Software Engineering Institute

Toward a Theory of Assurance Case Confidence

John B. Goodenough
Charles B. Weinstock
Ari Z. Klein

September 2012

TECHNICAL REPORT
CMU/SEI-2012-TR-002
ESC-TR-2012-002

Research, Technology, and System Solutions Program

http://www.sei.cmu.edu

Carnegie Mellon

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8721-05-C-0003 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

This report was prepared for the

SEI Administrative Agent
AFLCMC/PZE
20 Schilling Circle, Bldg 1305, 3rd floor
Hanscom AFB, MA 01731-2125

DM-0000054

Table of Contents

Abstract **vii**

1 Introduction **1**

2 Prior Work on Assurance Case Confidence **3**

3 Argumentation Theory **6**
3.1 Defeasible Reasoning 6
3.2 Eliminative Induction 10

4 A Framework for Confidence **12**
4.1 Overview 12
4.2 A Simple Example 13
4.2.1 Rebutting Defeaters for the Example 14
4.2.2 Undermining Defeaters for the Example 14
4.2.3 Undercutting Defeaters for the Example 15
4.2.4 Building a Confidence Case 15

5 Applying the Framework: Issues and Concerns **21**
5.1 Toward a Calculus of Confidence 21
5.2 What if a Defeater Cannot Be Eliminated? 22
5.3 Probability as a Measure of Confidence 22
5.4 Confidence and Argument Structure 23
5.5 Other Issues 23

6 Summary **24**

Appendix A Goal Structuring Notation **25**

Appendix B Glossary **27**

References **29**

List of Figures

Figure 1:	A Skeletal Assurance Case	1
Figure 3:	Confidence Argument for ACP1—Adapted from Hawkins	5
Figure 4:	The Components of an Argument	12
Figure 5:	Path Diagram with Basic Blocks	13
Figure 6:	Basic Block Assurance Case	14
Figure 7:	Basic Block Example with Assurance Claim Points	17
Figure 8:	Top Level of a Confidence Case for Figure 6	17
Figure 9:	The Confidence Case for Rebutting Defeaters	18
Figure 10:	The Confidence Case for Undercutting Defeaters	19
Figure 11:	The Confidence Case for Undermining Defeaters	20
Figure 12:	Summarizing How Much Confidence Is Justified in ACP2	20
Figure 13:	A Simple Argument with Confidence Numbers	21
Figure 14:	Goal Structuring Notation	26

List of Tables

Table 1: The Three Kinds of Defeaters 9

Abstract

Assurance cases provide an argument and evidence explaining why a claim about some system property holds. This report outlines a framework for justifying confidence in the truth of such an assurance case claim. The framework is based on the notion of *eliminative induction*—the principle (first put forward by Francis Bacon) that confidence in the truth of a hypothesis (or claim) increases as reasons for doubting its truth are identified and eliminated. Possible reasons for doubting the truth of a claim (*defeaters*) arise from analyzing an assurance case using defeasible reasoning concepts. Finally, the notion of Baconian probability provides a measure of confidence based on how many defeaters have been identified and eliminated.

1 Introduction

An assurance case is somewhat similar in content to a legal case. In a legal case, there are two basic elements. The first is evidence, be it witnesses, fingerprints, DNA, or the like. The second is an argument given by the attorneys as to why the jury should believe that the evidence supports (or does not support) the claim that the defendant is guilty (or innocent). In assurance cases, evidence (e.g., test results) relevant to a property of interest (e.g., safety, reliability, or security) is similarly combined with an argument showing how that evidence supports the property of interest.

A goal-structured assurance case specifies a claim regarding a property of interest, evidence that supports that claim, and a structured argument explaining how the evidence supports the claim. Given a claim and a supporting argument, our goal is to understand how much confidence we should have in the claim and why. For example, consider the skeletal assurance case shown in Figure 1.[1] Evidence in the case is offered to show that specific hazards have been eliminated. The implicit argument is that if these hazards have been eliminated, the system is safe. Given this case, how confident should we be that the system is actually safe, and what is the basis for this confidence? What does it mean to say we have some degree of confidence? Is it a measure of how likely the claim is to be true? Or is it a measure of how justified we are in believing the claim to be true, and if so, what is the basis for justification? If we need to improve our confidence in any of these claims, what should be done and why?

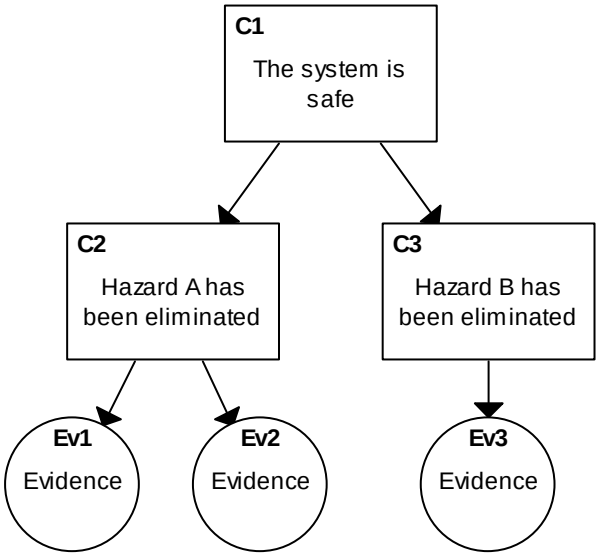

Figure 1: A Skeletal Assurance Case

Adequate answers to these and similar questions don't exist today. This report presents a framework for thinking about (and determining) confidence in assurance case arguments. The frame-

[1] The figure uses the goal structuring notation (GSN) invented by Tim Kelly [Kelly 1998]. We provide a brief summary of the notation in Appendix A.

work uses argumentation theory as developed in philosophy, jurisprudence, mathematics, and artificial intelligence to provide a justified basis for asserting some level of confidence in the truth of assurance case claims.

The next section of this report discusses related work regarding confidence in assurance cases. Section 3 presents relevant argumentation theory concepts drawn from disciplines where assessing support for a claim has been long studied. We will use these concepts in our confidence framework. Section 4 introduces our ideas regarding confidence along with an extended discussion of a simple program example. Section 5 briefly discusses some questions that arise when our framework is applied in certain assurance situations. Section 6 summarizes this report.

2 Prior Work on Assurance Case Confidence

Confidence is a slippery subject. There is an "I'll know it when I see it" quality to measuring it. The lack of a clear and precise definition of confidence is an impediment to achieving progress in measuring the impact of evidence and argument in assurance cases. This, in turn, makes it impossible to make intelligent decisions about where to apply scarce assurance resources.

Work on confidence and assurance cases dates back to at least 2003 when Bloomfield and Littlewood discussed a number of issues that intuitively affect the degree of confidence one would have in a case [Bloomfield 2003]. They proposed no quantitative measures or approach to assessing confidence. However, Bloomfield, Littlewood, and Wright later proposed a formal quantitative treatment of confidence. Their paper related confidence to the concept of safety integrity levels (SILs), which measure the risk of dangerous failure in safety-critical systems [Bloomfield 2007]. SILs range from one to four, with a higher number implying less risk (i.e., a lower probability of failure on demand). In their view of confidence, one is more confident in a system said to be operating at SIL3 if it has already been demonstrated to be operating at SIL4, that is, given a demonstration that a system is operating at a very high level of dependability, one can be said to have more confidence in a less stringent claim about its dependability.

In addition, Bloomfield, Littlewood, and Wright put forward two approaches for evaluating confidence. The first relies on confidence being measured as a function of evidence—confidence grows quickly in the initial stages of gathering evidence but eventually increases only marginally as more evidence is added. The second approach is to construct diverse supporting arguments showing that certain failure possibilities are adequately mitigated or eliminated. The idea is that one has more confidence in a claim that is supported by independent arguments, for example, by both a proof and test results. However, the paper does not discuss how to determine the amount of confidence gained.

Littlewood and Wright set out to show how diversity in arguments can increase confidence similarly to how diversity in systems can increase reliability [Littlewood 2007]. (A diverse one-out-of-two system is potentially more reliable than its constituent systems because if one fails the other compensates.) They found that while it is possible and straightforward to decompose an argument with multiple legs into multiple single-legged arguments, it is not at all straightforward to compose multiple single-legged arguments into a single multi-legged argument because of the need to model dependencies between the legs. They also showed, via a Bayesian Belief Network model, that confidence in a dependability claim arising from a diverse two-legged argument need not be greater than that arising from either of the single arguments, and in theory, can be less (e.g., if the second argument casts doubt on elements of the first).

Bloomfield and Bishop equate confidence to the likelihood of a claim being true and how such a likelihood depends on the supporting argument and evidence [Bloomfield 2010]:

> *For a given claim the confidence* [that is, probability of its being true] *– and its complement, doubt – will depend upon confidence/doubt in the truth of assumptions, in the correctness of reasoning, and in 'strength' of evidence.*

However, they note that computing and combining these confidence probabilities is not often done:

> *Current practice regarding confidence is often very pragmatic (e.g., 'traffic lighting' of evidence nodes in a graphical case).*

Wassyng and colleagues argue that safety cases, as they are currently developed, are not necessarily appropriate for software certification purposes [Wassyng 2011]. Their concern is that current safety-case approaches for software engineering lack the scientific and measurement principles found in other engineering disciplines such as civil engineering. There is much more to their discussion, but for our purposes their description of needed improvements to safety cases is the most important. In particular, they identify a critical need for repeatable and quantified methods that determine the level of confidence in a particular case. They state the following:

> *Establishing levels of confidence goes to the very heart of the problems with safety cases as they are currently defined. The key ingredient missing ... is the ability to make an objective, repeatable assessment of the confidence one should have in the safety of the system. The assessment procedure for determining one's confidence in a product's safety is no more and no less than a proper engineering method. ...[D]etermining levels of confidence in safety cases is "arguably" the most important issue for future investigation.*

Hawkins and colleagues directly address the confidence issue. They introduce the concept of an *assured safety argument* that explicitly splits a traditional safety case into two pieces [Hawkins 2011]. The first is the safety argument—an argument supported by evidence showing the desired safety property. The second is a *confidence* argument that provides an argument and evidence justifying a degree of confidence in the safety argument. The purpose of the confidence argument is to address uncertainties that underlie the safety argument. The reason for the split is that "…presenting both in an intermingled fashion typically results in a larger (often rambling) argument and makes grasping the critical structures difficult for the reader."

