Proceedings of the 35th European Safety and Reliability & the 33rd Society for Risk Analysis Europe Conference
Edited by Eirik Bjorheim Abrahamsen, Terje Aven, Frederic Bouder, Roger Flage, Marja Ylönen

doi: 10.3850/978-981-94-3281-3_ESREL-SRA-E2025-P0625-cd

Propagating knowledge strength through assurance arguments using three-valued logic to assess confidence in claims

Andreas Hafver, Dag McGeorge, Erik Stensrud, Frank Børre Pedersen

Digital Assurance, Group Research Development, DNV, Norway. E-mail: andreas.hafver@dnv.com

Torbjørn Skramstad

Department of Computer Science, NTNU, Trondheim, Norway.

Roger Flage

Department of Safety, Economics and Planning, University of Stavanger, Norway.

Assurance refers to the substantiation and scrutiny of claims about a system's capabilities and the risks associated with it. The assurance process involves formulating claims that capture stakeholders' interests in the system and building structured arguments to validate and verify the claims. The end goal is to determine if there are sufficient grounds for confidence in the claims, which requires a measure of confidence and a method for propagating it through the assurance argument. One common approach for propagating uncertainty through arguments is by use of probability theory and Bayesian Networks. However, the probability numbers used in such models do not capture uncertainties in the knowledge used to assign them. Many authors have therefore suggested alternative approaches based on extensions of probability theory, including Dempster-Shafer theory and subjective logic. However, such quantitative methods have been criticized based on ambiguity in interpretation and examples of seemingly inconsistent results. Another framework called Assurance 2.0 moves away from the focus on quantifying confidence and rather aims towards "indefeasible justification", meaning qualitative confidence that there are no overlooked or unresolved doubts that could change conclusions. In this paper, we propose to use the concept of knowledge strength as a practical way to assess confidence in claims. Specifically, a claim is considered true or false only if there is strong knowledge to substantiate it; otherwise, it is treated as uncertain. We then propagate confidence through the assurance arguments using three-valued logic. Inspired by Assurance 2.0, we emphasize the need for addressing doubts that could topple an argument and the need for incorporating counter evidence in the form of defeaters. Our proposed approach is demonstrated on an example of a machine-learning-based crack detection tool.

Keywords: Assurance, strength of knowledge, three-valued logic.

1. Introduction

Assurance cases are used in many industries to document critical properties or qualities of products, processes, or systems. ISO/IEC/IEEE 15026 (2019) defines 'assurance' as "grounds for justified confidence that a claim has been or will be achieved", and an 'assurance case' as a "reasoned, auditable artefact created that supports the contention that its top-level claim (or set of claims) is satisfied, including systematic argumentation and its underlying evidence and explicit assumptions that support the claim(s)". In simple English, an assurance case 'makes a case' for some claim(s) about a target product, process or system. If the case is about safety, it is sometimes called a 'safety case'. The case may be directed towards regulators, customers, or the general public.

A recurring discussion in the assurance case community is how to establish, express and communicate the level of confidence in claims, and how to propagate confidence through an assurance argument. A common approach is to use probabilities to express degree of confidence in claims and probability theory to reason probabilistically about claims, for example Guo (2003); Denney et al. (2011); Hobbs and Lloyd (2011); Zhao et al. (2012). Other quantitative approaches use extensions of probability theory such as Dempster–Shafer theory (for example Cyra and Górski (2008a,b); Guiochet et al. (2015)), subjective logic (for example Duan et al. (2015); Yuan et al.

(2017)) and evidential reasoning (for example Nair et al. (2014, 2015)). However, these approaches based on probability theory and extensions thereof have been criticized by Graydon and Holloway (2017) based on lack of clarity in interpretation and examples of seemingly inconsistent results. Others have tried to address this criticism to develop better approaches for confidence quantification (for example Wang et al. (2019); Nešić et al. (2021); Bloomfield and Rushby (2022)).

