# A New Approach to Creating Clear Safety Arguments

**Richard Hawkins[1], Tim Kelly[1], John Knight[2] and Patrick Graydon[2]**

[1]University of York, UK

[2] University of Virginia, Charlottesville, USA

**Abstract** We introduce *assured safety arguments*, a new structure for arguing safety in which the *safety argument* is accompanied by a *confidence argument* that documents the confidence in the structure and bases of the safety argument. This structure separates the major components that have traditionally been confused within a single safety argument structure. Separation gives both arguments greater clarity of purpose, and helps avoid the introduction of superfluous arguments and evidence. In this paper we describe a systematic approach to establishing both arguments, illustrated with a running example.

## 1 Introduction

In this paper, we introduce a new structure for arguing safety termed an *assured safety argument*. An assured safety argument has two components:

- a safety argument that documents the arguments and evidence used to establish direct claims of system safety
- a confidence argument that justifies the sufficiency of confidence in this safety argument.

These two components are both stated explicitly but separately. They are interlinked so that the justification for having confidence in individual aspects of the safety argument is clear and readily available but not confused with the safety argument itself. This separation eliminates several difficulties with traditional approaches and provides several advantages.

The role of a safety case is to provide:

> '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' (MoD 2007).

A safety argument must explain how the available evidence supports the overall claim of acceptable safety. Best practice, risk-based, safety arguments decompose this claim 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.

Unfortunately, both evidence and argument will typically be imperfect. For example, software testing may fail to support the claims for which it is cited for a variety of reasons including:

- inadequately defined test cases (e.g. that fail to fully capture the safety requirements)
- imperfect test coverage
- a faulty test oracle
- the failure of human testers to follow the test procedure faithfully
- testers inadvertently testing a different version of the system or component
- test results corrupted between collection and analysis.

There are numerous scenarios in which the reality of failures of the computer hardware and software together with the fallibilities of the test generation process could result in false conclusions (claims) being drawn from that evidence. Having sufficient confidence in safety claims is essential.

Any knowledge gap that prohibits perfect (total) confidence is referred to as an *assurance deficit*. In establishing an argument of safety it is first important to identify and acknowledge the assurance deficits that (inevitably) exist. Having recognised the assurance deficits, the goal is to explicitly manage them such that the overall confidence in the safety argument is considered acceptable.

Present practice is to develop a single, unified safety argument that does not distinguish the arguments of safety and confidence. This practice merges what are essentially two different but interrelated arguments. Both of these elements are essential to a *compelling* safety argument, but presenting both in an intermingled fashion typically results in a larger (often rambling) argument and makes grasping the crucial structures difficult for the reader. Clarity of presentation is important for all stakeholders even though their interests might differ. For developers, the distinction between the safety and confidence arguments would help provide clearer direction on the steps involved in constructing each argument and a better understanding of the necessary development and assurance steps. For reviewers, the distinction would help focus attention on those aspects of the argument that are weakly supported.

An assured safety argument separates the argument about assurance deficit into a separate confidence argument in order to address this problem. The *safety argument* documents the asserted arguments and evidence of risk reduction. The *confidence argument* documents the reasons for having confidence in the safety argument.

A truly risk-based safety argument must always be focused upon the identification and mitigation of hazards associated with the system. The safety argument

demonstrates how the risks associated with each hazard are managed. Everything cited in the safety argument should therefore have a direct role as part of the causal chain to the hazard. That is, all of the goals in the safety argument must be claims about the system or parts, properties, or properties of parts thereof. Artefacts from system development (e.g. test reports and, by extension, their contents) may be referenced only in solution or context elements. Strict adherence to this tight definition of a safety argument ensures the focus of the safety argument is clearly on the (direct) management of risk. We will describe later how safety arguments may be structured.

A confidence argument demonstrates 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 those uncertainties and explain why there is sufficient confidence in the safety argument. Figure 1 represents a complete assurance argument entailing the safety argument 'encapsulated' by a confidence argument. We will describe later how confidence arguments are used to create the overall assurance case for the system.

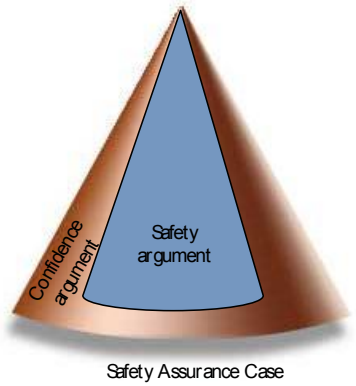


**Fig. 1.** A safety assurance case containing separate safety and confidence arguments

In the next section, we elaborate the difficulties that arise when both safety and confidence are argued in a single, integrated argument. In section 3, we discuss the construction of assured safety cases. In section 4, we illustrate our safety argument structure by presenting portions of a safety argument and confidence argument for a hypothetical drug infusion pump. Finally, we conclude in section 5.

## 2 The difficulties with a single argument

The present practice of including in a single argument elements that document both direct arguments of (product) mitigation and supporting arguments that are 'confidence-raising' leads to a number of difficulties including:

- Arguments tend to become large and unwieldy, because there is too much information in one argument. The entry criterion for the inclusion of an argument (or item of evidence) in the safety argument is often (too loosely), 'Does this have any possible bearing on the safety of the system?' Both direct arguments of risk reduction *and* (any) indirect arguments of confidence are admitted by this criterion. This can lead to voluminous, rambling, ad infinitum arguments.
- Both the safety argument and the confidence argument tend to be poorly prepared, because the lack of distinction between the two makes it more difficult to spot incompleteness or poor structure in either.
- Necessary elements of the argument are sometimes omitted, because the need for the specific elements is lost in the volume of the argument.
- Arguments become indirect and unfocused, and the link between elements of the argument and risk is often lost.
- Unnecessary material is sometimes included in arguments without proper consideration or explanation of its relevance – 'just in case'.
- Arguments become difficult to build, and weaknesses of the argument are sometimes not evident and so are easily overlooked.
- Arguments become difficult to review because of the size and lack of focus.

These difficulties are serious since they all detract from the basic purposes of using safety cases. We note that many of the problems with current practice in the application of safety cases were highlighted by (Haddon-Cave 2009).

Separation of the safety and confidence arguments offers the opportunity to mitigate these difficulties by providing different foci for safety and confidence. In addition, careful attention to linking the two arguments provides a mechanism for guiding analysis of the interrelationship between safety and confidence.

## 3 Constructing assured safety arguments

A safety argument must always be focused upon the identification and mitigation of hazards associated with the system. The safety argument demonstrates how the risks associated with each hazard are managed. Everything that is included as part of the safety argument should therefore have a direct role as part of the causal chain to the hazard. Anything that does not fulfil this role should not be included in the safety argument. Safety arguments are constructed by providing claims relating to the safety of the system. These claims are then broken into sub-claims that show how the top-level safety claim is demonstrated. The decomposition of

claims and sub-claims continues until a point is reached where a claim can be supported by citing a development or assessment artefact (e.g. a design analysis report or test report) as evidence. The strategy adopted when supporting a claim should be made explicit in the argument. The argument should also clearly state the context in which the argument is made, along with any assumptions that have been made. When arguments are communicated solely through narrative text it can often be difficult for a reader to identify the individual elements (e.g. distinct claims) and structure (e.g. asserted inferences) of the argument. It is therefore often clearer to represent a safety argument graphically. Figure 2 shows a simplified example of how a safety argument structure may be captured using the Goal Structuring Notation (GSN). We refer readers unfamiliar with the GSN notation to (Kelly and Weaver 2004).

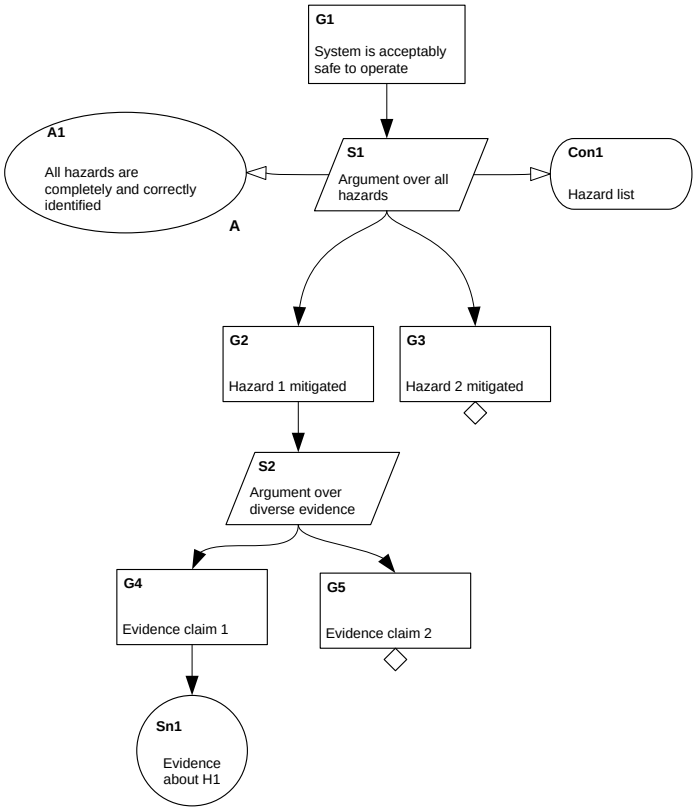


**Fig. 2.** An example safety argument represented using GSN

Although representing an argument graphically clearly disambiguates the structure and elements of the argument, it cannot ensure that the argument itself is 'good' or sufficient for its purpose. By exercising discipline over the permissible claims and evidence of the safety argument, and encouraging a systematic approach to the construction of a confidence argument, we can begin to address this issue.

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 is merely a documented position that collects together these assertions.) To be compelling, the argument must justify the truth of the assertions made. If an argument assertion cannot be justified, then the argument will not be believed (it will not provide the required assurance). The confidence argument provides the justification for argument assertions. In order to indicate the assertion in the safety argument that the confidence argument is associated with, the confidence argument is tied to a number of *Assurance Claim Points* (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 3 shows ACPs named ACP1, ACP2 and ACP3.

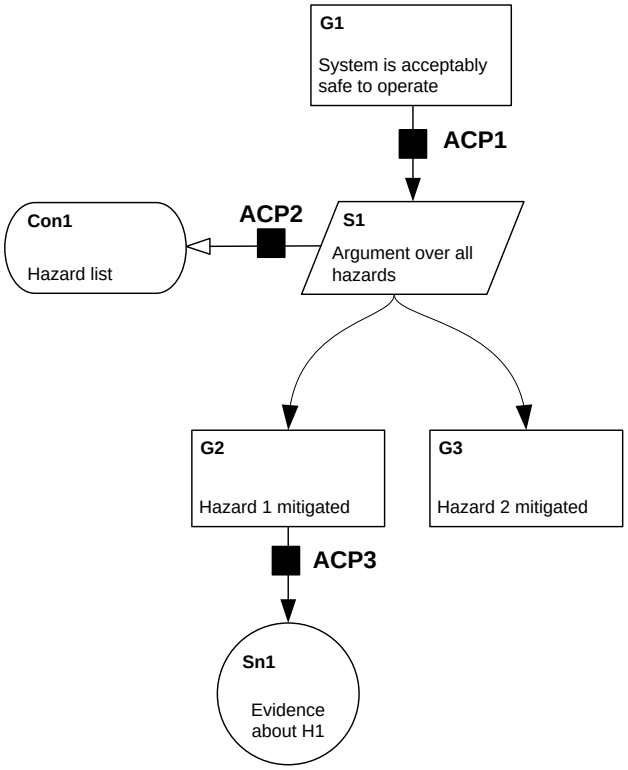


**Fig. 3.** Example of the use of ACPs

These ACPs correspond to three different types of assertion:

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

Below we discuss each of these three types of assertion in more detail.

### *3.1 Asserted inference*

Each time a claim is said to be supported by other claims in an argument, an assertion is being made that the inference is appropriate and sufficient. Only in deductive arguments do premise claims *prove* a particular conclusion. Instead, for inductive arguments, the assertion is that the probable truth of the premises is sufficient to establish the probable truth of the conclusion. Although safety cases can contain a mix of both deductive and inductive arguments, inductive arguments typically dominate. For example, Figure 4 shows (in GSN) the assertion that, given the applicable context, the sub-claims put forward to implement the chosen argument strategy are, if true, a sufficient basis upon which to infer the conclusion stated in the parent claim. To gain assurance in the adopted argument strategy, it is necessary to provide a confidence argument that demonstrates why the asserted inference should be believed. The ACP for an asserted inference is the link between the parent claim and its strategy or sub-claims.

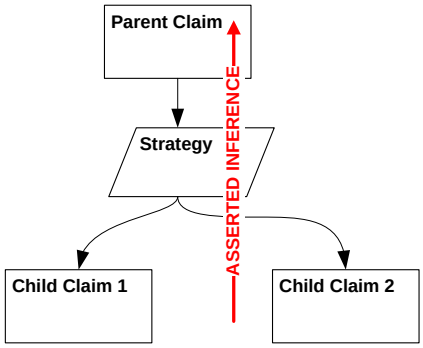


**Fig. 4.** Asserted inference

In the example shown below in Figure 5, the asserted inference is that if all hazards are mitigated then the system is acceptably safe to operate. The role of the confidence argument for ACP1 is to demonstrate why it should be believed that the two supporting claims of hazard mitigation are sufficient to draw the overall conclusion about system safety. We discuss how such a confidence argument may be constructed later.

### *3.2 Asserted context*

Each time contextual information (represented by context or assumption elements) is introduced into the argument, it is being asserted that the context is appropriate for the argument elements to which it applies. For example, consider a context reference to a list of failure modes for a particular piece of equipment. The introduction of this context element when arguing about the safety of that piece of

equipment implicitly asserts that the list of failure modes referred to is appropriate to the application and operating context in question.

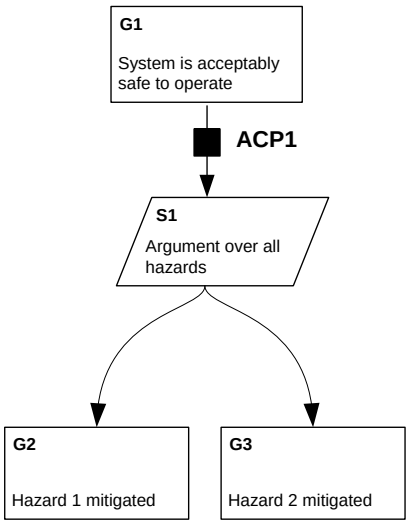


**Fig. 5.** ACP relating to an asserted inference

Figure 6 shows asserted context for an argument strategy. The assurance of the strategy depends upon the confidence that the context or assumption stated is appropriate for that strategy and its sub-goals. It is necessary to provide a confidence argument that demonstrates why it should be believed that the asserted context is appropriate. In addition to the appropriateness of the context, it is also necessary to provide an argument as to the trustworthiness of the context in question. The concept of trustworthiness relates to freedom from flaw. In the legal field the notion of integrity of evidence is often used to refer to the soundness or quality of the evidence put forward in a case. In considering the trustworthiness of an artefact, the processes used to generate that artefact are often considered (Habli and Kelly 2007). The ACP for asserted context is the link to the contextual element.

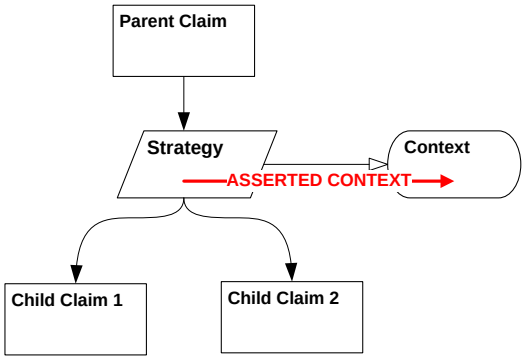


**Fig. 6.** Asserted context

In the example shown below in Figure 7 it is being asserted that the hazards given in the referenced hazard list are the relevant hazards. For this context to be appropriate there must be confidence that the hazard list is appropriate with respect to the system, application and context. The role of the confidence argument at ACP2 is therefore to demonstrate why it should be believed that citing this hazard list defines the appropriate context at this point in the safety argument. In addition, it is necessary to justify the trustworthiness of the hazard list. We discuss how such a confidence argument may be constructed later.

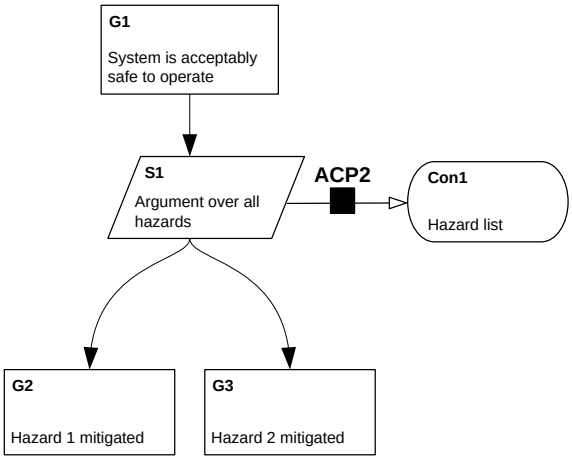


**Fig. 7.** ACP relating to an asserted context

For completeness, a confidence argument should be provided for both the inference *and* the context (ACP1 *and* ACP2), as shown in Figure 8. It is important to provide separate confidence arguments because each relates to a separate assertion.

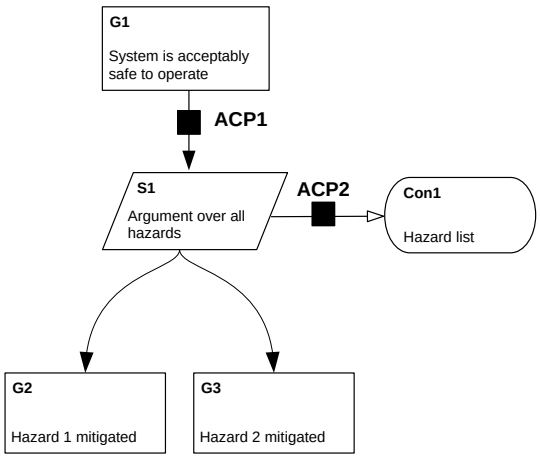


**Fig. 8.** ACPs relating to asserted inference and asserted context

### *3.3 Asserted solution*

Each time evidence is referenced as a solution to the argument, it is being asserted that the evidence put forward is sufficient to support the claim. Figure 9 shows an asserted solution to a safety claim. The assurance of the solution depends upon the confidence that the evidence is appropriate to support the claim, and the evidence is trustworthy. The ACP for asserted solutions is the link to the solution element.

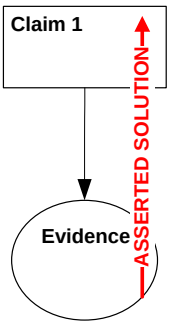


**Fig. 9.** Asserted solution

In the example shown below in Figure 10 it is being asserted that the stress testing results are sufficient to demonstrate that the defined operational forces can be tolerated. For this solution to be sufficient there must be confidence that the stress testing performed is good enough for this purpose. The role of the confidence argument at ACP3 is to provide this confidence. This will involve considering whether the stress testing of the type being referred to is adequate to support the claim and whether the stress testing procedure was followed faithfully. We discuss how such a confidence argument may be constructed later.

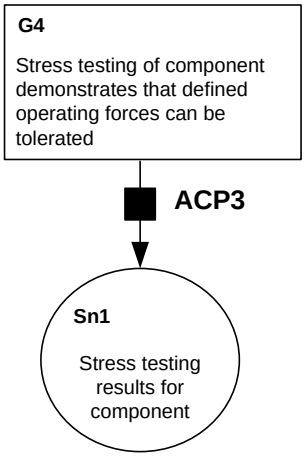


**Fig. 10.** ACP relating to an asserted solution

### *3.4 Confidence argument structure*

In all but a very few situations, the truth of the assertions put forward within a safety argument cannot be demonstrated with certainty. It is necessary to demonstrate that there is sufficient confidence in each assertion. This is the role of the confidence argument.

The issue of sufficiency with regard to confidence in an assertion is complex. The notion of reducing risk to a level that is As Low As Reasonably Practicable (ALARP) has become widely accepted (HSE 2001). Risk is a quantifiable entity (i.e. the expected loss). Confidence is also quantifiable (i.e. the probable truth of a claim). However, except where purely statistical evidence is used, to reason about confidence quantitatively requires first qualitative reasoning about the sources of uncertainty in arguments. In our approach we focus on these qualitative arguments, and leave the subsequent 'encoding' and quantification of these to those who believe that quantification can reap further utility.

We require a qualitative argument to demonstrate sufficient confidence in an assertion. This argument demonstrates why a sceptical audience should believe three important things about the assertion:

- There are grounds to support the probable truth of the assertion.
- Residual uncertainties (assurance deficits) in the assertion have been identified.
- The residual uncertainties (assurance deficits) in the assertion are insufficient to cause concern.

The first aspect of this argument considers the reasons why the assertion should be believed. This aspect is realized as the decomposition of a goal of the form 'the assertion <*x*> is true'. As in the safety argument, goal decomposition continues until the goal can be solved with evidence. Unlike the safety argument, however, the goals in this portion of the confidence argument are typically expected to be claims about properties of development artefacts (i.e. 'process' claims). For example, the decomposition of a solution assertion goal might contain arguments over the properties of test plans, development tools, and configuration management systems. Goal decomposition in this portion of the confidence argument should continue until no reasonable observer would deny that the artefact cited offers positive evidence in support of the goal claim.

The second aspect of the argument involves justifying that the uncertainties (assurance deficits) surrounding the assertion have been identified. The final (third) aspect of the argument, must argue the acceptability of the uncertainties (assurance deficits) that remain.

The identification of an assurance deficit identifies a gap in our knowledge relating to an assertion in the argument. One reason that assurance deficits are of interest is that they represent 'blind spots' in the argument – i.e. areas of the argument where no evidence has been presented. Should these 'blind spots' be eliminated (by providing the appropriate evidence) we may find that the evidence is positive (and supports the assertion made in the safety argument). However, we

may also find that the evidence is negative and forms counter-evidence to the safety argument. Recognising assurance deficits, therefore, helps identify the possible areas in the argument where counter-evidence *may* exist. (This guiding of the otherwise boundless search for counter-evidence is a useful side-effect of the identification of assurance deficits.) For example, consider a case where there is no control flow analysis evidence of the absence of infinite loops in some source code. When arguing that a return value will always be provided, we should consider the probability of the existence of counter-evidence to our claim (i.e. if we were to provide the control flow analysis – how probable is it that an infinite loop will be detected?)

It is necessary to identify assurance deficits as completely as practicable and to justify that the residual assurance deficits can be accepted. Creating an assured safety argument in the manner we have described makes it easier to identify the important assurance deficits, since the structure demands a systematic consideration of the weaknesses in the argument. It is possible to mitigate any identified assurance deficits by taking one of four actions:

- making changes to the design of the system, e.g. adding a hardware backup when it is impractical to demonstrate with adequate confidence that software has the properties necessary to ensure system safety
- making changes to system operation, e.g. by limiting the conditions under which the system is used
- making changes to the safety argument, e.g. adding an independent source of evidence
- generating additional evidence for the confidence argument, e.g. increasing the coverage of software functional tests.

It is important to note at this point that completely mitigating all assurance deficits is not normally achievable. In many cases it would be possible to go on forever generating additional evidence to try to gain some additional confidence. It is therefore necessary to make a judgment on when assurance deficits can be tolerated. To do this it must be shown that the cost (effort) expended in addressing an assurance deficit reflects the risk associated with that assurance deficit. The risk associated with an assurance deficit can be assessed by expert judgment of the likelihood of any event chains that would lead to the assertion being false and of how damaging it would be to the main safety argument if the claim were false. Considering the likelihood and severity of counter-evidence may help in making such judgements.

We show the potential structure of confidence arguments using the GSN pattern notation (Kelly 1998). To create argument patterns, GSN is extended to support multiplicity, optionality and abstraction. The multiplicity extensions shown in Figure 11 are used to describe how many instances of one entity relate to another entity. They are annotations on existing GSN relational arrows. The optionality extension is used to denote possible alternative support. It can represent a 1-of-n or an m-of-n choice. In Figure 11, one source node has three possible alternative sink nodes.

| | |
|---|---|
| n | A solid ball is the symbol for many (meaning zero or more). The label next to the ball indicates the cardinality of the relationship. |
| | A hollow ball indicates "optional" (meaning zero or one). |
| | A line without multiplicity symbols indicates a one to one relationship (as in conventional GSN). |

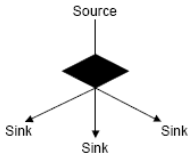


**Fig. 11.** GSN multiplicity and optionality extensions

The abstraction extensions shown in Figure 12 allow GSN elements to be generalised for future instantiation. The uninstantiated entity placeholder denotes that the attached element remains to be instantiated, i.e. at some later stage the abstract entity needs to be replaced with a more concrete instance. The undeveloped entity placeholder denotes that the attached element requires further development, i.e. at some later stage the entity needs to be decomposed and supported by further argument and evidence.


**Fig. 12.** GSN abstraction extensions

Figure 13 shows an example argument pattern for an asserted *inference* (e.g., ACP1 in Figure 5). This pattern demonstrates that there is sufficient confidence in the asserted inference by including a sub-argument:

- that the asserted inference is true
- that the assurance deficits relating to the asserted inference have been identified
- that any residual assurance deficits are acceptable.

The strategy used in the third sub-argument is to argue over the set of assurance deficits, and for each to show:

- the existence of significant counter evidence associated with the subject assurance deficit is considered unlikely
- the sensitivity of the remainder of the argument to the subject assurance deficit is acceptably low, i.e., the assurance deficit may be justified as acceptable when considered in the context of the other arguments and evidence in the safety case.

An example of how this pattern may be instantiated is included in section 4.

Figure 14 shows an example argument pattern for an asserted *solution* (e.g., ACP3 in Figure 10). The pattern demonstrates that there is sufficient confidence in the asserted solution by including a sub-argument that:

- the asserted solution is trustworthy
- use of the asserted solution is appropriate.

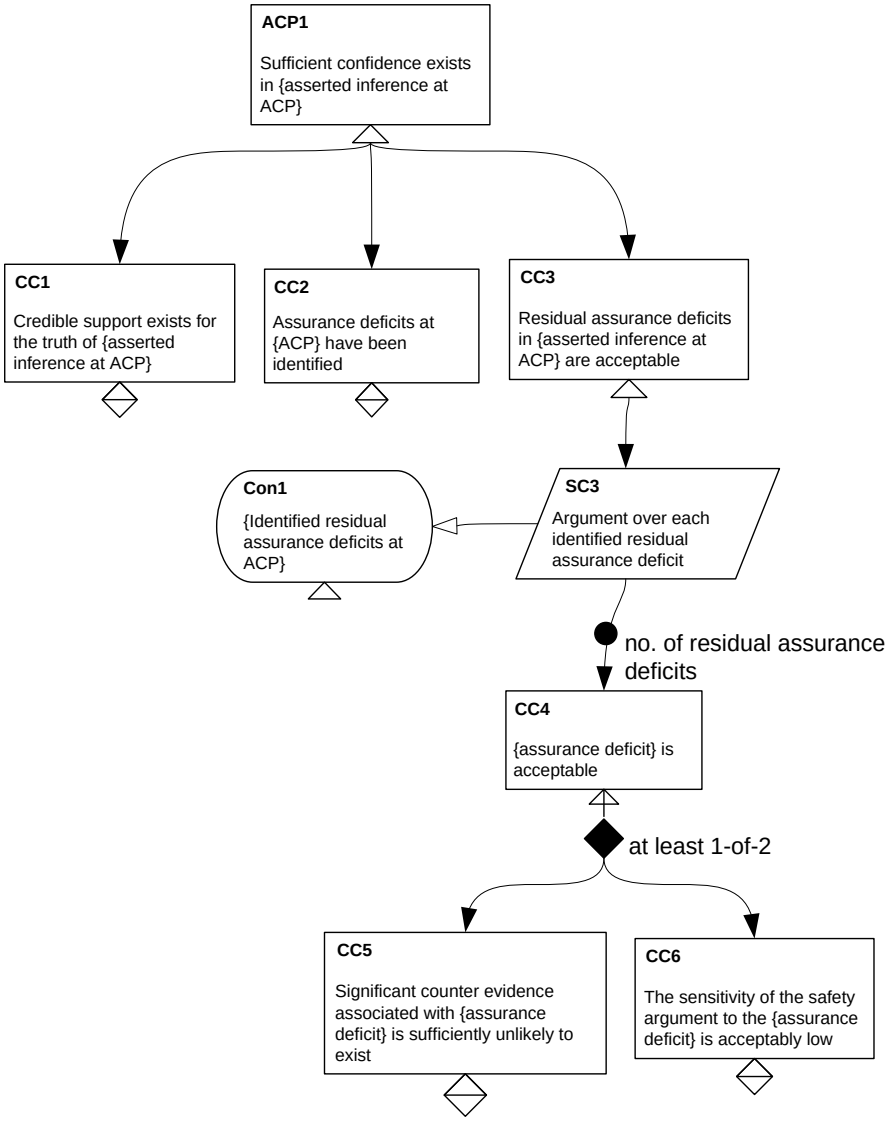


**Fig. 13.** Confidence argument structure for an asserted inference

Each of these sub-arguments has the same form as that used in Figure 13 and the same techniques for instantiation of the pattern could be used. The claims regarding the acceptability of the residual assurance deficits in each case (CC13 and CC23) would be supported using the same pattern as provided under CC3 in Figure 13. The distinction between these two sub-arguments is worthwhile since in general arguing the integrity of evidence is easier than arguing the appropriateness of the evidence. The explicit inclusion of both ensures attention is paid to both.

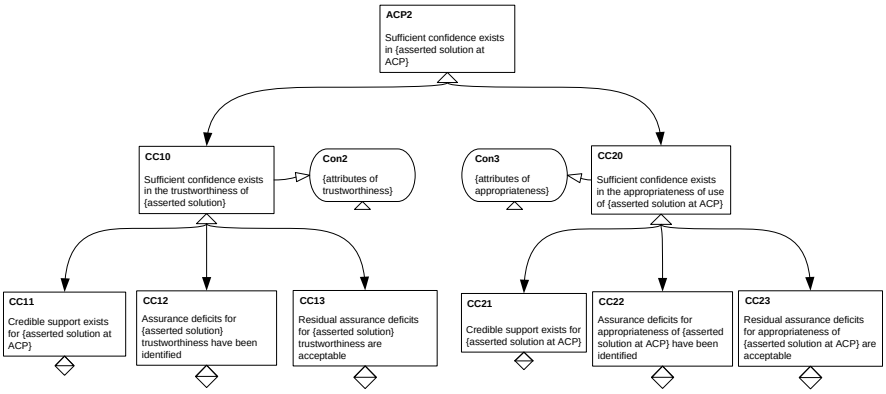


**Fig. 14.** Confidence argument structure for an asserted solution

## *3.5 The overall confidence argument*

The individual fragments of confidence argument, each addressing a particular assurance claim point in the safety argument, should be assembled together to form a single overall confidence argument (to accompany the single safety argument). To be truly comprehensive in the construction of this overall confidence argument would require that *all* of the assertions of the safety argument have an accompanying confidence (sub-)argument. This is illustrated in the three legs of the argument shown in Figure 15 (arguing confidence for *all* inferences, *all* context and *all* evidence used in the safety argument).

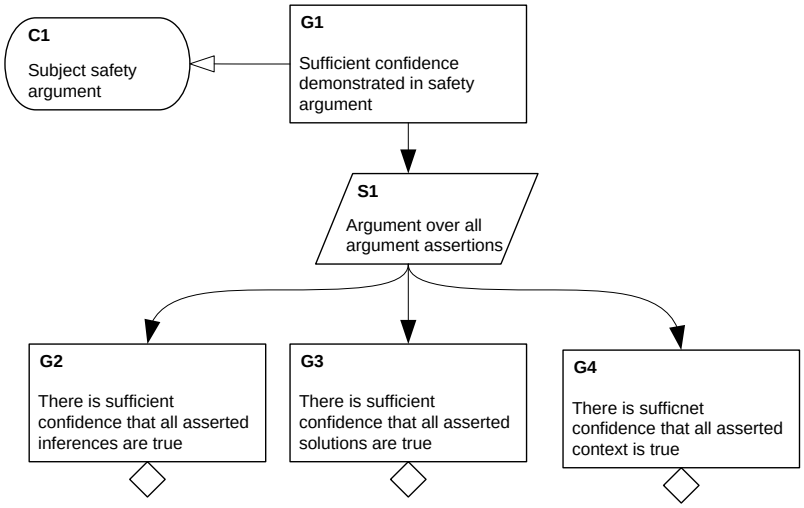


**Fig. 15.** Representing an overall confidence argument

In addition to this simple structure, there are a number of potentially important concerns at the level of the *overall* confidence argument. Firstly, arguing the sufficiency of the overall confidence in the safety argument can be more complex than the simple composition of arguments of sufficient confidence for each argument assertion (in the same way that arguing the acceptability of overall risk is more complex than simply arguing the acceptability of the risk posed by each individual hazard). For example, we have already highlighted in Section 3.4 that an assurance deficit for one argument assertion may be justified as acceptable when considered in the context of other arguments and evidence in the safety case. Such a justification of how shortfalls in one part of the safety argument are compensated by other arguments and evidence needs to be addressed at the level of the overall confidence argument. Secondly, it is useful to examine and justify whether the multiple lines of argument offered up in the safety argument (undesirably) share *common* underlying assurance deficits (i.e. there are common modes of failure in the argument). Thirdly, for large safety arguments it may simply not be practical to provide arguments of confidence for *every* assertion in the safety argument. Instead, some selection and prioritisation of the assertions of the safety arguments to be covered by the confidence argument may need to be performed. This prioritisation would be done most appropriately by addressing those assertions relating to the most significant arguments of risk reduction in the primary safety argument. Obviously, care must be taken when making any decisions regarding parts of the confidence argument to omit.

## 4 Example assured safety argument

To illustrate how an assured safety argument might be structured in practice, we show key aspects of an example argument created for a *hypothetical* insulin pump. Figure 16 shows the high-level structure of the safety argument. The claim that the insulin pump is adequately safe for routine use is supported by arguing over each of the identified credible hazards to which the patient might be subject.

To produce an assured safety argument, confidence argument fragments must be provided for each assurance claim point. In the example, the ACPs are:

**ACP.S1.** There is sufficient confidence that mitigating credible hazards will demonstrate that the insulin pump is adequately safe for routine use. Arguing over hazards is a widely accepted strategy in safety engineering, and this fragment of the confidence case is simple to construct.

**ACP.A1.** There is sufficient confidence that pump design is accurately documented. If the documented pump design does not faithfully represent the pump, then the argument presented may not be valid.

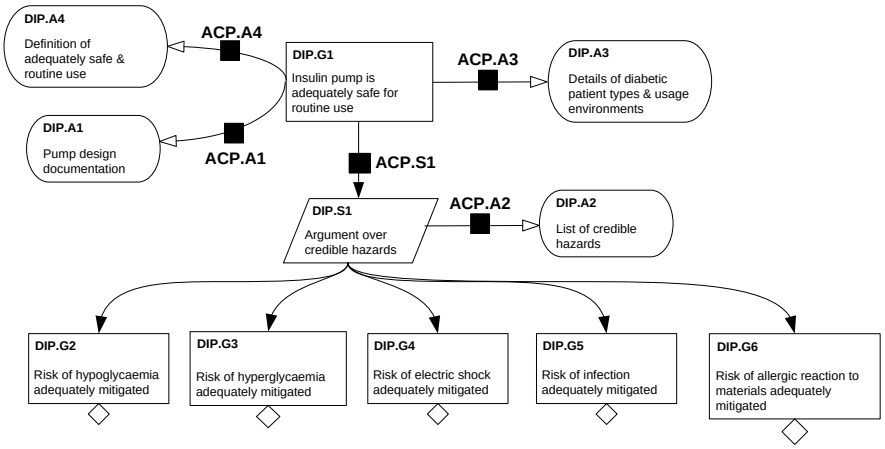


**Fig. 16.** High-level safety argument for an insulin pump

**ACP.A2.** There is sufficient confidence that the list of credible hazards is complete and correct. Inadequate definition of a hazard or omission of a hazard might invalidate the safety claim.

**ACP.A3.** There is sufficient confidence that the details of diabetic patient types and usage environments are accurately documented. Usage outside of the expected set of environments might invalidate the safety claim.

**ACP.A4.** There is sufficient confidence that the definitions of adequately safe and routine use are appropriate for the safety claim being made. If the scope defined by this context is not appropriate for the way in which the system is operated, for example if the device is used in an unplanned manner in a hospital, then the argument presented may not be valid.

We examine ACP.A1 in detail. A1 is a context, and to create a suitable confidence argument fragment we adapt the solution pattern shown in Figure 14. Figure 17 shows the sub-goals labelled CC1.3 and CC2.3 corresponding to sub-goals CC13 and CC23 from the pattern in Figure 14. The remainder of the pattern would be instantiated in a suitable way.

Subgoal CC1.3 states: 'Residual assurance deficits in the trustworthiness of the pump design document are acceptable.' The assurance deficits that we associate with the trustworthiness of the pump design document need to be enumerated and each included in the appropriate confidence argument fragment. In this example, we consider just two assurance deficits:

- the possible deficit introduced by the use of a commercial word processing tool (CC1.3.1), i.e., are we sufficiently confident that the document was not corrupted in some way by the word processor?

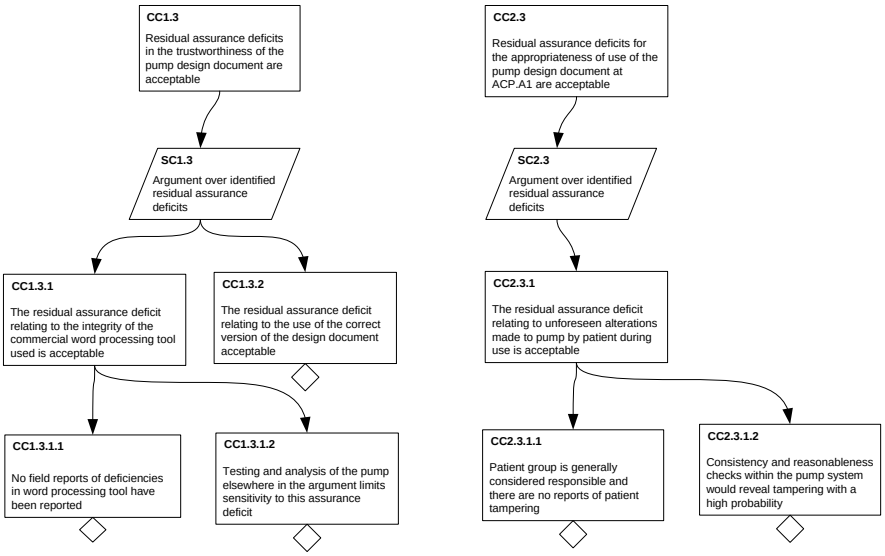


**Fig. 17.** Part of the confidence argument for ACP.A1

- the possible deficit introduced by the supply of the document for use (CC1.3.2), i.e., are we sufficiently confident that the correct document was actually referenced?

For both of these assurance deficits, we need to consider both counter evidence and sensitivity. In Figure 17, we show just a single claim for counter evidence and sensitivity and just for claim CC1.3.1. We argue a lack of counter evidence about the commercial word processor based on reported deficiencies, and we argue lack of sensitivity based on independent information about the design that will be generated by testing and analysis of the pump as built. Sensitivity is low because a defect in the document would be revealed from observations of the pump during testing and analysis.

A single assurance deficit for the appropriateness sub-argument is also shown in Figure 17 (claim CC2.3.1). The claim is: 'The residual assurance deficit relating to unforeseen alterations made to pump by patient during use is acceptable'. For this claim, the problem is that the documentation might be inappropriate because the pump has been locally modified. For lack of counter evidence in this claim, we cite the claim that there is no evidence that such tampering occurs. For sensitivity, we cite the claim that consistency and reasonableness checks by the pump during operation would reveal tampering with a high probability and would raise an alarm. Thus, the remainder of the safety argument is not especially sensitive to this possibility.

Part of the next level of the safety argument for the insulin pump, elaboration of goal DIP.G2, is shown in Figure 18. The strategy used in this elaboration is to

argue over the hazard of excess insulin in different delivery modes. Five assurance claim points are defined by this (incomplete) version of the elaboration:

**ACP.S2.** There is sufficient confidence that considering the risk of excess insulin during each possible delivery mode will demonstrate that the risk of hypoglycaemia is adequately mitigated.

**ACP.A5.** There is sufficient confidence that the list of delivery modes is complete and correct.

**ACP.S3.** There is sufficient confidence that arguing over patient commanded and uncommanded infusions will demonstrate that the risk of excess insulin during meal/correction bolus infusion is adequately mitigated. We might argue that 'commanded' AND 'uncommanded' is a tautology.

**ACP.A6.** There is sufficient confidence that the definition of commanded infusions is appropriate. Some modern insulin infusion pumps use a Bluetooth network connection to communicate. This definition of commanded infusions might be inappropriate if it does not make clear whether infusions resulting from security attacks over Bluetooth are commanded or not.

**ACP.A7.** There is sufficient confidence that the definition of uncommanded infusions is appropriate.

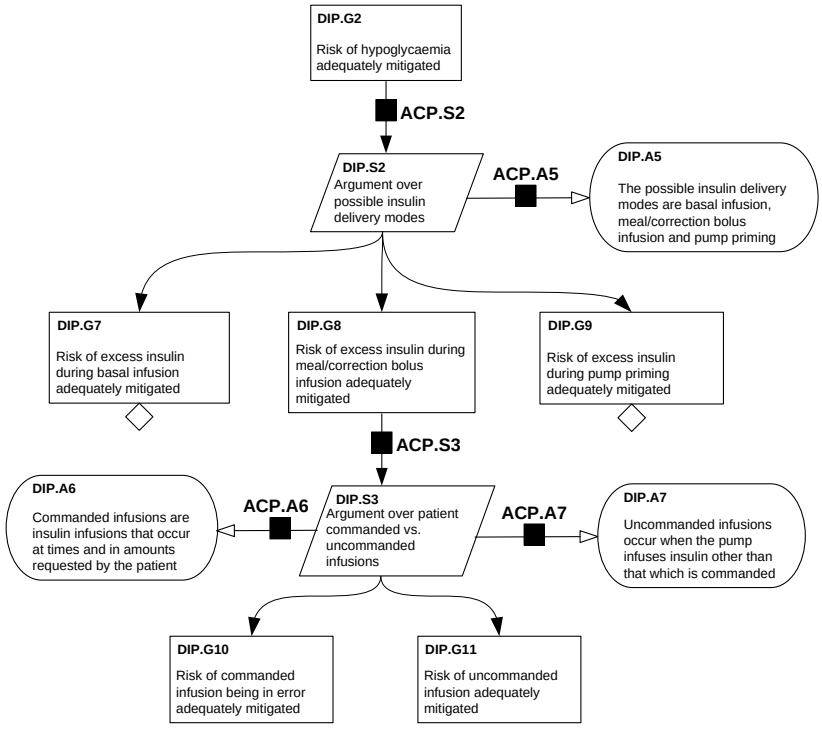


**Fig. 18.** Insulin pump safety argument elaboration of goal DIP.G2

## 5 Conclusions

It is currently commonplace for safety case authors to mix and confuse two types of argument within a single safety case argument structure – direct arguments of safety behaviour, and indirect 'confidence raising' arguments. There are a number of unfortunate consequences to this practice. Firstly, the confidence arguments are often weakly related to direct arguments of risk reduction. Secondly, the resultant arguments are often 'rambling', have poorly defined argument structure, and have unclearly defined stopping criteria. It is too easy to keep adding arguments and evidence ad infinitum, when the only entry criteria that seems to be being applied is, 'Does this have any possible bearing on the safety of the system?' Greater discipline is needed when deciding on how to structure the arguments of the safety case.

This paper introduces *assured safety arguments* as a mechanism to deal with this problem. This structure explicitly separates the safety case argument into two components – a *safety* argument and an accompanying *confidence* argument. The safety argument is allowed to talk only in terms of the causal chain of risk reduction, and is not allowed to contain general 'confidence raising' arguments. The confidence argument is constructed *relative* to this safety argument and clearly structured according to the assertions of the safety argument. Again, the confidence argument cannot be considered a 'free for all' and is not allowed to contain general 'confidence raising' arguments that cannot be clearly related to the structures of the core safety argument.

Of particular importance is the prospect of focusing the activities associated with certification on the two arguments in an assured safety case. Certification as defined by Defence Standard 00-56 (MoD 2007), for example, requires that a safety case provide:

> '... a compelling, comprehensible and valid case that a system is safe for a given application in a given environment'

The standard does not define 'compelling, comprehensible and valid', but intuition suggests that concern is with quality of the safety case. Using an assured safety case, officials charged with assessing a safety argument will have clear and distinct statements about the main properties of interest, the argument targeted at the primary safety claim and the argument targeted at the primary confidence claim.

We have limited our discussion in this paper to safety cases, but the concepts apply immediately to *any* property of interest. Thus, for example the notions of assured security cases or assured reliability cases are appropriate, and each would benefit from the explicit introduction of a confidence argument in the same way that a safety argument does. Naturally, the content of an assured security case would differ from the content of an assured safety case, but the overall structures and approaches would be identical.

Our preliminary experience of applying separation and developing explicit and separate confidence arguments has revealed that the approach yields the expected

benefits – greater clarity in (and consequently comprehension of) the arguments, and a reduction in size of the core safety argument.

**Acknowledgments** The authors would like to acknowledge the financial support of the Royal Academy of Engineering (through the Distinguished Visiting Fellowship Scheme) for the work reported in this paper.

**References**

Habli I, Kelly T (2007) Achieving integrated process and product safety arguments. Proceedings of 15th Safety Critical Systems Symposium.

Haddon-Cave C (2009) The Nimrod review. The Stationary Office. London

HSE (2001) Reducing risks, protecting people. Health and Safety Executive. HSE Books

Kelly T (1998) Arguing safety - a systematic approach to managing safety cases. PhD Thesis. Department of Computer Science, The University of York.

Kelly T, Weaver R (2004) The goal structuring notation - a safety argument notation. Proceedings of the Dependable Systems and Networks Workshop on Assurance Cases

MoD (2007) Defence Standard 00-56 Issue 4: safety management requirements for defence systems. HMSO