Hawkins and colleagues provide a notional structure and notation for stating a confidence argument. The confidence argument is attached to the safety argument by the use of *assurance claim points* (the small black squares labeled ACPi in Figure 2). Assurance claim points (ACPs) may be attached to inferences (the arrows connecting assurance case claims), contexts (explanatory information attached to other nodes), or evidence. An example of a (partial) confidence argument that might be created for ACP1 is shown in Figure 3.

Hawkins and colleagues take a qualitative view of confidence. Their confidence argument shows that (1) there is reason to believe in the probable validity of the safety argument, (2) remaining uncertainties have been identified, and (3) these residual uncertainties are not a cause for concern. For very large safety cases, it may not be practical to develop a confidence argument for every assurance claim point—prioritization by likely risk is necessary.

In summary, a variety of papers say that defining and measuring confidence in assurance claims is an important and unresolved issue. A framework for determining confidence is needed, and this report takes some initial steps in providing one.

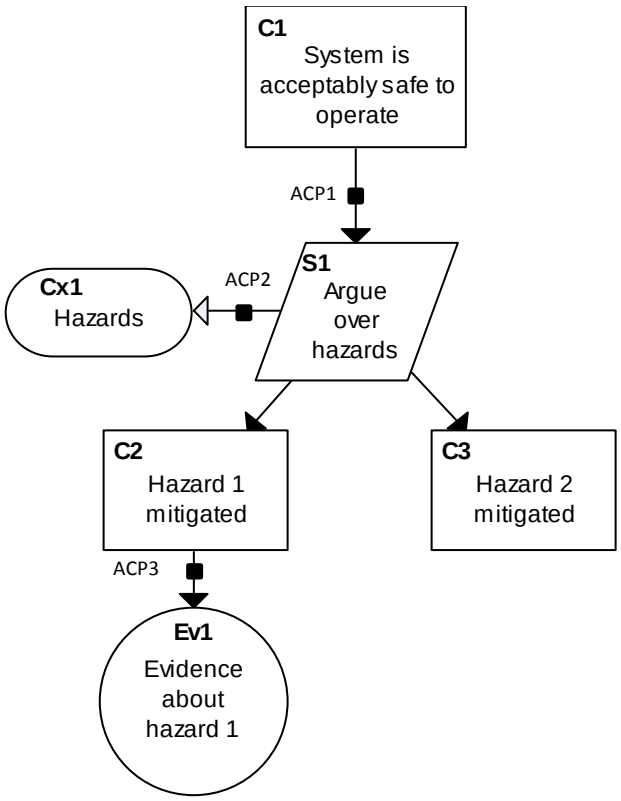

Figure 2: Assurance Claim Points—from Hawkins
[Hawkins 2011]

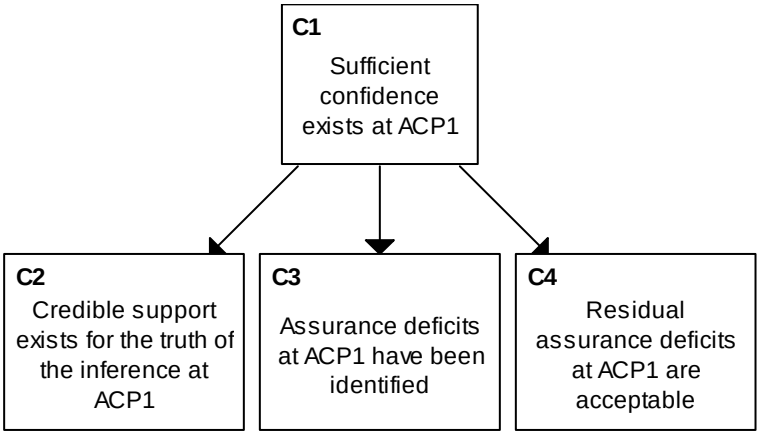

Figure 3: Confidence Argument for ACP1—Adapted from Hawkins
[Hawkins 2011]

3 Argumentation Theory

In the argumentation literature, much of the existing work related to notions of confidence has been concerned with weighing which hypothesis or conclusion in a given set of possibilities is most strongly inferable from some pool of evidence [Schum 2001]. For example, from a legal perspective, one might be interested in inferring which suspect, out of several potential criminals, robbed the bank, given evidence of where the suspects said they were at certain times, witness testimonials, and video surveillance. From an artificial intelligence perspective, one might be interested in how an artificial agent revises or maintains its belief in a reasonably derived conclusion as additional information becomes available.

The goal of these kinds of argumentation is to choose the hypothesis or conclusion, among several potential ones, that is the most acceptable or believable on the basis of the available evidence. Such argumentation theories are not directly relevant to assurance cases. In assurance case argumentation, rather than laying out several possible claims and searching for which is best supported by our existing collection of information, we start by asserting a single claim whose truth we wish to determine, and we gather evidence in support of that single claim. This difference means that certain complications that are important in argumentation theory (e.g., the interplay of argument and counter-argument) do not need to be addressed for assurance cases. Of course, other issues are directly relevant; for example, since arguments are based on a hierarchy of claims, we are also interested in understanding how confidence in higher level assurance case claims is affected by confidence in lower level claims and information.

Although our argumentation goal is different from that with which much argumentation theory is concerned, we can nonetheless draw on existing argumentation theories for a theory of assurance case confidence. In particular, we ground our theory on the intersection of two frameworks of reasoning—defeasible reasoning (discussed in Section 3.1) and eliminative induction (discussed in Section 3.2). Taken together, these frameworks provide both a conceptual vocabulary for identifying information that is relevant to measuring claim confidence and a basic method for justifying an increase in confidence.

In the remainder of this section, we introduce and explain argumentation principles and terminology that we later apply to assurance cases.

3.1 Defeasible Reasoning

In general, the argumentation in an assurance case follows a form of reasoning called *defeasible reasoning*:

> *Reasoning is defeasible when the corresponding argument is rationally compelling but not deductively valid. The truth of the premises of a good defeasible argument provide support for the conclusion, even though it is possible for the premises to be true and the conclusion false. In other words, the relationship of support between premises and conclusion is a tentative one, potentially defeated by additional information [MRL 2005].*

One logical form of a defeasible inference supporting a claim, C, is

[1] if E then (usually) C unless R, S, T, …

That is, for an assurance case, given evidence E, then claim C follows, unless this claim is invalidated because one or more of *R, S, T, etc*. hold. In defeasible reasoning, any claim is always subject to revision based on additional information. The ellipsis in the logical formulation above is significant as a means of indicating that the set of exceptions (*R, S, T,* …) to the inference from E to C is never completely known. These exceptions are known as *defeaters*.[2] Our fundamental principle is that each defeater is a source of doubt about the truth of the claim: As we are able to eliminate (i.e., falsify) defeaters, we eliminate doubt and thereby achieve increased confidence in a claim. Given that elimination of defeaters is fundamental to our framework, we need a method for identifying them. We draw from the defeasible reasoning literature to determine the different classes of defeaters that are applicable in assurance case argumentation.

The remainder of this section introduces the three kinds of defeaters we will consider in determining sources of doubt in assurance case arguments:

1. *Rebutting* defeaters are defeaters that eliminate belief in a claim by providing information that contradicts the claim.
2. *Undercutting* defeaters are defeaters that specify conditions under which the claim is not necessarily true even if the premises are true.
3. *Undermining* defeaters are defeaters that invalidate one or more of the premises (in which case even if the inference rule is valid and all rebutting defeaters have been eliminated, we still have a reduced basis for believing in the truth of the associated claim).

Each of *R, S,* and *T* in equation [1] are one (and only one) of these defeater types.

We use the following terminology in the remainder of this paper. An *inference rule* is a generalization that is (usually) considered to be true; for example, "All men are mortal" or "if X is a man then X is mortal." An *argument* is an instantiation of one or more inference rules with a specific premises and conclusions; for example, given the *premise* "Socrates is a man" and the rule "All men are mortal," the instantiated inference rule implies the conclusion "Socrates is mortal."

Stephen Toulmin was one of the first to attempt to formalize a model of argumentation on the assumption that argumentation claims can be defeated (or at least weakened) and qualified. In *The Uses of Argument*, Toulmin argued that formal logic, with its criteria for deductive validity, misrepresented and oversimplified how we actually reason in everyday contexts [Toulmin 1958]. From Toulmin's perspective, argumentation did not consist merely of airtight, logical leaps from a set of premises to a conclusion. Unlike in traditional logic, where a conclusion can be asserted as true only if it is guaranteed to be universally true by the truth of its premises and its inference rule, Toulmin saw truth in practical argumentation as much more dynamic and evolving. He saw the

[2] Just as it is unlikely that a hazard analysis will identify all potential hazards, we will never be able to identify all potential defeaters. For example, a system's operating environment may change, it may be used in ways never before imagined, and so on. The best we can hope for is to identify as many defeaters (or hazards) as is reasonable and practicable, and use that as a basis for argumentation/confidence discussions.

need to take into consideration argumentation structures in which true premises will not always imply the truth of a conclusion.