In this paper we do not take a stance regarding the merits and weaknesses of the different quantitative approaches cited above. However, we note that any quantitative method relies on assumption and choices made when assigning numbers to represent beliefs. Since a number does not reveal how it was assigned, one inevitably needs to do a qualitative assessment of the ’strength of knowledge’ behind the numbers (Aven, 2013). This knowledge strength is not carried along when numbers are propagated through quantitative models, and it is accordingly unclear how much one can trust the numbers computed by quantitative approaches. Moreover, there are many ways of interpreting probabilities, which calls for careful definitions and communication to avoid misguided conclusions.

Bloomfield and Rushby (2020) have proposed an ‘Assurance 2.0’ framework. In Assurance 2.0, the main measure of confidence is the logical soundness of the argument (Bloomfield and Rushby, 2022). By ‘logical soundness’ they mean that the premises of the argument are true and that the argument ‘makes logical sense’ given that the premises are true. They do not consider ‘true’ to be certain in a strict mathematical sense, but rather as ’reasonable or plausible’ from a perspective of Informal Logic (IL) and Natural Language Deductivism (NLD) (Bloomfield and Rushby, 2022). In establishing logical soundness, emphasis is put on the identification and evaluation of ‘defeaters’, which are doubts or objections that could topple the argument. Assurance 2.0 aims towards ‘indefeasible justification’, meaning qualitative confidence that there are no overlooked or unresolved doubts that could change conclusions. Although they also provide a probabilistic method for quantifying and propagating confidence through arguments, Bloomfield and Rushby recommend that probabilistic statements are rather incorporated into the claims themselves (e.g., using claims such as *‘The failure probability is less than* 10^{-3} *per year’*, where probabilistic analysis is used as evidence towards this claim). They also note that probabilistic representation of assurance arguments may be useful for sensitivity analysis to get a feeling for the influence of uncertainties.

1.1. *Contributions and outline of this paper*

In this paper, we propose a new method for assessing and expressing confidence in assurance cases that is inspired by Assurance 2.0 and makes use of the ’strength of knowledge’ concept as introduced and discussed in (Flage and Aven, 2009; Aven, 2018; Berner and Flage, 2016; Amundrud et al., 2017). Our method uses strength of knowledge as a measure of confidence, and three-valued (triadic/ternary) logic (Lane, 2001) for propagating this confidence through arguments. Three-valued logic extends two-valued logic (i.e., ‘true’ and ‘false’) with an additional value ’uncertain’ (or ’undecided’, ’unsupported’, etc.). We propose to assign ‘uncertain’ to claims whenever the knowledge is not strong enough to conclude that a claim is ‘true’ or ‘false’.

We do recognize that Assurance 2.0 already uses a form of three-valued logic. Specifically, Bloomfield and Rushby (2022) discuss the need for a ‘true’, ‘false’ and ‘unsupported’ value to handle defeaters, and note that ‘false’ cannot generally be propagated upwards in arguments. They also recognize that uncertainty in side-arguments and sub-arguments (see Section 2 for an explanation of the difference) should affect a claim differently, but they nevertheless treat them the same way based on the fact that, in two-valued logic, $side\ argument \Rightarrow (sub\ argument \Rightarrow claim)$ is equivalent to $(side\ argument \wedge sub\ argument) \Rightarrow claim$ (Bloomfield and Rushby, 2022). However, by moving to three-valued logic we can easily incorporate the desired effect that a claim should be ‘uncertain’ whenever the side argument is ‘uncertain’ or ‘false’, and equal to the conclusion of the sub-argument whenever the side argument is 'true' (i.e., the operation should be asymmetric for side arguments and sub arguments).

The remainder of this paper is organized as follows: In Section 2 we provide a brief introduction to assurance cases. In Section 3 we describe our proposed method for assessing and propagating confidence through arguments. In Section 4 we apply the method to a simple example, and in Section 5 we discuss the results. Finally, in Section 6 we conclude.

