# Reasoning About Confidence and Uncertainty in Assurance Cases: A Survey[⋆]

Lian Duan[1], Sanjai Rayadurgam[1], Mats P.E. Heimdahl[1],
Anaheed Ayoub[2⋆⋆], Oleg Sokolsky[2], Insup Lee[2]

[1] University of Minnesota, Minneapolis, USA
{lduan,heimdahl,rsanjai}@cs.umn.edu
[2] University of Pennsylvania, Philadelphia, USA
aae.anaheed@gmail.com, {sokolsky,lee}@cis.upenn.edu

**Abstract.** Assurance cases are structured logical arguments supported by evidence that explain how systems, possibly software systems, satisfy desirable properties for safety, security or reliability. The confidence in both the logical reasoning and the underlying evidence is a factor that must be considered carefully when evaluating an assurance case; the developers must have confidence in their case before the system is delivered and the assurance case reviewer, such as a regulatory body, must have adequate confidence in the case before approving the system for use. A necessary aspect of gaining confidence in the assurance case is dealing with uncertainty, which may have several sources. Uncertainty, often impossible to eliminate, nevertheless undermines confidence and must therefore be sufficiently bounded. It can be broadly classified into two types, *aleatory* (statistical) and *epistemic* (systematic). This paper surveys how researchers have reasoned about uncertainty in assurance cases. We analyze existing literature to identify the type of uncertainty addressed and distinguish between qualitative and quantitative approaches for dealing with uncertainty.

## 1 Introduction

Systems developed for medical, transportation, and infrastructure applications that significantly impact life, property, or environment typically need to gain the approval of an independent entity such as a regulatory body. This certification or approval process can be viewed as the manufacturer making the case that the system meets the criteria for certification or approval, and the third-party then independently assessing the case to arrive at a decision. Assurance cases provide a structure for making this case—using arguments supported by evidence to justify a claim, typically in a hierarchical fashion. Generally, the top-level claim is one about dependability properties of the system such as safety, trustworthiness

[⋆] This work has been partially supported by NSF grants CNS-0931931 and CNS-1035715

[⋆⋆] Currently employed at Mathworks

or reliability. An assurance case supports a claim *"x,"* such as *"the system is sufficiently safe"* or *"the system's software conforms to its requirements."*

Demonstrating such claims to the satisfaction of all concerned can be quite difficult. In stable and well-established fields such as avionics and medical devices, a prescriptive approach is commonly followed. A regulatory agency sets forth standards that must be followed (processes used during the development phase, tests that the system must pass, and so on), and the manufacturers must provide evidence showing that they followed the prescription [24]. Evidence collection to demonstrate adherence to prescribed standards is mandated by the prescription. However, parts of the argument linking the evidence produced during development to the ultimate claim being made about the system dependability are implicit. This could pose challenges to independent third-party assessors who have to fill in the missing pieces of the argument linking the supplied evidence to the claims. Assurance cases have attracted considerable interest and also have been adopted in domains where system safety is of particular concern.

The medical devices field, in particular, presents extra sources of challenges that must be considered due to the existence and nature of patients, who interact directly with the systems or become part of the system themselves. These patients are unpredictable unknowns. Additionally, the existence of patients affects how medical device companies approach their design and certification processes.

Depending on the goal of the assurance case, how it is structured and argued varies. In this survey, we focus on the safety assurance case. According to Bloomfield, a safety assurance case is a *"documented body of evidence that provides a convincing and valid argument that a system is adequately safe for a given application in a given environment"* [7]. The vendor provides a claim of safety and the evidence to sufficiently substantiate that claim. The case is then evaluated by a regulatory agency that must decide if the system(s) with that software can be used in the market. The responsibility to make the assurance case and demonstrate the safety of the system(s) using the software rests on the vendor instead of on the regulatory agency.

Confidence in the reasoning as well as in the evidence must be considered carefully when evaluating an assurance case. Grigorova and Maibaum introduce a working definition for confidence in an assurance case as *"the quality or state of being certain that the assurance case is appropriately and effectively structured, and correct"* [13]. This working definition applies to both the assurance case developers, who must have confidence in their case before the system is delivered, and the assurance case reviewers, such as a regulatory agency, who must develop adequate confidence in the case before they approve a system for use.

A necessary aspect of confidence is uncertainty; more uncertainty reduces confidence. Bloomfield suggests that the best way to indicate confidence and uncertainty is through the use of probability [7]. Higher uncertainty reduces the probability of the confidence. Since uncertainty is inherent in the world, when making assurance cases, uncertainty must be addressed, either implicitly or explicitly. Engineers and researchers have necessarily had to figure out how to deal with uncertainty in their various approaches to assurance cases. A possible approach is to classify uncertainty into two types—aleatory (statistical) or epistemic (systematic). Aleatory uncertainty relates to “the intrinsic randomness of a phenomenon” [19]. These uncertainties are the “known unknowns” and are quantified by probability distributions. Examples would be the overinfusion in an infusion pump or human error when calculating drug amounts. Another source of aleatory uncertainty that easily could be forgotten is residual ones - i.e., after a hazard has been deemed mitigated, there could be a small chance that it actually was not. Epistemic uncertainty is “presumed as being caused by a lack of knowledge (or data),” or the “unknown unknowns” [19]. Examples would include faults in logical reasoning that the reasoner was not even aware of, or a sequence of inputs that had not been anticipated in the development and design process. A goal for uncertainty researchers is to reduce epistemic uncertainty to aleatory so that it can be modeled. Recently, there has been more focus on quantifying epistemic uncertainty through methods such as Dempster-Shafer theory or Bayesian analysis [27].