According to Toulmin, argumentation begins with a claim (as it does in assurance cases), and we need to defend the claim with data (i.e., evidence) that supports it. Once we have data, we may also be asked to defend our inference from the data to the claim, that is, why the data are believed to support the claim. The mental step that authorizes our move from the data to the claim is called the warrant (W), which is often left implicit in argumentation. The warrant is the inference rule whose applicability we seek to establish when we use data to justify the truth or acceptability of a claim. Whereas traditional logics tend to *accept* the truth of warrants (e.g., All men are mortal) and use these warrants to infer conclusions from premises, the logical structure of the Toulmin model assumes that moving from data to claims in practical argumentation requires that the validity of a warrant be *established*. In Toulmin's model, a warrant is an inherently defeasible link that often requires its own justification.

In the process of supporting the claim, we may have to weaken the claim with a qualifier (Q) (e.g., "possibly" or "presumably") depending on how generally acceptable we think (the inference of) the claim is. A level of confidence in the claim is expressed in the Toulmin model by this linguistic qualifier; that is, the qualifier marks how justified we are in accepting the claim as true given the data that supports it. The major implication of Toulmin's model is that we become increasingly more justified in accepting the claim as true when there are no rebuttals (R) to defeat it. If we are unable to think of any conditions of rebuttal that could challenge the truth of a claim (given the data), or if we have identified potential rebuttals that we can reject, then we would not have to qualify the claim, or we might qualify it with the words *necessarily* or *certainly*.

In more recent developments in defeasible reasoning, Toulmin's notion of a rebuttal has come to be known as a "defeater." John L. Pollock defines a defeater as "information that can mandate the retraction of the conclusion of a defeasible argument" [Pollock 2008]. He distinguishes between two kinds of defeaters: rebutting defeaters and undercutting defeaters. *Rebutting defeaters* "attack an argument by attacking its conclusion." They give us reason to doubt that a conclusion is true. A rebutting defeater can be thought of as information from which we can infer a counterclaim to the conclusion of an argument. In this way, rebutting defeaters represent conditions under which the conclusion of a defeasible argument is necessarily false. For example, in (an adaptation of) one of the arguments that Toulmin analyzed, "Harry was born in Bermuda, so Harry is a British citizen," a possible rebutting defeater would be that Harry renounced his British citizenship [Toulmin 1958]. If we have evidence that Harry renounced his British citizenship, then the conclusion "Harry is a British citizen" is necessarily false even if the premise "Harry was born in Bermuda" is true. In general, rebutting defeaters for this example take the form "R, so claim C is false," ("R, so Harry is not a British citizen") where the inference R → ~ C is *not* defeasible. That is, R represents a condition under which C will *always* be false.

Undercutting defeaters, on the other hand, "attack the inference [rule] itself, without doing so by giving us reason for thinking that the inference [rule] has a false conclusion" [Pollock 2008]. In other words, undercutting defeaters give us reason to doubt that a conclusion can be inferred from a particular premise. The inference rule for the argument "Harry was born in Bermuda, so Harry is a British citizen" is "All persons born in Bermuda are British citizens." An undercutting defeater of this defeasible rule is that both of Harry's parents were citizens of only the U.S. at the time

of Harry's birth, that is, when Harry was born, they were aliens residing, but not legally settled, in Bermuda. Children born in Bermuda to two alien parents who are not legally settled in Bermuda are not, by birth alone, British citizens. But if we have evidence that Harry's parents were both U.S. citizens (i.e., aliens who are not legally settled in Bermuda), that information alone is not enough to rebut the claim "Harry is a British citizen" because Harry could have resided in the United Kingdom long enough to apply for and obtain British citizenship. If this were the case, then Harry would in fact be a British citizen (i.e., the conclusion would be true), regardless of the citizenship or residency of his parents, that is, even if the undercutting defeater were true.

The difference between rebutting and undercutting defeaters is subtle because both cast doubt on the validity of a conclusion. The essential difference is that a rebutting defeater says a conclusion is *necessarily* false, but an undercutting defeater says a conclusion is *not necessarily true*, that is, an undercutting defeater specifies conditions under which the conclusion, if true, is not adequately justified.

Undercutting defeaters for the argument "Harry was born in Bermuda, so Harry is a British citizen" can be viewed in two ways—one with respect to the argument's conclusion and another with respect to the argument's inference rule. In the following, U represents the undercutting defeater, "Harry's parents were not British citizens or legally settled in Bermuda when Harry was born."

- With respect to the argument's conclusion, the undercutting defeater can be viewed as attacking the conclusion "if U, then claim C can be false" ("if U, then Harry might not be a British citizens"), where the inference U → ~ C *is* defeasible, that is, U does not always imply ~C.
- With respect to the argument's inference rule, an undercutting defeater can be viewed as a condition under which the inference rule is not necessarily valid, that is, a condition under which the rule does not necessarily imply the specified conclusion. For our example, the impact of an undercutting defeater can be expressed as "U, so inference rule I is invalid" ("if U, then the rule, 'All persons born in Bermuda are British citizens,' is invalid"), where the inference U → ~ I is *not* defeasible.

Henry Prakken calls attention to a third kind of defeater: *undermining defeaters* [Prakken 2010]. Whereas rebutting defeaters and undercutting defeaters attack the conclusion and inference rule respectively, undermining defeaters attack the premises from which the conclusion is inferred. Like undercutting defeaters, the truth of an undermining defeater does not necessarily imply the falsity of the conclusion. A possible undermining defeater of the argument "Harry was born in Bermuda, so Harry is a British citizen" would be that Harry was born in London. If Harry were born in London, then the premise "Harry was born in Bermuda" is false, even though Harry could still be a British citizen. In general, undermining defeaters for this example take the form "M, so premise P is false" ("undermining defeater M, so Harry was not born in Bermuda") where the inference M → ~ P is *not* defeasible. That is, M represents a condition under which P will always be false.

Table 1 summarizes the three kinds of defeaters. To recap, in terms of the logical form of a defeasible inference presented in equation [1], "If Harry was born in Bermuda (E), then Harry is a British citizen (C) unless Harry renounced his British citizenship (R), both of Harry's parents were aliens in Bermuda at the time of Harry's birth (U), or Harry was actually born in London (M)." If

any of these defeaters is true, we cannot be confident, solely on the basis of the proffered evidence and inference rule, that Harry is a British citizen. On the other hand, if all these defeaters are false (and they are the only ones we have identified), we have no reasons to doubt the claim that "Harry is a British citizen."

The key point here is whether we have doubt about the claim. The fact that some of the defeaters may imply the claim is false is simply not relevant because we are only interested in eliminating reasons for doubting the conclusion, that is, we are only interested in eliminating defeaters. Harry's possible naturalized status as a British citizen does not enter into the validity of the argument about Harry's citizenship based on his place of birth, but it is relevant to the status of U as an undercutting defeater.

Table 1: The Three Kinds of Defeaters (if E then C unless R, U, M)

Type of Defeater	Attacks	Form	Example
Rebutting (R)	claim	R, so claim C is false	Harry renounced his British citizenship.
Undercutting (U)	inference rule	if U, then claim C can be true or false	Both of Harry's parents were aliens at the time of Harry's birth.
Undermining (M)	evidence	M, so premise E is false	Harry was born in London.

3.2 Eliminative Induction

The second framework of reasoning on which we ground our theory of assurance case confidence is called eliminative induction. As first proposed by Francis Bacon [Schum 2001] and extended by L. Jonathan Cohen [Cohen 1970, 1977, 1989], eliminative induction is basically a method for testing support for a claim or hypothesis. While previous work on calculating confidence in assurance cases relied on traditional Pascalian probabilities and Bayesian Belief Networks [Bloomfield 2007, Littlewood 2007, Bloomfield 2010], our framework for confidence is rooted in the Baconian system of probabilities as elaborated by Cohen. *Baconian* probabilities range from total doubt (no reasons for doubt eliminated) to complete acceptance (all identified reasons for doubt eliminated[3]), as opposed to Pascalian probabilities that range from disproof (an event is impossible) to proof (an event is certain).

In eliminative induction, we identify various possibilities for doubting the truth of a hypothesis, and then we gather evidence or perform analyses that eliminate each of these possibilities. Each eliminated possibility removes a reason for doubt and thereby increases our confidence in the hypothesis. If there are *n* possibilities for doubt and *i* of these have been eliminated, Cohen calls *i/n* the *Baconian probability* that the hypothesis is true [Cohen 1970]. (In our framework, the only possibilities for doubt are captured in the defeaters associated with a particular argument.)

To be more specific, consider a hypothesis H_1, and suppose we have identified *n* defeaters that cast doubt on the truth of H_1. As long as H_1 withstands all *n* defeaters (i.e., as long as all defeaters

[3] Although in principle, the number of reasons for doubt can always be increased, in practice, at any given time, the number of identified reasons for doubt is finite.

are shown to be false), we have no reason for doubting that H_1 is true. In general, the Baconian probability $B(H_1) = i/n$ means that H_1 has withstood i out of n defeaters, that is, i out of n reasons for doubt have been eliminated. In the case where $i = n$ and, therefore, $B(H_1) = n/n$, we have no reason to doubt H_1; in this situation, we say we have "total (Baconian) confidence" in H_1. On the other hand, in the case where $B(H_1) = 0/n$, we have zero confidence in H_1. In eliminative induction, zero confidence does not mean that H_1 has been disproven; it means we have not yet eliminated any defeaters relevant to H_1.