2. Assurance cases

An assurance case is a structured argument, supported by evidence, to substantiate claims about a system. The way assurance cases are structured has roots in Toulmin's argumentation theory (Toulmin, 1958). The assurance case can be hierarchical, where the evidence and justifications for one claim consists of further claims supported by arguments and evidence. To organize and present assurance cases, it is common to employ Claims-Arguments-Evidence (CAE) (Adelard, 1998) diagrams or Goal-Structure Notation (GSN) (Wilson et al., 1996). (In GSN language, goals correspond to claims, solutions to evidence, and strategies to arguments in the CAE framework.) For the purpose of this paper, we adopt the CAE notation as presented in the Assurance 2.0 manifesto (Bloomfield and Rushby, 2020). One reason for this is that the GSN notation currently does not (at least officially) support expansion of justification nodes into side arguments.

Assurance 2.0 simplifies the construction of assurance cases by allowing only five types of argument steps (Bloomfield and Netkachova, 2014), namely:

(i) Evidence Incorporation (i.e., connecting a claim to evidence that directly supports or defeats it)
(ii) Concretion (i.e., supporting a claim with another more concrete, specific and measurable claim)
(iii) Substitution (i.e., supporting a claim by replacing it with another equivalent or proxy claim)
(iv) Decomposition (i.e., breaking down a claim into parts connected according to some logic relationship)
(v) Calculation (i.e., supporting a claim about the value of some property by computing it from values of other properties).

One aspect that distinguishes Assurance 2.0 from previous approaches to assurance cases is the emphasis that reasoning steps (i.e. types (ii)-(v) above) ought to be deductive. This requires justification of each reasoning step. For this reason, we make a distinction between *sub claims*, which supports another claim according to one of the allowed reasoning step types, and *side claims*, which provides justification for the chosen argument step (e.g., why a substitution or calculation is valid). Assurance 2.0 also puts emphasis on avoiding confirmation bias and identifying weak spots in arguments by explicitly including doubts that could topple or invalidate the argument as 'defeaters' (discussed in detail by Bloomfield et al. (2024)). This idea of defeaters is not unique to Assurance 2.0, and is also emphasized by Goodenough et al. (2015). A defeater may enter both as sub claims and side claims, and the argument is not sound until all defeaters have been refuted or managed in some way.

3. Assessing confidence in assurance cases

In this section we outline a method for assessing confidence in assurance cases by first doing a qualitative assessment of confidence in leaf claims/defeaters, and then propagating this confidence through the argument steps using three-valued logic. We explain the confidence propagation in Section 3.1, and the strength of knowledge evaluation in Section 3.2.

3.1. *Confidence propagation*

In general, an assurance argument can be represented in terms of conjunction (and), disjunction (or) and negation (not) relations. These extend in an obvious way to three valued logic as shown in Figure 1. Our novel contribution in this paper is

A	Not A
F	T
U	U
T	F

A and B	B = F	B = U	B = T
A = F	F	F	F
A = U	F	U	U
A = T	F	U	T

A or B	B = F	B = U	B = T
A = F	F	U	T
A = U	U	U	T
A = T	T	T	T

If B=T then A else U	B = F	B = U	B = T
A = F	U	U	F
A = U	U	U	U
A = T	U	U	T

Fig. 1. Example of truth tables for three-valued logic operations (F='false',U='uncertain',T='true'). We propose the bottom right rule for propagating confidence through argumentation steps, where A is the sub-argument and B is the justification.

the operation for propagating confidence through arguments steps involving a justification, shown in the bottom right corner of Figure 1: Considering a claim C depending on an argument A and a justification B, We have $C = A\ if\ B = true\ else\ uncertain$. In other words, if the justification is not valid, then we cannot know if the logic concluding C from its sub claims hold, and we have to conclude that $C = uncertain$ irrespective of the status of the sub claims.

3.2. *Strength-of-knowledge assessment*

Aven has proposed strength of knowledge as a measure of justified confidence in the knowledge basis of risk assessments (Aven, 2013). Criteria for classifying knowledge as strong are given in (Flage and Aven, 2009; Aven, 2013; Berner and Flage, 2016), which say that all of the following conditions must hold to classify knowledge as strong:

(i) The assumptions made are seen as very reasonable.
(ii) There is broad agreement/consensus among experts.
(iii) Sufficient amounts of reliable data are available.
(iv) The phenomena involved are well understood; the models used are known to give good predictions.

In the context of assurance cases, we may regard claims (and defeaters) as instances of propositional knowledge (Klein, 1971). Accordingly, we may ask what the strength of knowledge behind a claim (or defeater) is. However, we find that the criteria listed above need to be refined when applied to assurance case claims. For every leaf claim in an assurance case (i.e., claims with no further claims supporting it), we wish to assess if we have sufficient confidence to treat the claim as 'true' or 'false', or if we consider it to be 'uncertain'. In Table 1 we propose a set of modified criteria for strong knowledge as a measure of confidence in a claim/defeater. We consider the knowledge to be strong if all these criteria are met, and we can confidently set the claim to either 'true' or 'false'. Note that the first two criteria (direction and interpretation) reflect what the evidence is saying, while the latter two (source and method) concerns wether we trust what it says.

In cases where the knowledge is not strong, or where there is no direct evidence related to a claim or defeater, we propose a criterion in Table 2 to determine wether we can confidently assume that the claim or defeater is 'true' or 'false'. If the latter criterion is not met, we should set the claim or defeater to 'uncertain'. This is inspired by an idea from (Aven and Kristensen, 2019) that one can use and control assumptions as a strategy to deal with weak specific knowledge.

Figure 2 summarizes a flow chart for how to evaluate leaf claims and defeaters. In cases where the knowledge behind a claim is not strong and we are not comfortable making assumptions on the claim or defeater directly, it might still be possible to refine the claim using more reasoning steps, such that the assurance case can be confidently concluded conditional on additional assumptions.

4. Example

To illustrate our proposed method for assessing confidence in assurance cases, we present a simple initial assurance case for a machine-

Table 1. Criteria for strong knowledge.

Direction	The evidence is consistent in one direction (towards 'true' or 'false'), the body of evidence is large enough and representative enough that the conclusion is reliable. Counter evidence has actively been sought and included.
Interpretation	The interpretation of the evidence is uncontroversial and not ambiguous. Efforts have been made to reveal distorting biases and tacit assumptions that could hide surprises and change the conclusion.
Source	The source of the evidence is trusted and there are no conflicts of interest that could raise doubts. If not, the evidence has been verified by another trusted party with no conflicts of interest.
Method	The methods used to obtain the evidence are recognized and have been applied correctly according to best practices. Alternatively, in cases where new or unconventional methods are used, these methods have been scrutinized and evaluated to be appropriate by qualified and independent experts.

Table 2. Criteria for treating a claim as an assumption .

Control	There are reliable means to either enforce the claim or detect if it becomes invalid and implement countermeasures. Alternatively, if the claim relates to something beyond our control, it is reasonable to demand that somebody else controls it (e.g., we could make the assurance case conditional on the requirement that the claim is 'true' or 'false').

learning-based crack detection (MLCD) tool, to be compared against crack detection by human expert's physical close up inspection (EPCI). The top claim we woul like to assess is: "The MLCD detects cracks in 2D images on par with EPCI". Since the MLCD depends on images taken of the asset under inspection, we substitute the top claim for the proxy claim "MLCD hit rate in 2D test images dataset is on par with expert's hit rate in the same dataset". This latter is something we have direct evidence for, however, the substitution requires a justification. In this case we propose the following justification: "Expert detection of cracks in images used by MLCD is representative of true EPCI performance". This claim is further broken down into sub claims and justified as shown in Figure 3.

Note that one of the nodes in Figure 3 is a defeater, namely "Experts might generally detect more cracks during physical inspection than in images". We depict defeaters as claim nodes with dashed boundary, unlike the Assurance 2.0 papers (Bloomfield and Netkachova, 2014; Bloomfield and Rushby, 2020, 2022; Bloomfield et al., 2024), which depicts defeaters as claims with a different boundary color. We use the boundary colors to represent confidence, namely green (confidently 'true' for claims or confidently 'false' for defeaters), red (confidently 'false' for claims or confidently 'true' for defeaters), and orange ('uncertain', when the knowledge is too weak to determine if the claim/defeater is 'true' or 'false').

In this example we concluded that the knowledge is strong for "MLCD hit rate in 2D test images dataset is on par with expert's hit rate in the same dataset" on the basis that the test evidence was clear and statistically valid, and because we trust the people and methods behind the evidence. Similarly, for the purpose of this paper, suppose that the knowledge is strong for the claim "The distribution of the test dataset is similar to the distribution of the collected inspection data".

The uncertainty lies in the defeater "Experts might generally detect more cracks during physical inspection than in images", which propagates up to the top claim. The conclusion on the top claim is also dependent on the assumption that "Coverage of collected close-up images equal or surpass coverage of pre-planned close-up areas for physical inspection" (which we are comfortable

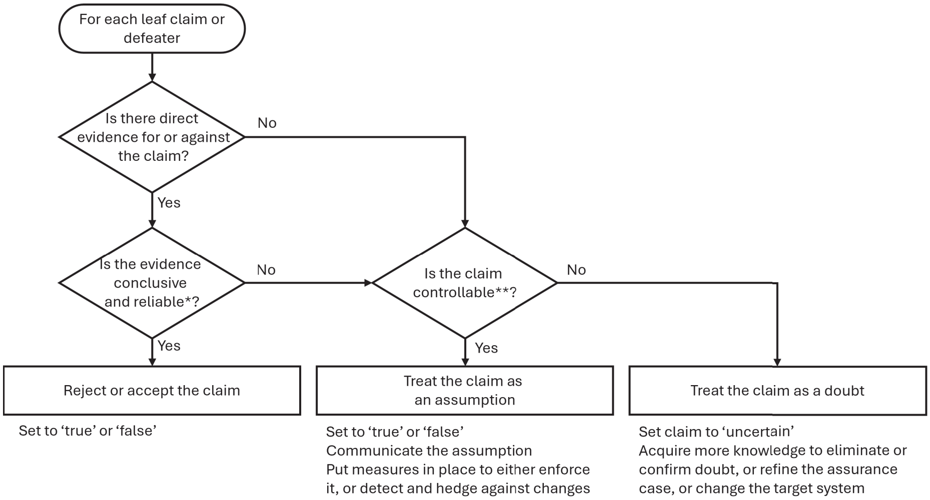

Fig. 2. Procedure for determining confidence in leaf claims/defeaters and how to treat them. (* see details in Table 1; ** see details in Table 2)

treating as an assumption because we believe we can control it and check it).

5. Discussion

Verheij (2005) discusses five types of rebuttal (defeaters), which Bloomfield et al. (2024) reduce to two types of defeaters (exact and non-exact). However, it appears to us that using our three-valued logic rule, these types of defeaters only differ by where they enter in the argument (in side or sub arguments).

Based on the confidence propagation rules explained in section 3.1, any 'uncertain' claims or defeaters will lead to the top claim also being 'uncertain'. One might therefor ask if all uncertainties are equally important? We argue that any argument with remaining uncertainties should not be accepted, so the mere presence of uncertainties should guide the assessor to refine the argument further. We do realize that our confidence propagation rule could be turned into a probabilistic rule (i.e., if we consider a probability of 0.5 to be full ignorance , then we could say that $P(C) = P(A)P(B) + 0.5(1 - P(B))$, where A is the argument for C in terms of sub claims, and B is the side claim (justification). This could allow us to do a sensitivity analysis to see which uncertainties influence the conclusion. However, we warn against this approach, because the use of probabilities may suggest a higher degree of precision in our expression of confidence than what the knowledge can support. This is exactly what we set out to avoid when we came up with our new approach.

6. Conclusion

We have proposed a method for evaluating confidence in assurance cases which simplifies previously proposed approaches by recasting the problem in terms of three-valued logic and strength-of-knowledge assessment. Our method is not yet tested and validated on real-world cases, but we hope this paper can stimulate discussions and synergies between the risk assessment and assurance case research communities.

References

Adelard, L. (1998). Ascad: Adelard safety case development manual. *Adelard LLP, London.*

Amundrud, Ø., T. Aven, and R. Flage (2017). How the definition of security risk can be made compatible with safety definitions. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability 231*(3), 286–294.

Aven, T. (2013). Practical implications of the new

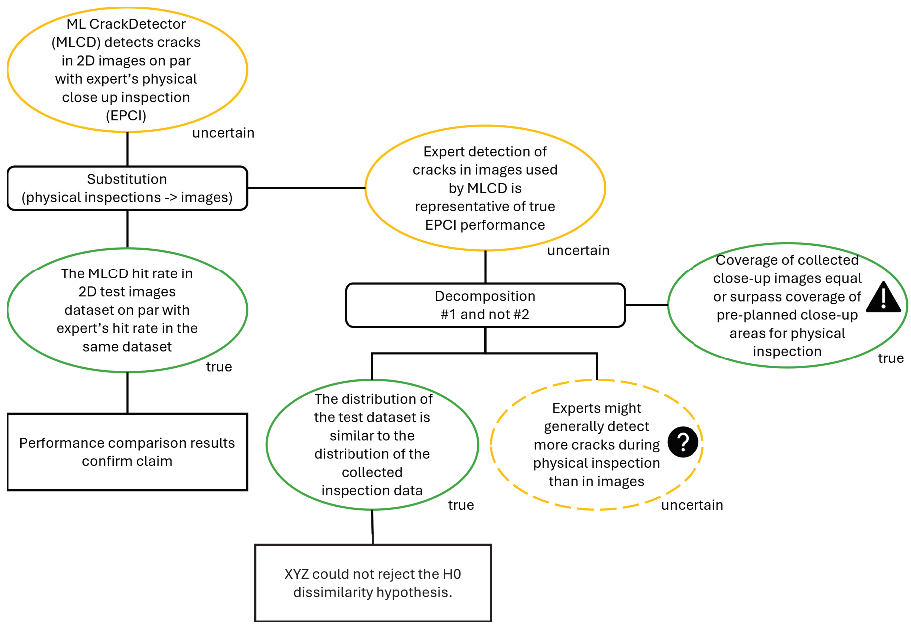

Fig. 3. Example of the initial iteration of an assurance case for a machine-learning-based crack detection tool. Assumptions that we are confident about are indicated with '!', and residual doubts are indicated with '?'.

risk perspectives. *Reliability Engineering & System Safety 115*, 136–145.

Aven, T. (2018). An emerging new risk analysis science: Foundations and implications. *Risk Analysis 38 (5)*, 876–88.

Aven, T. and V. Kristensen (2019). How the distinction between general knowledge and specific knowledge can improve the foundation and practice of risk assessment and risk-informed decision-making. *Reliability Engineering & System Safety 191*, 106553.

Berner, C. and R. Flage (2016). Strengthening quantitative risk assessments by systematic treatment of uncertain assumptions. *Reliability Engineering & System Safety 151*, 46–59.

Bloomfield, R. and K. Netkachova (2014). Building blocks for assurance cases. In *2014 IEEE International Symposium on Software Reliability Engineering Workshops*, pp. 186–191. IEEE.

Bloomfield, R., K. Netkachova, and J. Rushby (2024). Defeaters and eliminative argumentation in assurance 2.0. *arXiv preprint arXiv:2405.15800*.

Bloomfield, R. and J. Rushby (2020). Assurance 2.0: A manifesto. *arXiv preprint arXiv:2004.10474*.

Bloomfield, R. and J. Rushby (2022). Assessing confidence with assurance 2.0. *arXiv preprint arXiv:2205.04522*.

Cyra, L. and J. Górski (2008a). Expert assessment of arguments: A method and its experimental evaluation. In *Computer Safety, Reliability, and Security: 27th International Conference, SAFECOMP 2008 Newcastle upon Tyne, UK, September 22-25, 2008 Proceedings 27*, pp. 291–304. Springer.

Cyra, L. and J. Górski (2008b). Supporting expert assessment of argument structures in trust cases. In *9th International Probability Safety Assessment and Management Conference PSAM*. Citeseer.

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

Duan, L., S. Rayadurgam, M. P. Heimdahl, O. Sokolsky, and I. Lee (2015). Representing confidence in assurance case evidence. In *Computer Safety, Reliability, and Security: SAFECOMP 2015 Workshops, ASSURE, DECSoS. ISSE, ReSA4CI, and SASSUR, Delft, The Netherlands, September 22, 2015, Proceedings 34*, pp. 15–26. Springer.

Flage, R. and T. Aven (2009). Expressing and communicating uncertainty in relation to quantitative risk analysis. *Reliability & Risk Analysis: Theory and Application 2 (13)*, 9–11.

Goodenough, J. B., C. B. Weinstock, and A. Z. Klein (2015). Eliminative argumentation: A basis for arguing confidence in system properties. *Software Engineering Institute, Carnegie Mellon University, Pittsburgh, PA, Tech. Rep. CMU/SEI-2015-TR-005*.

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

Guiochet, J., Q. A. Do Hoang, and M. Kaaniche (2015). A model for safety case confidence assessment. In *Computer Safety, Reliability, and Security: 34th International Conference, SAFECOMP 2015, Delft, The Netherlands, September 23-25, 2015, Proceedings 34*, pp. 313–327. Springer.

Guo, B. (2003). Knowledge representation and uncertainty management: applying bayesian belief networks to a safety assessment expert system. In *International Conference on Natural Language Processing and Knowledge Engineering, 2003. Proceedings. 2003*, pp. 114–119. IEEE.

Hobbs, C. and M. Lloyd (2011). The application of bayesian belief networks to assurance case preparation. In *Achieving Systems Safety: Proceedings of the Twentieth Safety-Critical Systems Symposium, Bristol, UK, 7-9th February 2012*, pp. 159–176. Springer.

ISO/IEC/IEEE 15026 (2019). ISO/IEC/IEEE 15026: Systems and software engineering – systems and software assurance. Part 1: Concepts and vocabulary.

Klein, P. D. (1971). A proposed definition of propositional knowledge. *the Journal of Philosophy 68*(16), 471–482.

Lane, R. (2001). Triadic logic.

Nair, S., N. Walkinshaw, and T. Kelly (2014). Quantifying uncertainty in safety cases using evidential reasoning. In *Computer Safety, Reliability, and Security: SAFECOMP 2014 Workshops: ASCoMS, DECSoS, DEVVARTS, ISSE, ReSA4CI, SASSUR. Florence, Italy, September 8-9, 2014. Proceedings 33*, pp. 413–418. Springer.

Nair, S., N. Walkinshaw, T. Kelly, and J. L. de la Vara (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. IEEE.

Nešić, D., M. Nyberg, and B. Gallina (2021). A probabilistic model of belief in safety cases. *Safety science 138*, 105187.

Toulmin, S. E. (1958). *The Uses of Argument*. Cambridge University Press.

Verheij, B. (2005). Evaluating arguments based on toulmin's scheme. *Argumentation 19*, 347–371.

Wang, R., J. Guiochet, G. Motet, and W. Schön (2019). Safety case confidence propagation based on dempster–shafer theory. *International Journal of Approximate Reasoning 107*, 46–64.

Wilson, S. P., J. A. McDermid, C. H. Pygott, and D. J. Tombs (1996). Assessing complex computer based systems using the goal structuring notation. In *Proceedings of ICECCS'96: 2nd IEEE International Conference on Engineering of Complex Computer Systems (held jointly with 6th CSESAW and 4th IEEE RTAW)*, pp. 498–505. IEEE.

Yuan, C., J. Wu, C. Liu, and H. Yang (2017). A subjective logic-based approach for assessing confidence in assurance case. *International Journal of Performability Engineering 13*(6), 807.

Zhao, X., D. Zhang, M. Lu, and F. Zeng (2012). A new approach to assessment of confidence in assurance cases. In *Computer Safety, Reliability, and Security: SAFECOMP 2012 Workshops: Sassur, ASCoMS, DESEC4LCCI, ERCIM/EWICS, IWDE, Magdeburg, Germany, September 25-28, 2012. Proceedings 31*, pp. 79–91. Springer.