Researchers approach confidence and uncertainty in assurance cases usually through one of two ways—a qualitative analysis or a quantitative one. The qualitative view to dealing with aleatory uncertainty is to remove it, such as by narrowing the world-view to such a point that there is still sufficient confidence in the assurance case but there are no more unknowns. The qualitative view to dealing with epistemic uncertainty is to reason it away. The quantitative approach to uncertainty generally uses the fact that uncertainty reduces confidence—if one has 70% confidence in an argument, then one has 30% uncertainty.

This paper surveys how researchers have reasoned about confidence and uncertainty, first by organizing their work into whether they took a qualitative or quantitative approach, then by analyzing the sources of uncertainty present in their approaches and if those are aleatory or epistemic. Lastly, the paper concludes with a summary and a discussion of directions for future exploration; there has not been one approach that is fully adequate for the current state of assurance cases. As such, we believe that further work is needed to develop an uncertainty reasoning framework drawing on the strengths of current approaches.

## 2 Reasoning About Confidence and Uncertainty

Bertolino and Strigini [4] looked into the difference between the two extreme approaches to reasoning about software faults, “statistical” (quantitative) and “perfectionist” (qualitative). We follow the same premise for the rest of this section, but with a focus on how the research groups have reasoned about uncertainty. First, the work is separated into qualitative or quantitative approaches to reasoning about uncertainty (not qualitative or quantitative approaches to reasoning about assurance cases). Then, the subsection is further subdivided into three sections: (1) a summary of the research methods used, with a focus on how they reasoned about confidence; (2) a discussion of how the research groups have reasoned about uncertainty and into what category the uncertainties fall; and (3) a discussion of our views on these works.

### 2.1 Qualitative: Logical Argumentation

The structure of the argument plays an important role in the confidence in the assurance case. Researchers have focused on correct, logical argumentation structures to clearly convey how the claim, in a given context, can be inferred from the evidence provided. Uncertainty may then be dealt with by checking over the argumentation as well as by narrowing the context.

Tim Kelly presented the idea of using "argumentation structure" for assurance cases [17]. He used safety arguments to get the safety evidence needed to meet safety requirements. He developed the idea further into a formalized argumentation structure called the Goal Structured Notation (GSN)—a symbol-based language intended to help formulate assurance cases.

R.D. Hawkins et al. introduced the idea of the "assured safety argument" by separating an assurance case into two parts—the safety assurance case and the confidence case [14]. Instead of looking for ways to assess confidence implicit in an assurance case, they suggest making the argument for confidence explicit by constructing a second case—the confidence case—for the safety assurance case. The main idea is that the assurance case and the associated justification for it are two separate entities and should be treated as such. Since confidence is central to assessment, leaving it implicit or intermixed in a one-part assurance case can be a major source of confusion both for the party making the assurance case (how to justify the confidence) and the party reviewing it (how to factor in the confidence). The assurance case developers mark the locations on the "assurance argument" where more justification is needed, and these are then addressed in the "confidence argument." The direct and specific connection points between the two cases help ensure that the confidence argument contains no extraneous information beyond what is necessary to strengthen the assurance argument.

**Uncertainty Reasoning.** Kelly and Rob Weaver view uncertainty in a qualitative way [18, 29]. For epistemic uncertainty, if the argument made in the assurance claim is *sufficiently* well-structured, there should be no unknowns. If potential sources of epistemic uncertainty still exist, then one needs to go back into the assurance case and figure out which part needs to be better structured to remove uncertainty. Options to deal with aleatory uncertainty include resolving it, such as determining a way to eliminate it, or arguing that it does not impact the overall claim (at least not enough to change its credibility).

Kelly provides a "step-by-step" guide for assurance case reviewers where he points out inconsistencies for which reviewers should look, biases that might be inadvertently (or purposefully) inserted, and sources of weak or incorrect reasoning [18]. While not expressly stated as confidence or uncertainty, these two qualities exist innately for a reviewer when reviewing an assurance case. A developer can also introduce sources of epistemic uncertainty, thus unknowingly influencing evidence or arguments. Weak reasoning and biases, if caught by the reviewer, can reduce the confidence he or she may have in the case.

Hawkins also holds the view that the scope of the assurance case should be narrowed to the point that no aleatory uncertainty exists [15]. He defines safety assurance as *"[a] qualitative statement expressing the degree of confidence that a safety claim is true"* [15]. If the argument is believable or probable, and all of the uncertainties (whether aleatory or epistemic) are known (but cannot be neutralized for one reason or another), then there is still high confidence in the argument and one does not have to worry about the uncertainties. This "sufficient confidence" is established in the confidence argument part of the "assured safety argument" [14]. The confidence argument has multiple purposes—it has to establish sufficient overall confidence in the assurance case, it has to justify the corresponding parts of the assurance argument, and it has to address the uncertainties that exist in the assurance argument.

**Discussion.** If everything has been considered that possibly could be of interest in the "world," then, as Kelly, Weaver, Hawkins et al. suggest, there would be no uncertainty. Nevertheless, this is not realistic. Epistemic uncertainty always exists, especially when interacting with the real world. It is not feasible to account for every possible eventuality, simply because there will always be something that cannot be explained or anticipated. It makes good sense, however, to try: a reason why engineers must talk to domain experts to understand the world when building new products or improving old ones.

Hawkins' narrowing of the world view to eliminate aleatory uncertainty can be understood in a similar fashion. If one is building a new insulin pump, it would be reasonable to consider non-diabetic persons as outside the scope of the world, so the issue of, "What if a non-diabetic person used this machine?" would be eliminated with the solution, "That is outside the scope of the system under consideration". The risk here is that the narrowing of the scope may exclude situations or usage scenarios that may in actuality occur. For example, the extensive use of medical devices and pharmaceutics "off label" (used in situations for which approval has not been granted) could be excluded from an assurance case's scope to focus on a small and well understood patient population. This may, however, introduce uncertainty (both aleatory and epistemic) with respect to the reasonableness of this narrowing.

While the idea of separating the confidence case from the assurance argument achieves an important separation of concerns, since the assurance case itself has the propensity to grow rapidly and become increasingly complex (to create, check, and evaluate), the confidence argument is likely to be just as complex as the original assurance case. We suspect that, especially for safety-critical systems, dealing with uncertainties using qualitative approaches entails some inherent difficulties that cannot be completely eliminated by better assurance case structuring alone.

Unlike quantitative approaches, which attempt to represent confidence as a numeric value that might hide the nuances behind that number, the qualitative approach focuses on the reasoning and rationale behind any confidence value that could (some say arbitrarily) be placed in a confidence argument. Hawkins et al. point out that to arrive at a numerical value for assurance case confidence, one needs to go through the process of logical reasoning that is being used in their argument structures [14]. Therefore, one could view the quantitative approaches as requiring a foundation of qualitative analyses, without which numbers may be incorrectly adjusted to fit the end goal.

### 2.2 Qualitative: Baconian Probability

John Goodenough et al. approach confidence in assurance cases through *eliminative induction*—increasing confidence by removing sources of doubt and using Baconian probability to represent confidence [11]. First, they identify sources of doubt, called "defeaters"—that is, anything that could bring down confidence in the assurance case, and then work towards removing each source of doubt or proving that it is not relevant, ("eliminating" them). As more sources of doubt are eliminated, the confidence in the claim grows. A potential source of confusion with Baconian probability is that it must not be treated like actual probability. For instance, if there are 12 sources of doubt and 8 are eliminated, then $\frac{8}{12}$ sources of doubt are eliminated is the only valid conclusion, and we have $\frac{8}{12}$ confidence. This is to be considered as qualitatively different from $\frac{2}{3}$ as a confidence measure.

**Uncertainty Reasoning.** In this approach, sources of doubt are similar to sources of uncertainty. There is no specific separation between aleatory and epistemic—both can be sources of doubt. As doubts are identified and eliminated uncertainty decreases. Using the previous example, if 12 sources of doubt have been identified, and 8 sources have been eliminated, we still have $\frac{4}{12}$ (not the same as $\frac{1}{3}$) uncertainty in the assurance case. However, there is no way to remove all sources of doubt. They acknowledge that uncertainty will never be fully eliminated as it is not possible to find out all sources of doubt, recognizing, without naming, the existence of epistemic uncertainty.

**Discussion.** The use of eliminative induction and Baconian probability to deal with uncertainty in assurance cases is unique to the approach of Goodenough et al. Confidence is not the absence of doubt; it is showing the lack of basis for its presence. Thus, it depends on identifying the sources of doubts, which can be problematic for epistemic uncertainty. Further, eliminating a large number of identified doubts does not (or should not) necessarily proportionally increase confidence—presence of a large number of identified doubts may lead one to question if there are more—unidentified—ones. If simply eliminating more defeaters increases confidence, then one could add a large number of defeaters and eliminate them to artificially increase confidence. The authors acknowledge this issue, but argue that in reality there are only a finite number of those that are of consequence. In a way, they follow the world/scope-narrowing approach followed by the previous qualitative researchers.

Confidence values are accumulated over an assurance case by simply summing up the Baconian probabilities from the leaf nodes of the assurance argument structure. As an example, if we had a claim supported by three pieces of evidence, with confidence in the evidences at $\frac{9}{12}$, $\frac{2}{3}$, and $\frac{1}{2}$, then the overall confidence for the claim would be $\frac{12}{17}$. This approach is rather straightforward. However, this treats all defeaters as equally important, which may not be appropriate; as the authors acknowledge, some defeaters may need to have more weight than others in determining confidence.

An advantage of the Baconian approach over Pascalian approaches (Section 2.3) can be illustrated by an example provided by Goodenough et al. [11]. If some claim is based on four independent sub-claims, each with an associated confidence of 0.999, their (Pascalian) combination would result in a confidence of $0.999^4 = 0.996$ for the claim. If some evidence now establishes 3 out of those 4 sub-claims to be true (1 confidence), then the overall confidence increases to $1^3 \times 0.999 = 0.999$, not much of a growth. However, using Baconian probabilities, starting at $\frac{0}{4}$ overall confidence ("invalid sub-claim" as defeaters), the new evidence raises it to $\frac{3}{4}$, clearly highlighting the significance of the evidence.

### 2.3 Quantitative: Pascalian Probability

Assurance and confidence appear to be linked in the view of Robin Bloomfield et al. [7]. When reviewing assurance cases, there is always a bias, affected by how much the reviewer trusts what he or she is reviewing. While the qualitative researchers sought to lessen this bias by well structured arguments, Bloomfield et al. preferred to use probabilities. Since there is never certainty in the world, probabilities are the best way to show this uncertainty. Their work has brought forth the idea that multi-legged arguments can support each other and give higher confidence to a claim, and the idea that when one has a high degree of confidence that a system satisfies a high level safety assurance, one has an even higher degree of confidence that the same system satisfies a lower level of safety assurance. Bloomfield et al. also developed their own argumentation structure, called Claims, Arguments, and Evidence notation (CAE), to reason about assurance cases [6].

Bev Littlewood and David Wright believe that probability, especially Bayesian probability, is the best way to address confidence [22]. They use an idealized, reduced example to make a formal analysis of confidence using Bayesian Belief Networks. Specifically, they looked at whether two *diverse* legs of an argument would help increase the confidence in the assurance case and found the answer to be *yes*, most of the time.

Xingyu Zhao et al. categorically state that they believe the quantitative approach is better and aligned more closely with how humans think and reason [31]. To this end, they developed a framework to assess confidence in assurance cases. Starting with a structured argument meta-model (that supports GSN [17], CAE [6], and TRUST-IT [9]), they map the meta-model onto Toulmin's argumentation theory model [28]. This is converted to a Bayesian Belief Network (BBN) via four branches of reasoning – justified premises, adequate information, justified applicable warrant, and justified assumption that no exceptions apply. In the BBN, the leaf nodes are then set to prior probability values and the non-leaf nodes are assigned a conditional probability table whose numbers are from "field related statistical data" and "expert judgments" [31].

To John Rushby, an assurance case is composed of two components: communication and reasoning [25]. He argues that given enough parameters, it is possible to strive for the "possibly perfect" piece of software [24]. One can look towards a value for the "probability of perfection," which can then be related to confidence. He later argues, similar to Hawkins et al. [15], that if one narrows the scope enough to what is relevant, then perfection is possible as all uncertainties would have been eliminated. He approaches assurance cases and confidence recursively, building up confidence from the leaf nodes ( *"substantiated claims about a subsystem can be used as evidence in a parent case"*).

**Uncertainty Reasoning.** Bloomfield et al. say that we use probability *because* there is uncertainty. In their view, uncertainty surely exists in the environment, and it is best shown by using probabilities for confidence values [7]. Here, uncertainty is the complement of confidence in an assurance case—e.g., 0.8 confidence means 0.2 uncertainty.

Peter Bishop et al. address the issue of epistemic uncertainty in reliability cases as it relates specifically to the probability of failure on demand [5]. Uncertainty always exists, especially epistemic uncertainty, and it needs to be dealt with when developing assurance cases. Their approach is to suggest the use of Bayesian Networks, relating back to Bloomfield et al.'s initial suggestion of probability.

Littlewood and Wright's use of Bayesian Belief Networks indicate their recognition of epistemic uncertainty [22]. To them, confidence's complement is doubt, which they model with probabilities in a BBN. Their results from looking more into multi-legged arguments for assurance cases show that depending on the nature of the extra "legs" in the argument, these legs can increase or reduce confidence in the original claim, a fact which is not wholly intuitive.

The use of Bayesian Belief Networks by Zhao et al. point to a desire to quantify the epistemic uncertainty that exists in assurance cases [31]. The use of prior probability values is the result of reducing epistemic uncertainty to aleatory uncertainty for the leaf nodes, to be propagated back through the BBN.

Rushby asserts that there will always be unknowns [24]. He believes that we should formalize the assurance case review process as much as possible to reduce the amount of information that needs "human review." There is a notional perfect system, which is impossible to achieve, and so one should build towards that perfect system. Software is reliable and has predictable behavior, but its environment, riddled with epistemic doubt, is not [25]. He maintains that reasoning is a logical process which can be mechanized, while communication, involving human factors, is epistemic in nature. In Rushby's world, if the software never fails, we have reached perfection—a probability of one. He argues that the more verification and validation that has been done on a piece of software, the greater the "probability of being perfect." He views verification as part of the logical reasoning, which is reducible and can hopefully be formalized, and validation as dealing with epistemic doubt. The goal then is to reduce the epistemic doubt to the point where logical analysis can take over. Then, only the leaf nodes have epistemic uncertainty and once these are addressed the rest of the model can be formally analyzed.

**Discussion.** Bloomfield et al., along with Littlewood and Wright, say that multi-legged arguments can provide more confidence for a claim. Littlewood and Wright formally show this with a reduced example; they also show that confidence may be reduced by a seemingly supportive second leg argument. This issue is looked into further by in Section 2.6 – Ayoub et al. explore ways that evidence can combine to support or detract from the claim [2].

While probabilities help quantify confidence, using a single number ignores the subtle nuances in reasoning about uncertainty. The use of probability is a logical approach to quantifying confidence and uncertainty, but might be too coarse of an approach. More recent works address this issue specifically when addressing epistemic uncertainty, such as the one by Bloomfield et al. [5].

Zhao et al. develop a framework and they show several interesting results, but the specifics of the process employed are not clear from their work [31]. They provide simple guidelines and advocate using "common sense" to quantify uncertainty, but details seem to be lacking.

Rushby directly addresses the idea of "epistemic doubt", but does not view it in the same way we have approached it here [25]. In contrast to our categorization of uncertainty into epistemic and aleatory, he focuses on logic and doubt. As such, aleatory uncertainty is not specifically mentioned. He views the reasoning behind the connections between arguments as where there is no uncertainty (it is logic), while we argue that this is perhaps where much of the epistemic uncertainty exists.

### 2.4 Quantitative: The Confidence Toolkit

Lukasz Cyra and Janusz Gо́rski also developed their own argument structure and notation, similar to Kelly's GSN [17] or Bloomfield's CAE [6], called Trust-IT [9]. Where they differ from Kelly is the view that a sufficiently detailed, complete assurance case requires so much information that these argument structures become huge, unwieldy and difficult to navigate. The size makes reviewing the assurance case especially difficult. Their solution is to look at the language used in "expert assessments" and quantify it.

First, they generate a confidence versus decision plot. One axis is the "decision scale," which has four possible values representing decisions that could be made by the reviewer—rejectable, opposable (a soft reject), tolerable (a soft accept), or acceptable. The other axis is the "confidence scale," which has six possible values ranging from absolute "lack of confidence" to total "for sure" confidence. This is all plotted on a single two dimensional graph. When one has high confidence and acceptability, that is for sure a go (accept). When one has high confidence and opposability, then that indicates a stop (reject).

They next map this rectangular "assessment scale" onto a decision triangle – Josang's opinion triangle [16]. The "accept" and "reject" corners of the rectangle map onto the base of the triangle, and all decision points that correspond to "lack of confidence" map to a single point of uncertainty at the apex of the triangle. They conclude that when one has high confidence and an acceptable decision, there is a strong belief in the acceptability of the assurance case. Likewise, when there is high confidence and a rejectable decision, there is a strong disbelief in the acceptability of the assurance case. But when there is high uncertainty, no decision can be made. They effectively move from a two choice scale (accept or reject) to a three-choice scale (accept, reject, wait). The NOR-STA tool created by them provides a visual breakdown of a trust case, a decision, and the confidence/uncertainty in that decision and presents the decision scale on a user-friendly graphical interface. It provides a summary from the aggregation of expert assessors' opinions, along with a value for the uncertainty resulting from the aggregation. The intent is present all information to the final decision maker.

**Uncertainty Reasoning.** Cyra and Górski view uncertainty as lack of confidence and equate high levels of uncertainty with inability to make a decision [9]. A strong lack of confidence represents extreme uncertainty. Something that has been assessed as "with very low confidence opposable" is weak on multiple points. It is not a strong reject or accept on the decision scale, but it also has a fairly high level of uncertainty, casting doubt onto any decision that could be made. By mapping this graph onto Josang's triangle, the authors highlight the importance of uncertainty. Everything is in doubt when uncertainty dominates as it clouds the decision making process. They present a third option beyond "accept" or "reject", namely "wait" – wait for more information, until the uncertainty is removed and the confidence increased.

In the NOR-STA tool, the confidence in the assurance case is presented to the reviewer on a slider scale, while the uncertainty is shown on Josang's Triangle, so that with one quick glance, the reviewer can see clearly the confidence and the uncertainty associated with that confidence. The uncertainty represented is epistemic in nature, as it deals with the communication from the reviewers. Aleatory uncertainty may be taken into consideration by the reviewers in formulating their assessment but is not explicitly dealt with in the NOR-STA tool.

**Discussion.** Like Bloomfield et al., Cyra and Górski appear to view uncertainty as directly related to confidence. Additionally, they look at trust, represented by the belief or disbelief in the claim, which is wholly separate from confidence and uncertainty. Their use of the opinion triangle shows their approach to reasoning about uncertainty. It follows that when one has a lack of confidence, there is uncertainty. Therefore, one cannot make a decision when there is high uncertainty. Whereas one might look at Bloomfield et al.'s view as high uncertainty implies low confidence and thus a rejection, Cyra's approach is that a firm reject only applies when one has high confidence in the rejection. What they do not talk about is what to do when there are high levels of uncertainty. We believe it means that more information is needed, and no sound decision can be made until that uncertainty has been reduced or eliminated.

Their work also seeks to quantify language, which has a lot of subjectivity ("tolerable" versus "acceptable," for instance). Very few others, such as Lorenzo Stringini [26], have attempted to do something similar. However, human communication typically involves such subjective terms from which an objective decision must be derived. It is therefore useful to explore this deeper and the area holds potential for further study.

### 2.5 Quantitative: Multi-Component

Current popular assurance case tools such as GSN, CAE, or NOR-STA can very quickly and very easily balloon to such an extent that they are no longer beneficial to or navigable by the reviewer. In fact, too much information can obscure the argument or even be used to hide flaws. One quantitative approach by researchers is through the use of multi-component assurance cases. By separating the assurance case into multiple parts, the hope is that overall complexity is reduced while clarity of the arguments is increased.

Ewen Denney et al. continue Hawkins et al.'s work on a separate confidence case, but with a quantitative approach through the use of Bayesian Belief Networks [10]. This work represents an acknowledgement that numbers, when used prudently, can be helpful in reasoning about assurance cases. Their basic premise starts from a safety argument created using GSN. Then, following Hawkins et al., they create a confidence argument. Lastly, the nodes are quantified.

John Knight also embraced the idea of partitioning assurance cases, specifically, by separating the arguments into three different types—design, confidence, and operational [20]. Design arguments are used for a "desired safety property," confidence arguments deal with acceptability or believability of the components of assurance case, while operational arguments argue that assumptions made in the design argument will be true in an operational context.

Marc Bender et al. address confidence as one step in a multi-component software certification process [3]. They have split the process of certification into different categories of required information: evidence, confidence, determination, and certification. Similar to Knight, they address separation of concerns - when talking about evidence, and bringing about pieces of it as part of certification, one should only be concerned with the evidence, and *not* such qualities as trustworthiness, confidence, etc. When dealing with confidence, one only concerns oneself with aspects that relate directly to that. Evaluating evidence increases confidence, and this stops when one has reached the necessary satisfaction level. In their work, assurance cases are used as a representation of confidence. Just like how certification was broken down into four elements, confidence is also broken down into three necessary components: veracity about sources (trustworthiness), validity (soundness), and adequacy (sufficiency of knowledge).

Grigorova and Maibaum explore the analogy between confidence in assurance cases and the weight of evidence in law [13]. Looking at current assurance case confidence and their flaws in reasoning, the authors want to establish an exhaustive and complete compendium of relevant knowledge, one "living document" for each domain so that both developers and reviewers can see a baseline of what must be included. They start with a working definition of confidence that was introduced at the beginning of this paper, and stress that their goal is to achieve safety, not to satisfy a bunch of check boxes (focus on the process, not the end product), hence the suggestion for the compendium of evidence. Essentially, they are expanding upon Bender et al.'s multi-component approach by separating the *weight* of the evidence from other factors that may influence confidence.

**Uncertainty Reasoning.** Denney et al.'s use of BBNs show their recognition of the existence of epistemic uncertainty in their confidence arguments, as BBNs are generally considered the best way to treat epistemic uncertainty [27].

Bender et al. are of the viewpoint that uncertainty leads directly to a decrease in confidence [3]. They quantify the aleatory uncertainty by using random variables (usually normally distributed) with the means and variances affected by prior beliefs or parent nodes. Epistemic uncertainty is dealt with when a joint distribution is then computed on these random variables, and everything is tied together via a Bayesian Belief Network. They use the BBN to calculate confidence in the safety argument.

Greenwell et al. looked at fallacies made in system safety arguments [12]. As stated earlier, we believe that flaws in reasoning are sources of epistemic uncertainty, as there are multiple causes of such reasoning errors and, as demonstrated by Greenwell et al., such errors are difficult even for experts to correctly recognize and identify.

Rodes et al. equate confidence to belief in their work [23]. They create a generic framework for measuring a property (they use security but state that it can also be used for other areas, such as safety) via confidence. This confidence is directly related to belief, and can be reduced by doubt, which is caused by uncertainty. They discuss sources of doubt in assurance cases that include incorrect inferences, faulty evidence, or inaccurate goals. These are all sources of epistemic uncertainty.

Grigorova and Maibaum view uncertainty as the opposite of confidence [13]. They understand that it has a role in reducing confidence, and the use of weighted evidence would also impact uncertainty in different ways. By seeking to create a complete compendium of evidence, they attempt to turn epistemic uncertainty to aleatory, and ultimately eliminate some uncertainties.

**Discussion.** One issue with multi-component assurance cases is that while seeking to avoid the complications associated with complex and huge assurance cases, we possibly end up overloading reviewers with even more information, just presented differently. Instead of one complicated assurance case, the fear is that we now have three or four equally complicated assurance cases.

Denney et al. address a weakness in their approach when talking about evidence weights [10]. Like Goodenough et al. [11], the authors do not provide for a solution if some components of the argument are more important than others. Their assumption is the weights are equal. As seen earlier, Grigorova and Maibaum also looked into weighted evidence, as will Ayoub et al. [1]. Cyra and

Górski's NOR-STA tool allows assessors to assign weights to premises of certain warrants [9].

An issue in Grigorova and Maibuam's work is that they do not seem to be addressing directly who should bear this burden of considering evidence weights [13]. The authors do go back and forth a bit on this issue. Should it be the assurance case developer who makes sure all the evidence is accounted for or the reviewer who has to check that the evidence has all been included?

### 2.6 Quantitative: Confirmation Bias and Weighted Evidence

Anaheed Ayoub et al. start with the idea of confirmation bias, as noted by Leveson [21], which almost always exists when dealing with human judgment [1]. The basic premise is that we all have our preconceived notions, and these preconceptions influence our decisions and how we think in subtle, sometimes unrecognized ways. We are more likely to believe something if it aligns with what we have previously believed. So people who are structuring the assurance case might (possibly unintentionally) bias the evidence they use towards the claim and minimize evidence that might weaken it. Leveson further suggests that one way to combat confirmation bias when arguing that a system is safe is to use a counter-argument [21]. An example would be, instead of trying to argue that a system is safe, have the manufacturer argue that the system is unsafe. Then they are forced to consider evidence that previously would have been rejected.

Ayoub et al. take this idea in a different way, by focusing on the reviewers and having them argue for the sufficiency and the insufficiency of an argument [1]. The question they ask is, *"are the premises of the argument 'strong enough' to support the conclusions being drawn?"* The authors ask reviewers to assess a claim's sufficiency and insufficiency, forcing them to look at the why of what they think. Then, the difference from one of these two arguments is viewed as the uncertainty. For instance, given a claim "the device is safe to use," a reviewer evaluates its sufficiency at 80%, or 0.8. For the same claim, the reviewer is asked to evaluate its insufficiency (or, thought of differently, the reviewer is asked to evaluate the claim "the device is NOT safe to use"). He or she evaluates this at 10%, or 0.1. This means there is another 10% of the reviewer's opinion that is not accounted for - this is the uncertainty. By asking the reviewer to evaluate the insufficiency of the claim, Ayoub et al. seek to avoid confirmation bias.

Ayoub et al. also sought to create "a systematic approach to justifying confidence" in safety arguments [2]. Just as Weaver et al. sought to create a systematic approach to evaluating safety arguments [29], Ayoub et al. now apply similar ideas to the confidence arguments as introduced by Hawkins et al. [14]. The authors encourage a prescriptive approach (using what they call a "common characteristics map") to identifying the system hazards and deficits in the assurance case, mitigating these defects, then putting it all together in the confidence case. This all starts with the idea of masses and weights.

**Uncertainty Reasoning.** Ayoub et al. use masses and weights as representations of confidence in [2]. Higher confidence means a higher mass, implying a better or stronger argument. Arguments and evidences are weighted depending on their strengths. As part of their approach, they separated the evidence that the reviewers are examining into four types: alternate, disjoint, overlap, and containment, based partially on Dempster-Shafer theory. When the mass is zero, full uncertainty is implied. Such a missing mass, when used in a formula to calculate the overall sufficiency of an argument, can affect the total mass (and thus the total confidence), depending on the weight of the missing mass. The use of Dempster-Shafer theory implies that the authors acknowledge the existence of epistemic uncertainty. Aleatory uncertainties affect the mass of the evidence.

**Discussion.** This intriguing approach needs further development and refinement. It is counter-intuitive for some people to think about arguing for insufficiency, and it is not easy for most people to argue against themselves. It is even more counter-intuitive to think about all of the confidences adding up to one. One wonders if it is possible for a reviewer to give his or her sufficiency an 80% confidence and his or her insufficiency a 50% confidence. Where would the uncertainty be then? We argue that this should be allowed, and the excess 30% confidence over 100% would represent the uncertainty. There are a lot of limitations that the authors themselves admit.

While previous researchers mentioned the importance of weighting evidence, Ayoub et al. are the only group to actually consider it thus far, and they provide a logical solution. They also look further into how different pieces of evidence can combine to either increase or reduce confidence, in a similar fashion to how other researchers looked at multi-legged arguments.

## 3 Conclusions

We have briefly surveyed recent work on safety assurance cases and their potential usefulness in safety critical systems. A key issue that plagues assurance cases is uncertainty, which necessarily exists in the real world. Uncertainly can be categorized as aleatory or epistemic. Researchers have used a variety of methods to handle uncertainty inherent in assurance arguments supporting system dependability claims. We have attempted to categorize the uncertainty addressed by various methods, which is somewhat complicated by additional factors – the model used and the situation in which the method is applied play a role.

Qualitative approaches usually try to narrow the scope of the assurance case world to such a point that all uncertainties have been either eliminated or shown to be inconsequential. Quantitative approaches acknowledge that uncertainty always exists and must be dealt with, usually through its impact on confidence values. Since aleatory and epistemic uncertainty vary in how they are treated quantitatively, a clear distinction is crucial.

In our view, there has not been one approach that is fully adequate for the current state of assurance cases, though many approaches have novel ideas that address important considerations. It is our belief that to comprehensively handle uncertainty, one may need to employ a combination of approaches, perhaps including new techniques that have yet to be developed. While we do not yet have such a solution, we think it must have some essential ingredients.

Weighting evidence enables a more accurate reflection of the real world, and as such, is a minimum necessity in assurance cases. If we can get to a point in an assurance case where the leaf nodes only contain aleatory uncertainty, and epistemic uncertainty only exists in the other nodes, then perhaps it would be feasible to apply formal reasoning methods to the argumentation between the leaf and non-leaf nodes.

We believe that probability should be used to quantify uncertainty, especially if any sort of mechanization process is desired. How weighted evidence ties into Bayesian Belief Networks is an area ripe for exploring.

Another issue that must be addressed is the lack of consistency in the terminology and working definitions used by different researchers, which makes comparison techniques and combining approaches challenging. While this is to be expected in a new area, we believe the field is now maturing to the point where a common foundation would better serve further study.

A third issue is how little research has focused on the human factors, with some giving up before even starting, and others trying to minimize this as much as possible so as to avoid dealing with it. As this is an aspect that is unavoidable, it is worth a deeper look. A fourth issue is that uncertainty is a vast research area in its own right. Before dwelling deeply into reasoning about uncertainty in assurance cases, it would serve well to understand the foundational principles in reasoning about uncertainty in general. Lastly, the importance of proper assurance case technique must be underscored. It is far too easy for assurance case creators to go overboard and put too much and unnecessary information in an assurance case, potentially losing the argument that is being made or obscuring shortcomings that might exist in the evidence. Assurance cases should have clear, convincing, and preferably concise arguments.

While we have presented a survey of approaches to reasoning about confidence and uncertainty in assurance cases, an obvious question would be which method is superior? To answer this question requires a comparative evaluation between techniques. While we feel that this is beyond the scope of the original intentions of this work, we would like to present some thoughts on this issue.

Context is of utmost importance when choosing the best method to analyze confidence and uncertainty in assurance cases. The source of the uncertainty will also have an impact on the method used to address it. As we have shown, there are a variety of methods to convey, analyze, and evaluate confidence and uncertainty. At times, one method might prove to be more informative than others, while this same method might be insufficient in other situations.

For infusion pumps in the United States which expect to undergo FDA regulation, which require assurance cases as part of the approval process, exact confidence and uncertainty values might be less important as the overall reasoning [8]. As such, a qualitative approach might be reasonable. The purpose of the assurance case is for the manufacturer to show that they have reasoned about their device, similar to the logical reasoning as espoused by Kelly, Hawkins, Weaver, et al.

In the UK, the Healthy and Safety at Work Act uses the phrase "so far as is reasonably practicable" - which is a qualitative goal [30]. As such, while it is possible to use a quantitative analysis and then map it onto a qualitative scale, one could reasonably use logical argumentation to show that the goal has been achieved without the use of any specific quantifiers.

On the other hand, a manufacturer that is attempting to show that their system is an improvement over another might benefit more from a quantitative approach. If more confidence or less uncertainty was demonstrated, quantitatively,

## References

1. Anaheed Ayoub, Jian Chang, Oleg Sokolsky, and Insup Lee. Assessing the overall sufficiency of safety arguments. In *Safety-Critical Systems Club*, 2013.
2. Anaheed Ayoub, BaekGyu Kim, Insup Lee, and Oleg Sokolsky. A systematic approach to justifying sufficient confidence in software safety arguments. In *SAFECOMP*, 2012.
3. Marc Bender, Tom Maibaum, Mark Lawford, and Alan Wassyng. Positioning verification in the context of software/system certification. In *Proceedings of the 11th International Workshop on Automated Verification of Critical Systems*, 2011.
4. A. Bertolino and L. Strigini. Assessing the risk due to software faults: Estimates of failure rate versus evidence of perfection. In *Software Testing, Verification and Reliability*, 1998.
5. Peter Bishop, Robin Bloomfield, Bev Littlewood, Andrey Povyakalo, and David Wright. Towards a formalism for conservative claims about the dependability of software-based systems. In *IEEE Transactions on Software Engineering, Volume 37, Issue 5*, 2011.
6. Robin Bloomfield and Peter Bishop. Safety and assurance cases: Past, present, and possible future - an adelard perspective. In *Making Systems Safe*, 2010.
7. Robin E Bloomfield, Bev Littlewood, and David Wright. Confidence: its role in dependability cases for risk assessment. In *International Conference on Dependable Systems and Networks*, 2007.
8. Rick Chapman. Safety assurance for embedded software in infusion pumps. Presented as a keynote talk at FHIES/SEHC, 2014.
9. Lukasz Cyra and Janusz Górski. Supporting expert assessment of argument structures in trust cases. In *9th International Probability Safety Assessment and Management Conference PSAM*, 2008.
10. Ewen Denney, Ganesh Pai, and Ibrahim Habli. Towards measurement of confidence in safety cases. In *2011 International Symposium on Empirical Software Engineering and Measurement*, 2011.
11. John B. Goodenough, Charles B. Weinstock, and Ari Z. Klein. Toward a theory of assurance case confidence. Technical report, Carnegie Mellon, 2012.
12. William S. Greenwell, John C. Knight, C. Michael Holloway, and Jacob J. Pease. A taxonomy of fallacies in system safety arguments. In *International System Safety Conference*, 2006.
13. S Grigorova and T S E Maibaum. Taking a page from the law books: Considering evidence weight in evaluating assurance case confidence. In *Software Reliability Engineering Workshops*, 2013.

14. R.D. Hawkins, Tim Kelly, John Knight, and Patrick Graydon. A new approach to creating clear safety arguments. In *Advances in Systems Safety*, 2011.
15. R.D. Hawkins and T.P. Kelly. Software safety assurance - what is sufficient? In *4th IET International Conference on Systems Safety*, 2009.
16. Audun Jøsang and Tyrone Grandison. Conditional inference in subjective logic. In *Proceedings of the 6th International Conference on Information Fusion*, 2003.
17. Tim Kelly. *Arguing Safety-A Systematic Approach to Safety Case Management*. PhD thesis, The University of York, 1998.
18. Tim Kelly. Reviewing assurance arguments - a step-by-step approach. In *Safety Management Requirements for Defence System*, 2007.
19. Armen Der Kiureghian and Ove Ditlevsen. Aleatory or epistemic? does it matter? In *Journal of Structural Safety*, 2008.
20. John Knight. Private e-mail communication, 2014.
21. Nancy Leveson. Cost-effective safety certification of software-intensive systems. In *Seventh Software Certification Consortium*, 2011.
22. Bev Littlewood and David Wright. The use of multilegged arguments of increase confdience in safety claims for software-based sytems: A study based on a bbn analysis of an idealized example. In *IEEE Transactions on Software Engineering*, 2007.
23. Benjamin D. Rodes, John C. Knight, and Kimberly S. Wasson. A security metric based on security arguments. In *WETSoM '14*, 2014.
24. John Rushby. Formalism in safety cases. In *Making Systems Safer*, 2010.
25. John Rushby. Logic and epistemology in safety cases. In *Proceedings of SafeComp 32*, 2013.
26. Lorenzo Strigini. Engineering judgement in reliability and safety and its limits: what can we learn from research in psychology. Technical report, Centre for Software Reliability Technical Report, 1996.
27. Laura P. Swiler, Thomas L. Paez, and Randall L. Mayes. Epistemic uncertainty quantification tutorial. In *Proceedings of the IMAC-XXVII*, 2009.
28. Stephen Toulmin. *The Uses of Argument*. Cambridge University Press, 1958.
29. Rob Weaver, Jane Fenn, and Tim Kelly. A pragmatic approach to reasoning about the assurance of safety arguments. In *8th Australian Workshop on Safety Critical Systems and Software (SCS'03)*, 2003.
30. Peter Wilkinson. The use of safety cases in certification and regulation by nancy leveson  a review by peter wilkinson. Technical report, US Chemical Safety Board, 2014.
31. Xingyu Zhao, Dajian Zhang, Minyan Lu, and Fuping Zeng. A new approach to assessment of confidence in assurance cases. In *Computer Safety, Reliability, and Security*, 2012.