One can also view Baconian probabilities as measures of the amount of information we have that is relevant to removing reasons for doubt. $B(H_1) = 0/n$ means no information is available, and $B(H_1) = n/n$ means we have all the information needed to remove all identified reasons for doubting the truth of H_1. Of course, just as a hazard analysis is subject to the addition of more hazards with experience, the set of defeaters always remains subject to the addition of more reasons for doubt. But in practice, we can only work with the doubts (or hazards) identified.

It is important to note that, in the Baconian system of probabilities, i/n is neither a reducible fraction nor a fractional representation of a numerical value between zero and one (as in Pascalian probability). The Baconian probability 2/4 means that, so far, the hypothesis has withstood two out of four potential reasons for doubting its truth, and we are currently missing information to eliminate the other two defeaters. This Baconian probability cannot be represented as 1/2 because such a representation would imply that there are only two relevant defeaters (instead of four) and that the hypothesis has withstood only one of them (instead of two). The probability 1/2 also suggests we are currently missing information to eliminate or falsify only one other defeater (instead of two). Likewise, the probability $B(H_1) = 2/4$ cannot be represented as 0.5 because such a (Pascalian) representation fails to account for how many relevant defeaters there are and how many of them the hypothesis has withstood.

In sum, defeasible reasoning provides a framework for identifying the reasons for doubt used in eliminative induction (namely, the defeaters), and eliminative induction provides a basis for justifying confidence in a claim, namely, by demonstrating that the defeaters are not valid. Our framework for assurance case confidence (which we will present in the next section) is rooted in Toulmin's theoretical assumption that claims in practical argumentation (i.e., outside of formal logic) can and should be qualified [Toulmin 1958]. As Toulmin acknowledges, the truth of a claim is tentative and can change as more information becomes available. The role of Toulmin's qualifiers, then, is to guard against overstating the truth of a claim at any point in time. In our framework, we qualify the truth of an assurance claim by attributing a level of confidence to it. We determine this level of confidence for each claim by identifying the relevant defeaters and assessing how many of them we can eliminate by using the information that is currently available.

4 A Framework for Confidence

4.1 Overview

In Section 3.1, we discussed how an argument consists of a premise, a conclusion, and an instantiated inference rule. In an assurance case structure such as that shown in Figure 4, the downward arrows between claims or between claims and evidence represent the instantiation of an inference rule. Somewhat counter to intuition, the elements at the ends of the arrows are premises, and the claim at the start of the arrow is the conclusion; for example, Ev1 and Ev2 are premises for the conclusion captured in C2. The diagram can be read as follows: "The system is safe *because* hazards A and B have been eliminated. These hazards have been eliminated *because* Ev1 and Ev2 *imply* that hazard A has been eliminated and Ev3 *implies* that hazard B has been eliminated." The italicized words represent the implicit inference rules (the warrants, in Toulmin's terms) being used to support the conclusions. The inference rule connecting C1 with C2 and C3 is something like "If all hazards have been eliminated, a system is safe." Usually the collection of arrows between a claim and one or more premises represents a single, implicit inference rule, but sometimes, for example, in the case of so-called "multi-legged" [Bloomfield 2003] or "diverse" [Kelly 1998] arguments, more than one inference rule may be at work. As we mentioned in the Introduction, our goal is to provide a justified basis for explaining how much confidence we should have in the truth of C1.

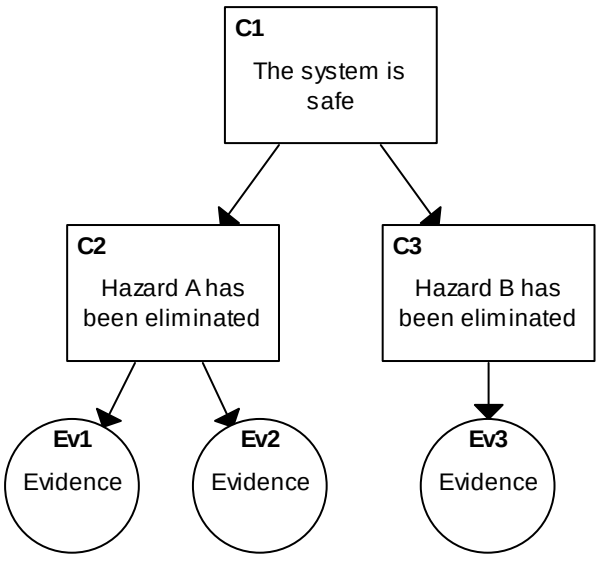

Figure 4: Components of an Assurance Case

We use the following general definition:

Confidence (in the truth of a claim): belief (in the truth of a claim)

As stated in Section 3.2, our framework for assessing confidence is based on the principle of eliminative induction, that is, the principle that a conclusion is justified only to the extent that reasons for doubting its truth have been eliminated. The reasons for doubt in our framework are the defeaters from argumentation theory discussed in Section 3.1. Therefore, in our framework, confi-

dence in a claim can vary, depending on the number of defeaters and how many are eliminated. No confidence means no defeaters have been eliminated, that is, no reasons for doubting the truth of the claim have been eliminated. As we eliminate defeaters, confidence grows until we have no further reasons to doubt the claim.

We distinguish between "justification for a claim" and "justification for *belief* in a claim." *Justification for a claim* is provided by the argument and supporting evidence for a claim. *Justification for* belief *in a claim* is provided by identifying and eliminating defeaters relevant to the claim.

Note that the degree of confidence in a claim has little or nothing to do with the (Pascalian) probability that a claim is true. Confidence is just a function of how many reasons for doubt have been identified and removed. We will return to this point after considering several examples.

4.2 A Simple Example

To ground the discussion throughout this section, consider Figure 5, which shows a program with four basic blocks.[4] We define an *egregious* error as one that is detected by *every* execution of a statement containing such an error, that is, *every* execution fails when encountering an egregious error.[5]

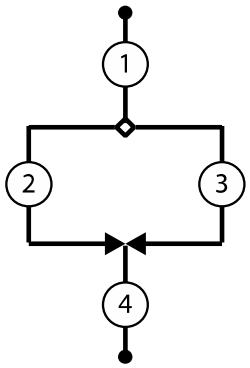

Figure 5: Path Diagram with Basic Blocks

Suppose we run one test case and it executes successfully. On the basis of this single test, how much confidence should we have that there are no egregious errors in the program?

To start, let's consider a possible assurance case for this example (see Figure 6). This case reasons that if there are no egregious errors in any basic block, there are no egregious errors in the program. The single test case shows successful execution of three of the four basic blocks. Let's consider the possible rebutting, undermining, and undercutting defeaters associated with claim C2 in order to determine how much confidence we should have in claims C1 and C2.

---

[4] A basic block is a sequence of instructions with only one entrance and one exit.

[5] In an interpretive language, calling an undefined function would be an egregious error. In a compiled language, using the wrong formula could be an egregious error. Note that a basic block can contain a non-egregious error, that is, one that is revealed by some, but not all, executions of the code.

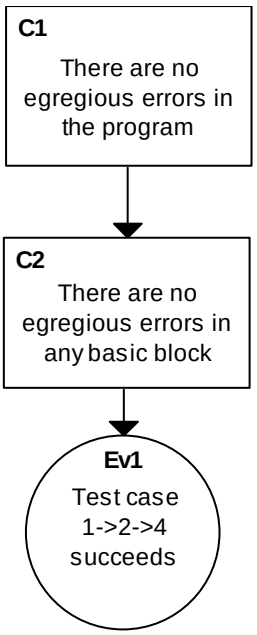

Figure 6: Basic Block Assurance Case

4.2.1 Rebutting Defeaters for the Example

There are four rebutting defeaters for claim C2, one for each basic block. Each rebutting defeater has the same form: "There is an egregious error in block *i*." Such a defeater is eliminated *only* by successfully executing block *i* at least once.

The test case given in Ev1 successfully executes three of the four basic blocks, so three of the four possible ways of rebutting claim C2 have been eliminated, and we have no information about the fourth possible rebutting defeater. Prior to the test's successful execution, no defeaters were eliminated, so our level of confidence was zero out of four (expressed as 0/4). After execution of the test, our level of confidence increased to 3/4, that is, B(C2|Ev1) = 3/4. It is tempting to say our confidence is 75%, but remember that this is not what the notation means. Instead, we say, "The Baconian probability of C2, given Ev1, is 3 out of 4, which is expressed as 3/4." Baconian probability pertains to the *amount of information* we have about possible reasons for doubting that a claim is true.

Note that if a test does not eliminate any defeater, the test is irrelevant; it cannot increase our confidence in any claim.

4.2.2 Undermining Defeaters for the Example

Undermining defeaters attack the premises of an inference. In the case of our simple example, we have just one premise so far: a test has executed blocks 1, 2, and 4 successfully.

Here are some typical defeaters for test results:

- The test oracle is wrong, that is, the declaration of test success is invalid. Eliminating this defeater requires gathering evidence about the reliability of the method used to determine test success or failure.

- The test results are not for the current system. In a real system, this defeater might be eliminated by reviewing the configuration management mechanisms being used and the extent to which one can be confident that test results continue to apply to a system as it undergoes changes.

An additional defeater for C2 would be "Blocks 1, 2, and 4 have not, in fact, been executed by the test case" or more generally

> Assertions about what blocks have been executed are incorrect.

To know which blocks have been covered by a particular test case requires some kind of monitoring or analysis. This defeater calls into question whether the monitoring or analysis has been done correctly.

Note that the truth of any of these three defeaters would mean the inference rule does not provide support for the stated claim because its premises are not valid. Confidence in the conclusion is therefore reduced to the extent that these defeaters have not been eliminated.

Suppose two of these three defeaters have been eliminated, and we have no information about the third. Considering just the undermining defeaters and using the principle of eliminative induction, our confidence in C2 could be stated as 2/3. Considering just the rebutting defeaters, our confidence in C2 is 3/4. How these should be combined to arrive at an overall confidence score is discussed later.

