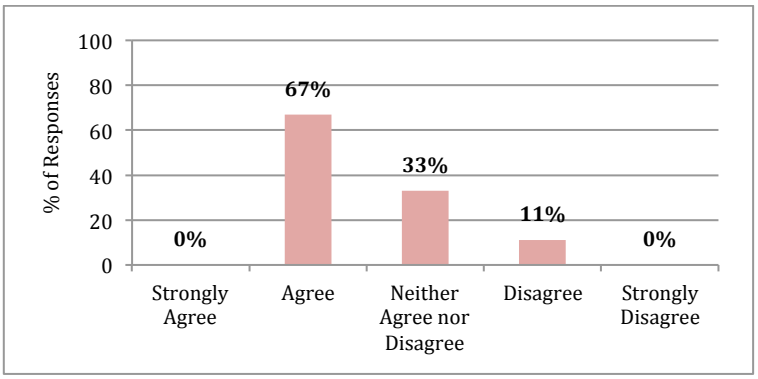


Figure 17. Use of the approach will lead to more accurate safety evidence assessments.

Regarding the perceived ease of use of the approach and the tool support, 44% (4 out of 9) of the participants indicated that it is easy to express any doubt or ignorance about a particular safety evidence assessment with the approach. When asked if it is easy to customise a safety evidence assessment in the tool, 33% (3 out of 9) of the participants indicated that they strongly agree. The final question in this category asked the participants if it is easy to interpret the results produced by the approach. 56% (5 out of 9) of the participants indicated that they strongly agree. Figure 18, 19 and 20 shows the % of responses for each of the three questions in this section respectively.

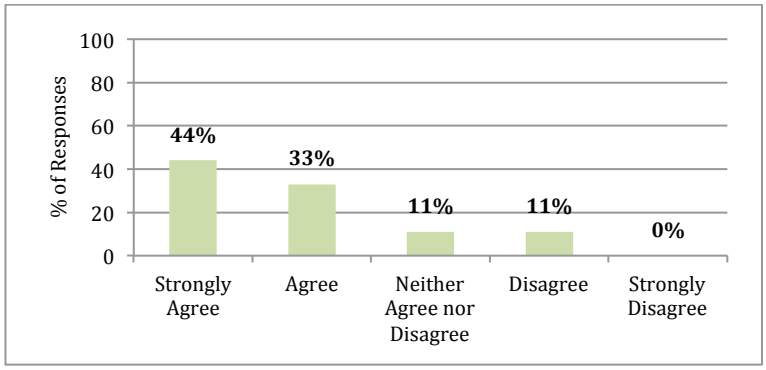


Figure 18. It is easy to express any doubt or ignorance about a particular safety evidence assessment with the approach.

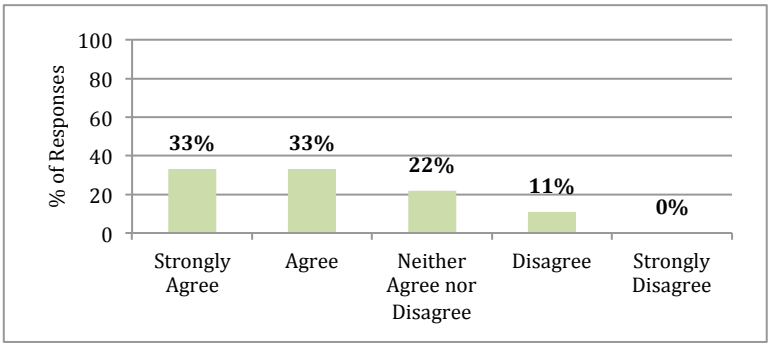


Figure 19. It is easy to customize a safety evidence assessment in the tool.

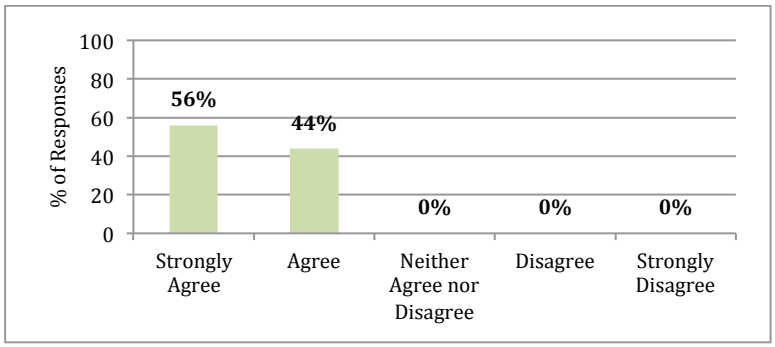


Figure 20. It is easy to interpret the results produced by the approach.

Finally, one question was asked to the participants about the intention of use of the approach and tool support. 56% (5 out of 9) of the participants indicated that they agree that they would use the approach for safety evidence assessment tasks if it were made available to them within their organization. Figure 21 shows the % of responses for this question.

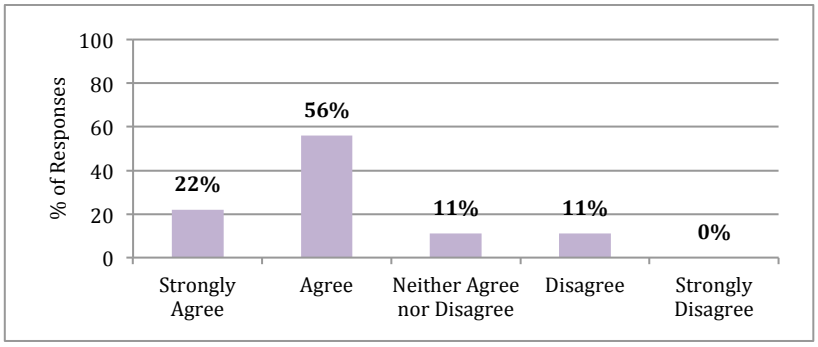


Figure 21. I would use the approach for my safety evidence assessment tasks (if it were made available to me within my organization).

In summary, the answers to the perceived usefulness, perceived ease of use and intention of use of the approach and tool support suggests that the approach was generally viewed as being easy to understand and use. The participants also indicated it would be advantageous to use the approach within their organizational context.

VI. Related Work

A recent survey on the state-of-the-practice of evidence assessment shows that expert judgement is one the most commonly used technique for evidence assessment [31]. In spite of its high reliance, little has been studied about expert judgement in safety assessment context in general and more specifically in assessing safety evidence [34]. [9].Oour previous work attempted to understand how safety experts involved in the assessment process interpreted and understood three evidence assessment criteria - *completeness*, *sufficiency* and *overall confidence* [4]. Through this study, we identified several factors related to the evidence creation and verification process that influenced expert decisions concerning the acceptance of safety evidence. We also identified through the study that most of these factors were implicit in the assessment process.

Safety case development and assessment has been of much interest in research over the past years. One method to construct safety argument using GSN is the *Six Step* method [6]. However, this method does not explicitly consider the confidence of the safety argument and the confidence in the safety evidence [7]. Other strands of work have provided the various criteria and factors that should be considered to determine the confidence in safety evidence and arguments [8]

A new approach for creating clear safety cases was introduced in [2]. The approach suggests building a secondary confidence argument that explicitly states the reason for having confidence in the evidence cited. The paper argues that a confidence argument helps in explicitly reasoning about the confidence established in a primary safety argument. The paper acknowledges that there will be uncertainties associated with aspects of the safety argument or supporting evidence and the role of the confidence argument is to explicitly address those uncertainties and explain why there is sufficient confidence in the safety argument. The paper presents an indicative argument pattern for the confidence argument that is based upon the identification and management of assurance deficits (uncertainties). Combining this approach of a secondary confidence argument with the various factors that influence the assessor's confidence, we propose a more exhaustive pattern that covers the various reasons for having confidence in the evidence.

Other attempts have been made to measure confidence in safety cases qualitatively [10][11]. One approach similar to the one proposed in this paper can be found in [12]. It uses common concerns associated to an argument and systematically builds

confidence arguments based on them. The limitation of this approach, as acknowledged by the authors, is it only covers *Trustworthiness* factors. Other concerns related to, for example *Appropriateness,* are yet to be categorised. Past studies have detailed the notion of uncertainty in safety cases [13][14] and provided ways to handle them e.g., using Bayesian Belief Networks (BBN) [15][16]. Although plausible, BBN rely heavily on their probability tables, which in turn rely on the availability of prior probability information. This reliance upon the prior probability information, which is often complicated to obtain given the scarcity of priors, makes it difficult to provide a thorough assessment on confidence when the assessor is ignorant or doubtful.

ER is an example of Dempster-Schäfer (DS) theory [17][18]. DS has been applied to a multitude of diverse `Multi-Criteria Decision Analysis problems' such as environmental impact assessments [19], assessment of weapons systems capabilities [20], and safety analysis [21]. An approach based on trust cases to support and improve expert assessment of arguments is proposed in [22]. The approach is developed in connection with the Trust-IT methodology [29][30]. Similar to our proposed approach, the trust case approach is based on DS theory and it provides a way to issue assessments and their aggregation depending on the types of inference used in arguments. What differentiates our work from trust case based approach is the use of an explicit secondary argument structure for demonstrating the reason for having confidence in the evidence. DS theory and aggregated assessment alone may not be sufficient to demonstrate sufficient confidence in the assessment of the evidence or the argument. The approach proposed in this paper proposes a systematic confidence argument pattern that allows explicit presentation of the various factors that provide confidence in the evidence and aggregates the individual beliefs into a final assessment. Through our approach, the final assessment can be broken down into lowest level reasons for having confidence, making any uncertainty explicit.

VII. CONCLUSIONS

This paper has proposed a novel approach to automatically construct confidence arguments and quantify confidence using Evidential Reasoning. The approach enables safety experts to assess evidence by explicitly reasoning the various factors related to the evidence creation and verification process. The approach also allows experts to indicate their uncertainty or doubt related to the assessment. As part of our approach, we proposed a confidence argument pattern that details the various reasons for establishing confidence in the evidence. The confidence argument pattern decomposes the abstract notion of overall confidence in the evidence into lower level sub-claims regarding the trustworthiness and the appropriateness of the evidence. The proposed approach then enables experts to provide individual belief functions for the lowest-level claims by questioning each of the factors. With the help of the ER algorithm, we then propagate (aggregate) the belief functions to higher level claims, until it provides an aggregate belief function for an overall confidence claim with which explicitly captures any uncertainty in the expert´s judgement from the lower-level confidence ratings. The final result of the approach is an explicit confidence argument structure that visually and quantitatively presents the confidence in the evidence assessed.

As a proof of concept, the proposed approach has been developed into a prototype tool named EviCA (Evidence Confidence Assessor). The proposed approach and the tool support were evaluated using Technology Acceptance Model (TAM) by conducting a survey with safety experts who are directly involved in safety case development and evidence assessment. A total of nine experts responded to questions relating to the perceived use of the approach, perceived usefulness of the approach and future intention to use the approach. The overall results of the evaluation suggest that the participants perceived the approach to be useful and easy to use. The results also indicate that participants would use the approach and tool support if it were available in their organizational context.

As future work, we would like to further extend the proposed confidence argument pattern by evaluating it with more checklists used in practice and with further safety experts. We would like to evaluate the approach and the tool support with a case study. We would also like to develop a classification of questions for each confidence factor in the pattern for different evidence type. Improving and updating the tool support would also be part of the on going and future work.

ACKNOWLEDGMENT

We acknowledge funding from the FP7 programme under grant agreement n° 289011 (OPENCOSS) and from the Research Council of Norway under Project No. 203461/030. The authors would like to thank the industrial collaborators in OPENCOSS for their useful feedback.

REFERENCES

[1] Interim Defence Standard 00-56 Part 1 - Issue 5, in, UK MOD (2014)

[2] Hawkins. R, et.al.: A new approach to creating clear safety arguments. In Advances in Systems Safety (pp. 3-23) (2011)

[3] Yang J.-B., Xu D.-L.: On the evidential reasoning algorithm for multiple attribute decision analysis under uncertainty, IEEE Transactions on Systems, Man, and Cybernetics, Part A, vol. 32,no. 3 (2002)

[4] Nair. S, et.al.: Understanding the practice of Safety Evidence Assessment: A Qualitative Semi-Structured Interview Study, Technical report, Simula Research Laboratory (2014)

[5] Davis, Fred D.: A technology acceptance model for empirically testing new end-user information systems: Theory and results. Diss. Massachusetts Institute of Technology (1985).

[6] Kelly, T.: A six-step Method for Developing Arguments in the Goal Structuring Notation (GSN). Technical report. York Software Engineering, UK (1998)

[7] Hawkins, R., Kelly, T.: Software Safety Assurance – What Is Sufficient?. In: 4th IET International Conference of System Safety (2009

[8] Menon, C., Hawkins, R., McDermid, J.: Defence standard 00-56 issue 4: Towards evidence-based safety standards. In: Safety-Critical Systems: Problems, Process and Practice, pp. 223–243. Springer, London (2009)

[9] Weaver, R.: The Safety of Software - Constructing and Assuring Arguments. PhD thesis, Department of Computer Science, University of York (2003)

[10] Bloomfield, R., Littlewood, B., Wright, D.: Confidence: Its Role in Dependability Cases for Risk Assessment. In: 37th Annual IEEE/IFIP International Conference on Dependable Systems and Networks, DSN 2007, pp. 338–346 (2007)

[11] Denney, E., Pai, G., Habli, I.: Towards Measurement of Confidence in Safety Cases. In: International Symposium on Empirical Software Engineering and Measurement (ESEM 2011). IEEE Computer Society, Washington, DC (2011)

[12] Ayoub, Anaheed, et al. "A systematic approach to justifying sufficient confidence in software safety arguments." *Computer Safety, Reliability, and Security*. Springer Berlin Heidelberg, 2012. 305-316.

[13] Denney. E, Pai. G.: A lightweight methodology for safety case assembly. In Computer Safety, Reliability, and Security (pp. 1-12). Springer Berlin Heidelberg (2012).

[14] Weaver, R., et.al.: Gaining confidence in goal-based safety cases. In Developments in Risk-based Approaches to Safety (pp. 277-290) (2006)

[15] Denney. E, et.al.: Towards measurement of confidence in safety cases. In ESEM (2011)

[16] Wu, Weihang, and Tim Kelly. "Combining bayesian belief networks and the goal structuring notation to support architectural reasoning about safety." *Computer Safety, Reliability, and Security*. Springer Berlin Heidelberg, 2007. 172-186.

[17] Dempster A. P.: A generalization of Bayesian inference, Journal of the Royal Statistical Society, Series B, vol. 30, pp. 205–247 (1968)

[18] Shafer. G.: A Mathematical Theory of Evidence. Princeton University Press, (1976)

[19] Wang, Y. M, et al. "Environmental impact assessment using the evidential reasoning approach." *European Journal of Operational Research* 174.3 (2006): 1885-1913.

[20] Jiang, J. et al. "Weapon system capability assessment under uncertainty based on the evidential reasoning approach." *Expert Systems with Applications*38.11 (2011): 13773-13784.

[21] Wang, J., J. B. Yang, and P. Sen. "Safety analysis and synthesis using fuzzy sets and evidential reasoning." *Reliability Engineering & System Safety* 47.2 (1995): 103-118.

[22] Cyra, Lukasz, and Janusz Górski. "Expert assessment of arguments: A method and its experimental evaluation." *Computer Safety, Reliability, and Security*. Springer Berlin Heidelberg, 2008. 291-304.

[23] Kelly, Tim P., and John A. McDermid. "Safety case construction and reuse using patterns." *Safe Comp 97*. Springer London, 1997. 55-69.

[24] Kelly, T.: Arguing safety – a systematic approach to managing safety cases. PhD thesis. Department of Computer Science, University of York (1998)

[25] GSN Committee. "Draft GSN Standard. Version 1.0." (2010).

[26] R. Hawkins, J. McDermid, Software Systems Engineering Initiative, SSEI-TR-0000041, *Software Safety Evidence Selection and Assurance*, Issue 1, University of York, October 2009.

[27] Walkinshaw, N. "Using evidential reasoning to make qualified predictions of software quality." *Proceedings of the 9th International Conference on Predictive Models in Software Engineering*. ACM, 2013.

[28] Turner, Mark, et al. "Does the technology acceptance model predict actual use? A systematic literature review." *Information and Software Technology* 52.5 (2010): 463-479.

[29] Gorski, J., Jarzebowicz, A., Leszczyna, R., Miler, J., Olszewski, M.: Trust Case: Justifying Trust in IT Solution, Elsevier, Reliability Engineering and System Safety, Volume 89, 33- 47 (2005)

[30] Gorski, J.: Trust Case – a Case for Trustworthiness of IT Infrastructures, Cyberspace Security and Defence: Research Issues, NATO ARW Series, Springer-Verlag, 125-142 (2005).

[31] Nair. S, et.al.: Evidence Management for Compliance of Critical Systems with Safety Standards: A Survet on the State of Practice, Technical report, Simula Research Laboratory (2014)

[32] RSC Guidelines, Railways Safety Commission, http://www.rsc.ie/publications/rscguidelines.html (accessed September 2014).

[33] Habli, I, Kelly, T, Achiveing intergrated process and product safety arguments, Proceddings of 15th Safety Critical Systems Symposium (2007).

[34] Nair, S. et al An Extended Systematic Literature Review on Provision of Evidence for Safety Certification, Information and Software Technology, Volume 56, Issue 7, July 2014, Pages 689–717

APPENDIX A – EVICA TECHNOLOGY ACCEPTANCE SURVEY QUESTIONNAIRE

This survey is intended to evaluate the user acceptance of a novel approach for evidence assessment and a prototype tool support named EviCA (Evidence Confidence Assessor) developed for the same. The aim of this questionnaire is to obtain feedbacks regarding the acceptance of the proposed technology and the tool support by gathering feedback on three main aspects:

- Perceived Ease of Use: “the degree to which a person believes that using a particular method would be free of effort”
- Perceived Usefulness: “a person’s subjective probability that using a particular system would enhance his or her job performance”
- Intention to Use: “the extent to which a person intends to use a particular system”

The questionnaire is directed to safety experts and practitioners who are directly involved in safety case development and assessment and safety evidence assessment. By responding to this survey, you will provide us with very valuable inputs regarding the acceptance of this approach in practice in future.

Filling the questionnaire should take about 5 - 8 minutes.

Although the responses will be held confidential and anonymous, we would need an identification information, which is the name of the organization for which the respondents work in order to be bale to analyze and interpret the responses correctly. Nonetheless, the name of the organization will never be published or disclosed.

Respondents that are interested in the results of the study can obtain them by contacting Sunil Nair (sunil@simula.no).

The research leading to this paper has received funding from the FP7 programme under the grant agreement no 289011 (OPENCOSS) and from the Research Council of Norway under the project Certus SFI.

Thank you very much for your participation in the study.

Section 1 – Background Information

Name (Optional) –
Email (for follow up) –

1. Are you interested in receiving the results from the survey?
□ Yes
□ No

2. Is your organization involved in safety certification/assurance/assessment?
□ Yes
□ No

3. How much experience do you have with certification/assurance/assessment - related activities?
□ Less than 6 months
□ More than 6 months but less than 12 months
□ More than 1 year but less than 2 years
□ More than 2 years

4. As part of your work, do you assess safety evidence information?
□ Yes
□ No

5. As part of your work, do you develop or assess safety case documents?
□ Yes
□ No

Section 2 – Perceived Usefulness (Questions were randomized)

6. The ability to express ignorance or doubt in the approach is useful for the assessment of safety evidence.
□ Strongly Disagree
□ Disagree
□ Neither Agree nor Disagree
□ Agree
□ Disagree
□ Other

7., The range of safety evidence assessment factors covered in the approach is adequate.
□ Strongly Disagree
□ Disagree
□ Neither Agree nor Disagree
□ Agree
□ Disagree
□ Other

8. Use of the approach will lead to more accurate safety evidence assessments.
□ Strongly Disagree
□ Disagree
□ Neither Agree nor Disagree
□ Agree
□ Disagree
□ Other

Section 3 – Perceived Ease of Use (Questions were randomized)

9. It is easy to express any doubt or ignorance about a particular safety evidence assessment with the approach.
□ Strongly Disagree
□ Disagree
□ Neither Agree nor Disagree
□ Agree
□ Disagree
□ Other

10. It is easy to customize a safety evidence assessment in the tool.
□ Strongly Disagree
□ Disagree
□ Neither Agree nor Disagree
□ Agree
□ Disagree
□ Other

11. It is easy to interpret the results produced by the approach.
□ Strongly Disagree
□ Disagree
□ Neither Agree nor Disagree
□ Agree
□ Disagree
□ Other

Section 4 – Intention of Use

12. I would use the approach for my safety evidence assessment tasks (if it were made available to me within my organization).
□ Strongly Disagree
□ Disagree
□ Neither Agree nor Disagree
□ Agree
□ Disagree
□ Other

Any additional Comments: