

Balancing the Formal and Informal in Safety Case Arguments

Ibrahim Habli, Tim Kelly
Department of Computer Science
University of York
York, YO10 5GH, UK
+44 1904 325477
{ibrahim.habli, tim.kelly}@york.ac.uk

ABSTRACT

In many safety-critical industries, developers and operators are required to construct and present well reasoned arguments that their systems achieve acceptable levels of safety. These arguments (together with supporting evidence) are typically referred to as a “safety case”. Safety arguments historically have been communicated through narrative text, leading often to problems of comprehension and communication. Over the last twenty years, there has been increasing interest in using structured argumentation notations such as GSN (Goal Structuring Notation) or CAE (Claims-Argument-Evidence) to communicate the structure of the argument. Whilst such arguments are structured, they remain informal. There is increasing interest in exploring how these informal arguments may be modelled in formal logic, potentially opening up benefits of forms of analysis not possible with informally recorded argument. This position paper discusses a number of considerations in balancing the role of informal and formal logic in modelling safety case arguments.

Keywords

Safety cases, argumentation, logic.

1. INTRODUCTION

The purpose of a safety case can be defined in the following terms:

A safety case should communicate a clear, comprehensive and defensible argument (supported by evidence) that a system is acceptably safe to operate in a particular context.

The safety case argument is used to *demonstrate* how someone can reasonably conclude that a system is acceptably safe from the evidence available. A safety case is a device for *communicating* ideas and information, usually to a third party (e.g. a regulator or an independent assessor). In order to do this convincingly, it must be as clear as possible.

The following three types of arguments are commonly found within a safety case [1] [2]:

- (Causal) Behavioural **arguments of risk management**, i.e. how the causes of hazards are eliminated or mitigated, or how the consequences of hazards are mitigated.
- **Confidence arguments** – arguments that provide confidence in the adequacy of the details of the risk management argument, e.g. justifying the adequacy of hazard identification techniques, or the sufficiency of verification results presented.

- **Arguments of conformance / compliance** with safety standards, regulations, and legislation – where compliance is not straightforward it is necessary to justify how a project, system design and operation have addressed legal and regulatory obligations.

Safety arguments historically have been communicated in safety cases through narrative text. The biggest problem with the use of free text is in ensuring that all stakeholders involved share the same understanding of the argument. Without a clear and shared understanding of the argument, safety case management is an inefficient approach to regulation. To address this problem, structured argumentation notations such as GSN (Goal Structuring Notation) [3] or CAE (Claims-Argument-Evidence) [4] have been increasingly adopted as a means of communicating the structure of the argument. For example, GSN explicitly represents the individual elements of a 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).

Although representing an argument using a notation such as GSN clearly disambiguates the structure and elements of the argument, it cannot ensure that the argument itself is ‘good’ or sufficient for its purpose. A safety argument includes a number of *assertions*. These assertions relate to the sufficiency and appropriateness of the inferences declared in the argument, the context and assumptions used and the evidence cited. (A documented safety argument can be said to be merely a documented position that collects together these assertions, which should then be reviewed and challenged by the system stakeholders, e.g. operators, developers, analysts and regulatory bodies.)

The underlying domain (subject matter) of safety case arguments means that they are rarely *provable* deductive arguments. Instead they are more commonly inductive. A *deductive* argument is an argument that proceeds without any room for probability [5]. An *inductive* argument is one that is based upon the estimation of the probable truth of the premises [5]. In an inductive argument, the probable truth of the premises is passed through the argument to the conclusion.

2. FORMALISING THE INFORMAL

There is increasing interest in exploring how these informal safety arguments may be modelled in formal logic (e.g. [6] [7] [8]), potentially opening up benefits of forms of analysis not possible with informally recorded arguments. The informality of the underlying reasoning present in safety assurance cannot be

eliminated (e.g. justification of the adequacy of the interpretation of a safety regulation, or the domain experience of personnel involved in hazard analysis). However, the informal arguments can be represented by formal logic. For example, it is possible to attempt to define formal¹ semantics for notations such as GSN and CAE. This would enable automatic derivation of a formal characterisation of an argument, which could be used to uncover reasoning flaws using proof or model checking. Such an approach would need to accommodate assumptions and side conditions concerning the main inferences of a safety argument.

3. CONSIDERATIONS

Whilst formalisation of safety case arguments may be seen as the natural ‘next step’ in improving precision, there are a number of potential considerations to be addressed in this endeavour, including, but not limited to, the following:

How does formalisation of arguments address current needs in safety case practice?

Current safety case practice is not without criticism and challenges. The Nimrod Review [9] cited a number of concerns with current safety case practice, including problems of focus (inadequate attention to the core concern of risk management), scale and outsourcing of production. The principal concern with the specific example of the Nimrod XV230 safety case was that it was quickly constructed and lacked sufficient evidence to justify the claims made (regarding the acceptable mitigation of all identified hazards), i.e. the principal problem was that of validity (was it ‘accurate’?). Others have highlighted the potential for bias (e.g. confirmation or selection bias) in the development of safety arguments and evidence.

Whilst the automation possibilities of formalisation (such as the ability to verify the ‘well-formedness’ of an argument) offer benefits that help address some of these challenges (e.g. scale), other problems remain (e.g. those concerning the validity and veracity of arguments recorded – whether formally or informally). To go beyond the position of simply recording assertions concerning the subject domain, and being able to reason about the truth of those assertions relies upon establishing a connection between formal models of the subject domain as well as formal models of the argument (i.e. formal evidence as well as argument).

Others have highlighted the problems of fallacious reasoning in safety case practice [10]. In the field of argumentation and informal logic, the problem of fallacies is well recognised. Damer in [11] describes the following five categories of argument fallacy:

- (1) structural flaw in the argument;
- (2) irrelevance of a premise;
- (3) unacceptability of a premise;
- (4) insufficiency of the combined premises of an argument to establish its conclusion; and
- (5) failure to give an effective rebuttal.

¹ At present the semantics of notations such as GSN, or assurance case standards such as the OMG SACM is defined by natural language (e.g. as in the GSN standard) or by metamodel (e.g. as in the SACM standard).

Formalisation would help in dealing with the first category. The remainder are issues of validity and may prove difficult to address with formalisation. Annotation of informal logic with information concerning argument schemes [12] may enable formalisation to help with consistency checking with prior knowledge of acceptable (types of) argument structure. However, such an approach would rely upon correct annotation (i.e. to enable checking would rely upon the author knowing that there was an issue to be checked).

Are all types of safety case argument equally amenable to formalisation?

Informal arguments can be represented formally. However, formalisation often involves axiomatising (informal) aspects of the argument at the ‘edge’ of our argument. A good example of this is the often-used style of a safety case argument that is structured over identified hazards. The logic that says (adequately) addressing each of the identified hazards leads to a claim of acceptable system safety rests on the assumption that all hazards have been identified. It is easy to say that if this is true, the argument is valid. However, the claim of having identified all hazards is not itself provable. Instead it relies itself on (inherently) informal arguments and evidence (such as the adequacy of the domain experience of those involved in identifying the hazards). Of course, it is possible to further formalise these informal arguments. However, doing so ‘kicks the can down the road’ and will ultimately result in a further set of axioms covering the informal aspects of the formalised argument.

It can be argued that a valuable service has been performed by ‘annexing’ the informal arguments to an easily identified location (a form of reductionism). However, a possible concern that an illusion of formality is created through having hidden problematic informal and subjective arguments behind an abstraction. It can be questioned whether a formalised ‘core’ with informality pushed to the periphery of the formalisation is advantageous or dangerous for evaluation and review of safety case arguments. Tool support (through formalisation) for checking arguments is potentially useful. However, formalisation will not reduce perhaps the most significant aspect of the review burden – namely individual review and acceptance of subjective (informal) assertions of the argument.

Does the subject matter of a safety case argument affect the value of formalisation?

Whilst many safety case domain arguments are inductive in nature, it is possible for deductive arguments to form part of a safety case. This is particularly true when the subject matter domain of the safety argument is itself logical (e.g. concerning the behaviour of the software in a software-intensive system). Where the underlying behaviour of the subject domain is deductive, asserted inferences can become provable inferences. However, this raises another question: When safety case arguments (or at least portions of them can become provable) are they perhaps not better represented as *evidence* (i.e. proof), rather than as informal logic?

It could be argued that the value of a safety case is to represent the informal logical ‘glue’ that pulls together different forms of the evidence (including deductive results – proof being one such example). It is possible to question whether deductive arguments are best represented using formalised models of informal argumentation notations.

What are the tradeoffs between accessibility and precision in safety arguments?

One position is that the value of safety cases lies in *facilitating accessibility* of safety arguments amongst key stakeholders. Informal argumentation notations (e.g. with the use of graphical representations) are designed with this position in mind. Another position is that the value of safety cases lies in removing ambiguity and improving *precision* in the description of safety arguments. Formalisation of arguments could be argued to support this position. Both these positions (accessibility and precision) are important, and *can* be reinforcing. However, there may be a point where precision and accessibility start to be in conflict – i.e. precision is achieved at the expense of comprehensibility.

4. SUMMARY OF POSITION

Safety case arguments are often informal (inductive, non-provable) arguments expressed using informally defined argumentation notations (such as GSN and CAE). There is growing interest in formalisation of the representation of safety arguments. Formalisation of the representation of inherently informal arguments raises points of discussion concerning the value gained over merely ‘structured’ (model-driven) approaches, trade-offs between precision and accessibility, and whether all forms of argument are equally amenable to formalisation. These points can and should be debated and considered as work on formalisation of safety case arguments progresses.

5. REFERENCES

- [1] Hawkins, R. Kelly, T., Knight, J., and Graydon P. 2011. A New Approach to Creating Clear Safety Arguments, *Advances in Systems Safety*, Safety-critical Systems Symposium SSS '11, (Southampton February 2011).
- [2] Graydon, P., Habli, I., Hawkins, R., Kelly, T., and Knight, J. 2012. Arguing Conformance, *IEEE Software*, 29, 3, (May/June 2012).
- [3] Goal Structuring Notation (GSN) Community Standard, Issue 1, 2011, available www.goalstructuringnotation.info
- [4] Bloomfield, R.E., et al. 1998. *Ascad – Adelard Safety Case Development Manual*.
- [5] Cass, M., and Le Poidevin, R. 1993. *A Logic Primer*. Vortex Publishing.
- [6] Rushby, J. 2013, Logic and Epistemology in Safety Cases. *International Conference on Computer Safety, Reliability and Security, SAFECOMP 2013* (Toulouse, September 2013).
- [7] Denney, E., and Pai, G. 2013. A Formal Basis for Safety Case Patterns. *International Conference on Computer Safety, Reliability and Security, SAFECOMP 2013* (Toulouse, September 2013).
- [8] Matsuno, Y. 2014. A Design and Implementation of an Assurance Case Language. *44th Annual IEEE/IFIP International Conference on Dependable Systems and Networks DSN* (June 2014).
- [9] Haddon-Cave, C. 2009. *The Nimrod Review: an independent review into the broader issues surrounding the loss of the RAF Nimrod MR2 aircraft XV230 in Afghanistan in 2006 report*. Report No. No 1025 2008–09, Published by the Stationery Office (TSO).
- [10] Greenwell, W., Knight, J., Holloway, M., and Pease, J. 2006. A Taxonomy of Fallacies in System Safety Arguments. *2006 International System Safety Conference (ISSC)*, System Safety Society.
- [11] Damer, T. E. 2009. *Attacking Faulty Reasoning: A Practical Guide to Fallacy-free Arguments*. (6th ed.). Belmont, California: Wadsworth.
- [12] Walton, D., 2008. *Argumentation Schemes*, Cambridge University Press.