4.2.3 Undercutting Defeaters for the Example

Undercutting defeaters attack the inference rule used to assert that a conclusion is true if the premises hold. If an undercutting defeater is true, then nothing can be inferred about the truth of the associated conclusion.

The inference rule connecting the successful test result(s) to claim C2 is "If all basic blocks are executed successfully at least once, no basic block contains an egregious error." This is, formally, a tautology, that is, given the definition of egregious error, the conclusion is just a restatement of the premise. So there can be no condition under which the premise holds, and the conclusion is false. However, the inference rule linking C2 to C1 is "No egregious error in any basic block implies there are no egregious errors in the program." This rule is equivalent to

$$\forall\, x \in B(\mathrm{P}), \sim E(x) \rightarrow \sim E(\mathrm{P})$$

where $B(\mathrm{P})$ is a set of basic blocks identified as belonging in program P, $E(\mathrm{P})$ is "egregious error in program P," and $E(x)$ is "egregious error in basic block x." If P contains basic blocks that are not included in $B(\mathrm{P})$, that is, if not all basic blocks have been correctly identified, there could be an egregious error in an undetected basic block, so the conclusion $\sim E(\mathrm{P})$ would not necessarily be true. Consequently, an undercutting defeater here is "Not all basic blocks have been identified."

4.2.4 Building a Confidence Case

Up to now, we have illustrated our notion of confidence by considering defeater types individually, but our goal is to show why we should have a certain degree of confidence in some system claim. That is, we seek to make a confidence claim and provide an argument supporting that claim

of confidence. As Hawkins and colleagues have suggested, this is best done by separating the *system* case (with claims about system properties) from the *confidence* case (with claims about confidence in the system claims) [Hawkins 2011]. They propose a particular form of confidence case which we discussed in Section 2. Our framework for assessing confidence, based on eliminative induction and the identification of defeaters, builds on their ideas about using an assurance case structure to present confidence arguments, but focuses on showing relevant defeaters and the evidence used to eliminate them. In this section, we present an assurance case structure for supporting a confidence claim.

Our framework is similar to the Hawkins approach in that we both make confidence claims and use an assurance case structure to justify them. However, the generic structure and content of our confidence cases are different from that proposed by Hawkins (see Figure 3) [Hawkins 2011]. The structure shown in Figure 3 shows the start of a confidence argument associated with assurance claim point 1 (ACP1) in Figure 1.

The three claims in the argument apply to any ACP associated with an inference:

1. Credible support exists for the truth of the inference at ACPi.
2. Assurance deficits at ACPi have been identified.
3. Residual assurance deficits at ACPi are acceptable.

Each of these claims supports the claim "Sufficient confidence exists at ACPi."

This generic structure could be framed in terms of our ideas about confidence by equating the identification of

- "assurance deficits at ACPi" with the identification of a set of defeaters relevant to the system claim associated with ACPi
- "credible support" with arguments showing that the defeaters are eliminated
- the acceptability of "residual assurance deficits" with an argument that "sufficient confidence" exists despite the existence of uneliminated defeaters

The content of the confidence case would then be the set of defeaters for the associated system assurance case plus the arguments/evidence showing which defeaters are eliminated.

Our framework for evaluating confidence, however, suggests a different form of confidence case, which we will illustrate with respect to the basic block example. Consider the assurance case in Figure 7, which is the same as Figure 6 but with assurance claim points added. Given our focus on identifying and eliminating defeaters as a measure of confidence, the top-level structure for the ACP2 confidence case would appear as follows.

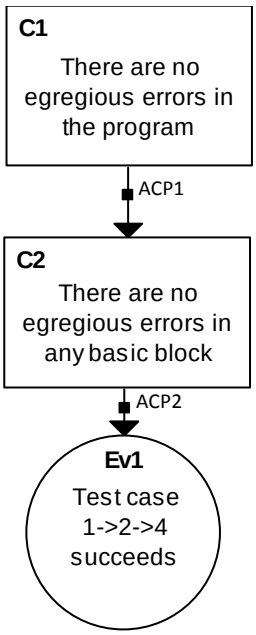

Figure 7: Basic Block Example with Assurance Claim Points

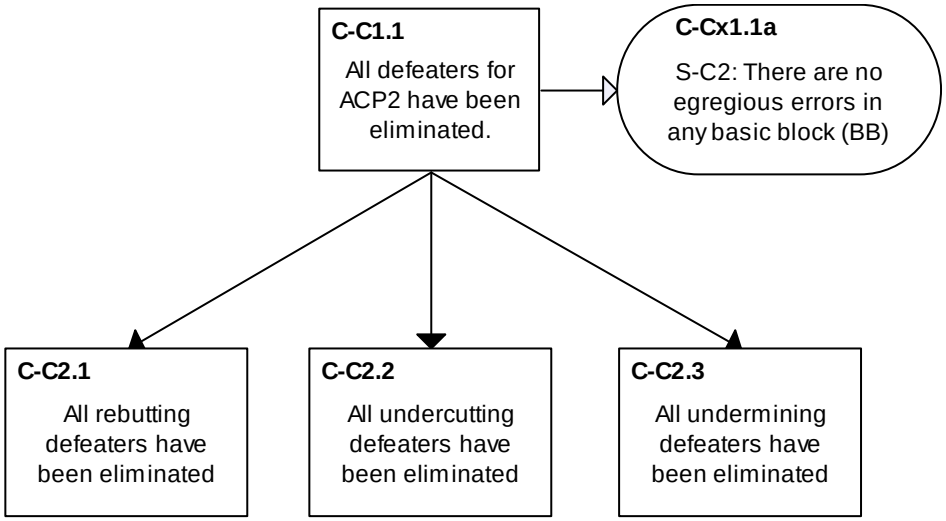

Figure 8: Top Level of a Confidence Case for Figure 7

Since we define confidence in terms of eliminating defeaters, the top-level confidence claim is naturally that all defeaters relevant to the ACP have been eliminated.

We preface the node numbers in the confidence case with "C-" to show that these nodes belong to a confidence case. Since the nodes in a confidence case are related to the nodes in a system case, having a notation that distinguishes the two is convenient. When we need to refer to a node in the system case (e.g., a node in Figure 7), we preface the reference with "S-" to remind the reader that

we are talking about the system case. For example, in node C-Cx1.1a (the context node[6] associated with the top-level claim), we repeat the claim associated with ACP2 in Figure 7, and we call the repeated claim "S-C2" to show that we are talking about a claim in the assurance case for the system.[7]

The argument associated with C-C2.1 is shown below.

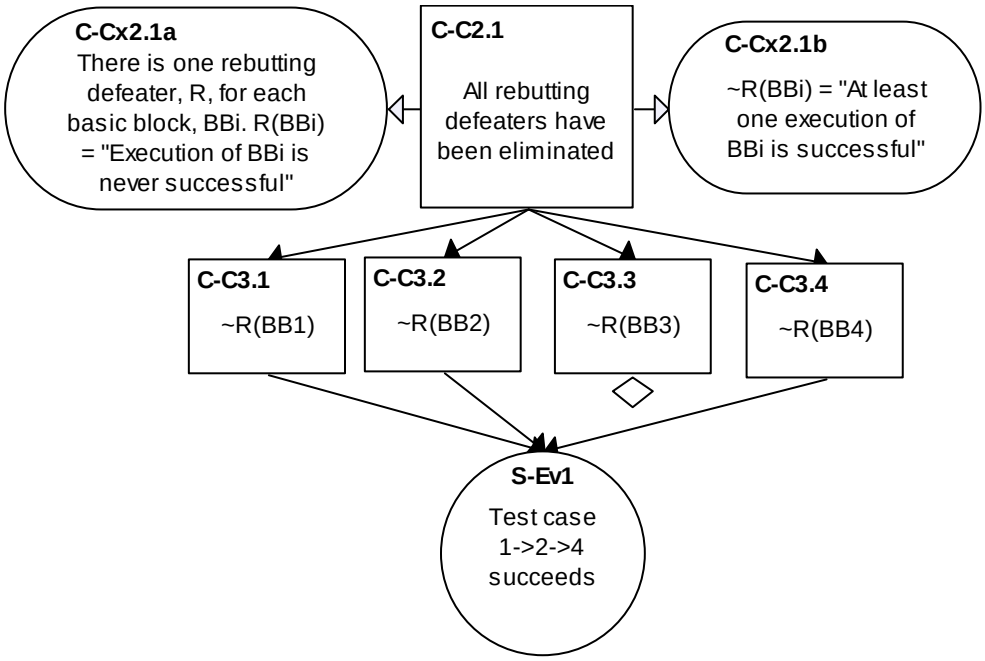

Figure 9: The Confidence Case for Rebutting Defeaters

We have provided two context nodes for claim C-C2.1. The leftmost context node gives the form of rebutting defeater that is relevant to claim S-C2. The rightmost context node explains what constitutes the negation of these rebutting defeaters. Since our goal is to eliminate rebutting defeaters, claims C-C3.1 through C-C3.4 express the requirement that each rebutting defeater be falsified.

The "S" prefix for the evidence node, S-Ev1, indicates that it is the same evidence that appeared in the assurance case involving ACP2. The confidence case shows why this evidence increases our confidence in the truth of claim S-C2. The small diamond at the bottom of C-C3.3 means further development is required, that is, that this defeater is not yet eliminated.

The development of the second subclaim for ACP2 is shown in Figure 10. We use the claim about undercutting defeaters as an opportunity to state explicitly (in C-Cx2.2a) the inference rule that is associated with ACP2.

---

[6] Context nodes are shown as rounded rectangles.

[7] We don't label all the nodes with "S-" in the system case; this would be too much a deviation from current practice. We only use the "S-" notation as a convenient way of distinguishing the system case from the confidence case in our discussion.

Our intent in this part of the confidence case is to remove reasons for doubting the validity of the inference rule. To capture the idea that the inference rule could be invalid under some conditions, we posit a generic definition of the undercutting defeater and eliminate it by showing that the rule is a tautology. The formal proof is the evidence provided in C-Ev5.1.

Finally, we develop the claim about undermining defeaters in Figure 11. The context node associated with C-C2.3 lists the various undermining defeaters that we have considered, and the corresponding claims negating these defeaters are shown beginning with C-C3.6. In this example, we have not provided any further argument or evidence to eliminate these defeaters, as indicated by the diamonds attached to the claims.

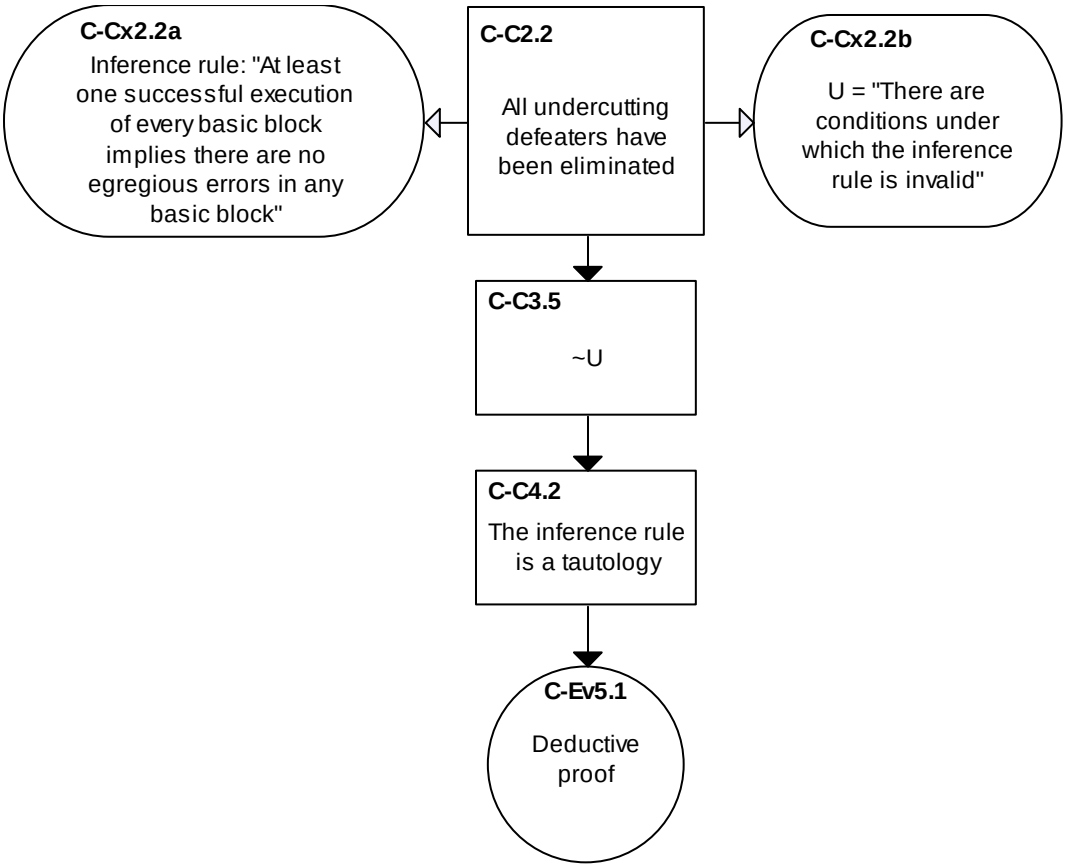

Figure 10: The Confidence Case for Undercutting Defeaters

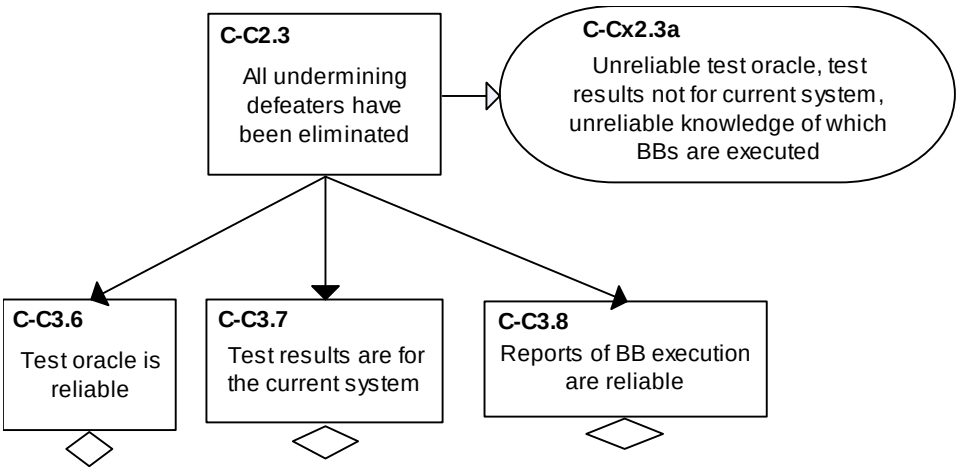

Figure 11: The Confidence Case for Undermining Defeaters

We can now summarize the status of the defeaters for ACP2 by decorating the claims with the Baconian probabilities based on how many defeaters exist and have been eliminated:

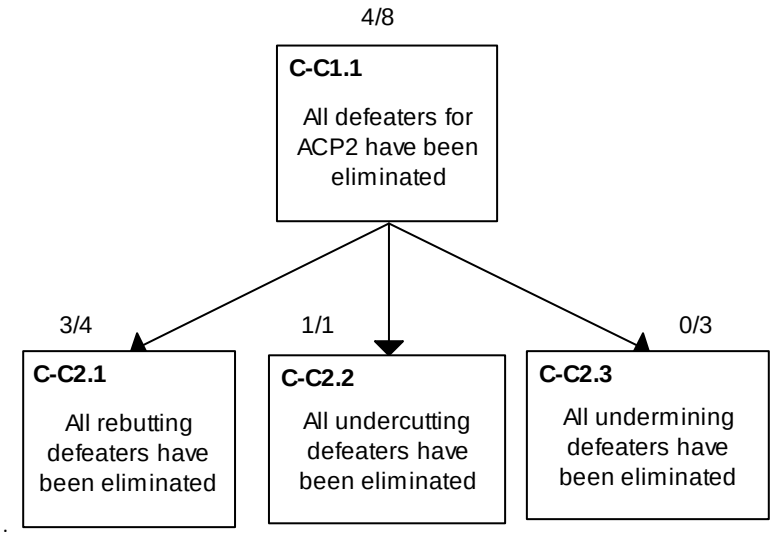

Figure 12: Summarizing How Much Confidence Is Justified in ACP2

The Baconian probability, 4/8, says that of the eight defeaters associated with ACP2, only four have been eliminated, that is, there are four remaining "assurance deficits." The uneliminated defeaters provide reasons for doubting the truth of S-C2. Although it appears that the Baconian probability for the top-level claim is calculated by just adding the numerators and denominators of the subclaim probabilities, 4/8 actually represents the union of all the defeaters associated with the subclaims. Since there are no duplicates or dependencies among the defeaters, all these defeaters apply to S-C2. Consequently, eight defeaters apply to S-C2 of which four have been eliminated, so the Baconian probability associated with C-C1.1 is 4/8. Since this confidence claim is associated with S-C2, the Baconian probability for S-C2 is 4/8.

5 Applying the Framework: Issues and Concerns

In this section, we briefly present some thoughts on how the framework can be applied in various circumstances, and we discuss some issues requiring further investigation.

5.1 Toward a Calculus of Confidence

Figure 12 showed how the confidence "score" for the top-level claim in a confidence argument depends on the confidence "scores" of the sub-arguments. Figure 12 represents our current thoughts on the calculus of confidence, that is, we simply combine the defeaters and determine whether they have been eliminated up the argument tree. In the more general case, where the defeaters for *different* system claims may be identical, duplicate defeaters will need to be eliminated from the higher level Baconian probability.

Suppose we are able to assert a confidence B(C2) = *i/m* and B(C3) = *j/n* as shown in Figure 13. That means that *i* out of *m* defeaters have been eliminated for one subclaim and *j* out of *n* have been eliminated for the other. In addition, there might be defeaters that apply just to C1; in particular, there will always be at least one undercutting defeater for the inference rule connecting C1 with C2 and C3. To determine the confidence associated with C1, we need to consider the union of all defeaters that apply to the subclaims of C1 plus any new defeaters applicable only to C1.

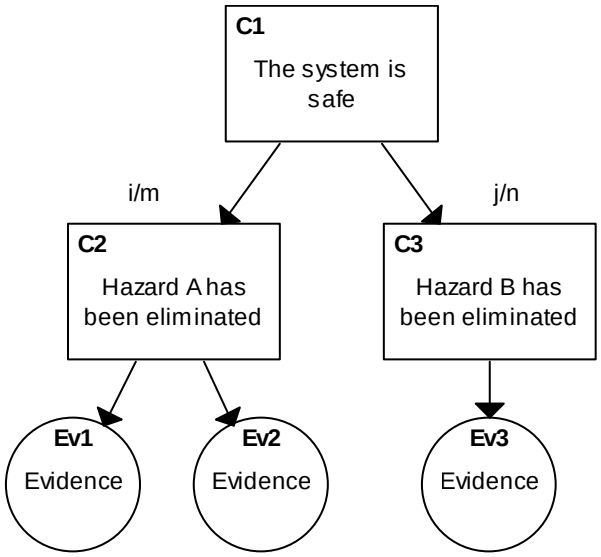

Figure 13: A Simple Argument with Confidence Numbers

If no defeaters are shared between C2 and C3 (that is, if the defeaters are independent), then we can assert a confidence B(C1) = $(i+j+k)/(m+n+p)$ for the parent argument, where *p* represents the number of unique defeaters associated with C1 and *k* is the number of these that have been eliminated. If all defeaters for one subclaim are shared by the other, then there are only max (*m*,*n*) unique defeaters for the parent argument, and the number of shared eliminated defeaters is max (*i*,*j*). Thus we have B(C1) ranging from $(\max(i,j)+k)/(\max(m,n)+p)$ to $(i+j+k)/(m+n+p)$. A closer study of the specific eliminated defeaters for each sub-argument can narrow the range.

Not all defeaters are equally likely. For example, the defeater "Harry's parents are aliens" may be more likely than "Harry has renounced his British citizenship." So far, our confidence numbers have been calculated by simply counting the number of defeaters, but the number really should account for relative defeater likelihood or importance. This is an issue we are studying, and it will be the subject of a future report.

5.2 What if a Defeater Cannot Be Eliminated?

We know that in the course of gathering evidence to support a system claim, we sometimes find that a claim may not be true under all circumstances. For example, suppose we find that a system occasionally fails at low temperatures. This means that a broad claim, such as "The system works reliably," is no longer valid. How should this be handled?

There are really only two choices: fix the system so the system works at low temperatures, or change the claim so the failure is irrelevant. For example, we could tell the user not to use the system when the temperature is below freezing. This means we have modified the general claim to take into account the test failure, for example, the claim may now read "The system works reliably as long as it is not used when temperatures are below freezing." Now the failed test is irrelevant to the modified claim.

5.3 Probability as a Measure of Confidence

As mentioned in Section 2, Bloomfield and others have suggested that confidence should be a measure of how likely it is (in a Pascalian sense) that a claim is true. We noted in discussing our definition of confidence (in Section 4.1) that our view of confidence is different.

The difference can be easily illustrated using our basic block example. The likelihood that a basic block will contain an egregious error is quite small, say, one in a thousand (0.001). In a program with four basic blocks, the likelihood of no egregious error is therefore $0.999^4 = 0.996$. So, without running any tests at all, we are pretty "confident" that there are no egregious errors. After running our single test, the likelihood of no egregious error rises to 0.999, that is, the growth in confidence does not seem very great. Using Baconian probabilities, our confidence with respect to just the rebutting defeaters has risen from 0/4 (no information) to 3/4 (3 reasons for doubt removed). To us, characterizing confidence in terms of doubt elimination gives a more insightful view of why we should feel more confidence in the claim after providing certain kinds of evidence.

Nonetheless, when considering top-level claims such as "The system is safe," we would like to know how likely it is that the claim is true. In such a case, we would reword the claim to incorporate the desired degree of certainty, for example, "The system is safe with 99.99% probability." The associated confidence case would present reasons for doubting the truth of such a claim and why we believe these reasons have been eliminated, arriving at a Baconian measure of confidence in the claim.

Statistical testing may be used to support a claim that the probability of failure on demand is less than some desired value. A defeater for a statistical testing argument would be something like "More than one failure has been observed in 15,000 tests." Clearly such a defeater is eliminated by showing that at least 15,000 tests have been successful. Note that in our framework, running

30,000 successful tests would not increase our Baconian confidence because once a defeater has been eliminated, no additional evidence is necessary.

5.4 Confidence and Argument Structure

The number of defeaters applicable to an argument depends on the structure and claims used in the argument. For instance, in our simple program example, we argued that there were no egregious errors in a program because there were no egregious errors in any basic block. We could have argued equally well that there are no egregious errors in the program because there are no egregious errors in any execution *path* through the program. For most programs, there are, of course, far more paths than basic blocks, so this might not be a sensible argumentation approach. However, in our simple example, there are only two paths through the program. Consequently, there are only two rebutting defeaters for a claim that "No program path contains an egregious error." Our single test of path $1 \rightarrow 2 \rightarrow 4$ serves to eliminate one of the two defeaters. So after running this test, our confidence (with respect to rebutting defeaters) is 1/2, whereas when the argument was phrased in terms of basic blocks, our confidence was 3/4.

This example shows why Baconian probabilities cannot be compared directly and the care that must be taken when interpreting their meaning. Whether our confidence is expressed as 1/2 or 3/4, there is still one remaining source of doubt to be considered. We cannot suppose that the argument structure and evidence yielding a confidence of 3/4 is somehow better or more stringent than the structure yielding a confidence of 1/2. Baconian probabilities cannot be used for such comparisons, since different supporting arguments will have different defeaters and, consequently, different confidence values for a given set of evidence.

5.5 Other Issues

Much work remains to demonstrate the usefulness of our framework in practice and its usefulness in explaining how confidence grows in certain circumstances. Moreover, the concepts need to be tested with practical examples to see how easy it is to identify and eliminate defeaters, whether the ideas scale, and other adoption issues. Nonetheless, thinking about confidence in terms of defeaters and eliminative induction does give a good basis for understanding why certain evidence increases our confidence in the truth of a claim.

6 Summary

This report has presented a framework for understanding and measuring confidence in assurance cases. We have drawn on the principles of eliminative deduction and defeasible reasoning, and adapted them to the more constrained domain of assessing confidence in assurance cases.

In the Introduction, we presented a skeletal assurance case and asked the following questions:

> *How confident should we be that the system is actually safe, and what is the basis for this confidence? What does it mean to say we have some degree of confidence? Is it a measure of how likely the claim is to be true? Or is it a measure of how justified we are in believing the claim to be true, and if so, what is the basis for justification? If we need to improve our confidence in any of these claims, what should be done and why?*

Our framework provides answers to these questions. Having some degree of confidence in a claim means having some degree of belief in the truth of the claim. The basis for our degree of belief is determined by how many reasons for doubt (defeaters) we have identified and how many we have eliminated. If we want to increase confidence, we need to eliminate more defeaters.

We have only started development of the framework. Further examples and application are needed to determine how helpful it will be in practical cases. However, we have found it to be a helpful way of thinking about assurance confidence and expect to develop it further.

Appendix A Goal Structuring Notation

This appendix contains basic information regarding the goal-structured assurance case and the Goal Structuring Notation (GSN) developed by Tim Kelly and his colleagues at the University of York in the United Kingdom [Kelly 1998]. A goal-structured assurance case specifies a claim regarding a property of interest, evidence that supports that claim, and a detailed argument explaining how the evidence supports the claim.

In the case illustrated in Figure 13, the top-level claim is "The top-level claim is true." From that claim flows an argument that supports the top-level claim. The argument consists of one or more subsidiary claims that, taken together, make the top-level claim believable. These lower level claims are themselves supported by additional claims until finally a subclaim is to be believed because evidence exists that clearly shows the subclaim to be true.

Claims (C1 through C4 below) are phrased as predicates; they are either true or false. Evidence nodes (Ev1 through Ev3 below) are stated as noun phrases. Other elements shown in the sample are

- the diamond under "C3," which indicates that the claim requires further development
- the triangle under evidence "Ev3," which indicates that the evidence is parameterized and needs to be instantiated in an actual case
- the diamond within the link between the claim "C2" and evidence "Ev1" and "Ev2," which indicates (as labeled) that either "Ev1" or "Ev2" applies or both do
- the parallelogram labeled "S1," which contains an explanation of the strategy used to structure the argument
- a rounded rectangle labeled "Cx1," which provides context information
- an oval with an "A" under it labeled "A1," which is a stated assumption

The last two elements provide explanatory information about the claim to which they are attached.

In this report we do not use all the elements shown in Figure 13, and these are not the only elements of a goal-structured assurance case. We refer the interested reader to Kelly's GSN description for a more complete treatment [Kelly 1998].

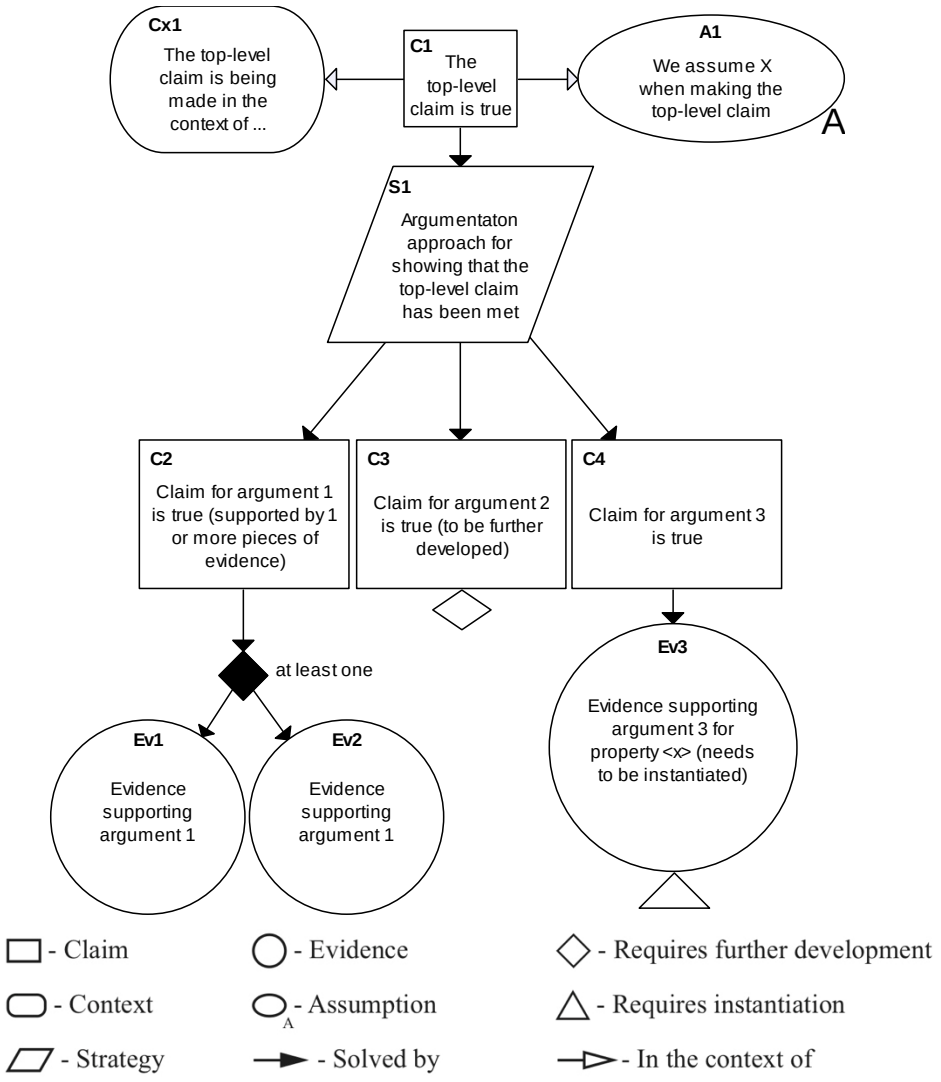

Figure 14: Goal Structuring Notation

Appendix B Glossary

argument
instantiations of one or more inference rules with specific premises and conclusions. In an assurance case, conclusions are claims and premises are either evidence or other claims.

assurance case
a structured set of arguments and informative elements showing how the truth of a claim is justified by evidence or subordinate arguments. An assurance case provides defeasible reasons for believing that a claim is true.

Baconian probability
a system of probabilities that range from total doubt (no reasons for doubt eliminated) to complete acceptance (all identified reasons for doubt eliminated)

claim
an assertion. In assurance cases claims are statements that are asserted to be true. In a goal-structured assurance case, claims are stated as predicates, that is, statements that are true or false (e.g., “The system is safe.”)

conclusion
an assertion that follows from specific premises and the application of an inference rule. In an assurance case, claims are the conclusions of supporting arguments. More generally, a conclusion is the consequent of an inference rule.

confidence (in the truth of a claim)
belief (in the truth of a claim). Increased confidence in a claim means increased belief in the truth of the claim. In our framework, confidence increases only by eliminating defeaters associated with an argument.

confidence case
an assurance case that provides justification for belief in a system case claim by identifying and eliminating defeaters

defeasible reasoning
a form of reasoning where the corresponding argument is rationally compelling but not deductively valid. The tentative relationship of support between premises and conclusion is potentially defeated by additional information.

defeater
a source of doubt about the truth of a claim

eliminative induction
identification of possibilities for doubting the truth of a hypothesis, followed by the gathering of evidence or performance of analyses that eliminate each of these possibilities. Each eliminated possibility removes a reason for doubt and thereby increases our confidence in the hypothesis.

eliminating a defeater
A defeater is eliminated by an argument showing that a defeater is false.

evidence
data. In a goal-structured assurance case, evidence is stated as a noun phrase (e.g., “Test results”).

goal structuring notation (GSN)
a notation for graphically presenting assurance cases

hypothesis
a top-level claim. Philosophical discussions are often concerned with how to decide which of several competing hypotheses is best supported by available evidence, but in an assurance case, we are only concerned with deciding whether the top-level claim is adequately supported by an argument that is linked to evidence. We use the term “hypothesis” when alluding to the philosophical basis for confidence.

inference rule
a generalization that is considered to be true under most conditions (e.g., “All persons born in Bermuda are British citizens.”) An inference rule is instantiated in an argument by applying it to specific premises.

Pascalian probability
a system of probabilities that range from disproof (an event is impossible) to proof (an event is certain). This is the “traditional” view of probability.

premise
in an assurance case, either a subclaim or evidence in an instantiated inference rule. More generally, the antecedent of an inference rule.

rebutting defeater
a defeater that eliminates belief in a claim by providing information that contradicts the claim

safety case
a system case for safety

system case
an assurance case whose top-level claim is about a system’s properties of interest

undercutting defeater
a defeater falsifying an inference rule, that is, a defeater for an inference rule that specifies conditions under which a claim is not necessarily true even if the premises are true

undermining defeater
in our framework, information that invalidates one or more pieces of evidence

References

URLs are valid as of the publication date of this document.

[Bloomfield 2003]
Bloomfield, R. & Littlewood, B. "Multi-Legged Arguments: The Impact of Diversity upon Confidence in Dependability Arguments," 25-34. *Proceedings of the International Conference on Dependable Systems and Networks (DSN 2003)*. San Francisco, CA, June 2003. IEEE, 2003.

[Bloomfield 2007]
Bloomfield, R., Littlewood, B., & Wright, D. "Confidence: Its Role in Dependability Cases for Risk Assessment," 338-346. *Proceedings of the International Conference on Dependable Systems and Networks (DSN 2007).* Edinburg, U.K., April 2007. IEEE, 2007.

[Bloomfield 2010]
Bloomfield, R. & Bishop, P. "Safety and Assurance Cases: Past, Present and Possible Future – an Adelard Perspective," 51-67. *Making Systems Safer*: *Proceedings of the Eighteenth Safety-Critical Systems Symposium.* Bristol, U.K., February 2010. C. Dale and T. Anderson (eds.), Springer-Verlag, 2010.

[Cohen 1970]
Cohen, J. *The Implications of Induction*. Methuen, 1970

[Cohen 1977]
Cohen, J. *The Probable and the Provable.* Clarendon, 1977.

[Cohen 1989]
Cohen, J. *An Introduction to the Philosophy of Induction and Probability.* Clarendon, 1989.

[Hawkins 2011]
Hawkins, R., Kelly, T., Knight, J., & Graydon, P. "A New Approach to Creating Clear Safety Arguments"185-197. *Advances in Systems Safety*: *Proceedings of the Eighteenth Safety-Critical Systems Symposium.* Southampton, U.K., February 2011. C. Dale and T. Anderson (eds.), Springer-Verlag, 2011.

[Kelly 1998]
Kelly, T. "Arguing Safety – A Systematic Approach to Safety Case Management." PhD diss., University of York, Department of Computer Science, 1998.

[Littlewood 2007]
Littlewood B. & Wright, D. "The Use of Multilegged Arguments to Increase Confidence in Safety Claims for Software-Based Systems: A Study Based on a BBN Analysis of an Idealized Example." *IEEE Transactions on Software Engineering 33*, 5 (May 2007): 347-365.

[MRL 2005]
Metaphysics Research Laboratory, Center for the Study of Language and Information, Stanford University. *Stanford Encyclopedia of Philosophy: Defeasible Reasoning*, 2005. http://plato.stanford.edu/entries/reasoning-defeasible/

[Pollock 2008]
Pollock, J. "Defeasible Reasoning." *Reasoning: Studies of Human Inference and Its Foundations,* J.E. Adler and L.J. Rips (eds.), 451-469. Cambridge University Press, 2008.

[Prakken 2010]
Prakken, H. "An Abstract Framework for Argumentation with Structured Arguments." *Argument & Computation 1*, 2 (2010): 93-124.

[Schum 2001]
Schum, D. *The Evidential Foundations of Probabilistic Reasoning.* Northwestern University Press, 2001.

[Toulmin 1958]
Toulmin, S. *The Uses of Argument*. Cambridge University Press, 1958.

[Wassyng 2011]
Wassyng, A., Maibaum, T., Lawford, M., & Bherer, H. "Software Certification: Is There a Case Against Safety Cases," 206-227. *Monterey Workshops 2010, Lecture Notes in Computer Science 6662.* R. Calinescu and E. Jackson (eds.), Springer-Verlag, 2011.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
(Leave Blank)	September 2012	Final

4. TITLE AND SUBTITLE	5. FUNDING NUMBERS
Toward a Theory of Assurance Case Confidence	FA8721-05-C-0003

6. AUTHOR(S)

John B. Goodenough, Charles B. Weinstock, Ari Z. Klein

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
Software Engineering Institute Carnegie Mellon University Pittsburgh, PA 15213	CMU/SEI-2012-TR-002

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFLCMC/PZE 20 Schilling Circle, Bldg 1305, 3rd floor Hanscom AFB, MA 01731-2125	ESC-TR-2012-002

11. SUPPLEMENTARY NOTES

12A DISTRIBUTION/AVAILABILITY STATEMENT	12B DISTRIBUTION CODE
Unclassified/Unlimited, DTIC, NTIS	

13. ABSTRACT (MAXIMUM 200 WORDS)

Assurance cases provide an argument and evidence explaining why a claim about some system property holds. This report outlines a framework for justifying confidence in the truth of such an assurance case claim. The framework is based on the notion of *eliminative induction*—the principle (first put forward by Francis Bacon) that confidence in the truth of a hypothesis (or claim) increases as reasons for doubting its truth are identified and eliminated. Possible reasons for doubting the truth of a claim (*defeaters*) arise from analyzing an assurance case using defeasible reasoning concepts. Finally, the notion of Baconian probability provides a measure of confidence based on how many defeaters have been identified and eliminated.

14. SUBJECT TERMS	15. NUMBER OF PAGES
assurance case, defeasible reasoning, eliminative induction, confidence, argumentation	41

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL