Software Engineering Institute
Carnegie Mellon University

Eliminative Argumentation: A Basis for Arguing Confidence in System Properties

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

February 2015

TECHNICAL REPORT
CMU/SEI-2015-TR-005

Software Solutions Division

http://www.sei.cmu.edu

Copyright 2015 Carnegie Mellon University

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.

Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the United States Department of Defense.

This report was prepared for the
SEI Administrative Agent
AFLCMC/PZM
20 Schilling Circle, Bldg. 1305, 3rd floor
Hanscom AFB, MA 01731-2125

NO WARRANTY. THIS CARNEGIE MELLON UNIVERSITY AND SOFTWARE ENGINEERING INSTITUTE MATERIAL IS FURNISHED ON AN “AS-IS” BASIS. CARNEGIE MELLON UNIVERSITY MAKES NO WARRANTIES OF ANY KIND, EITHER EXPRESSED OR IMPLIED, AS TO ANY MATTER INCLUDING, BUT NOT LIMITED TO, WARRANTY OF FITNESS FOR PURPOSE OR MERCHANTABILITY, EXCLUSIVITY, OR RESULTS OBTAINED FROM USE OF THE MATERIAL. CARNEGIE MELLON UNIVERSITY DOES NOT MAKE ANY WARRANTY OF ANY KIND WITH RESPECT TO FREEDOM FROM PATENT, TRADEMARK, OR COPYRIGHT INFRINGEMENT.

This material has been approved for public release and unlimited distribution except as restricted below.

Internal use:* Permission to reproduce this material and to prepare derivative works from this material for internal use is granted, provided the copyright and “No Warranty” statements are included with all reproductions and derivative works.

External use:* This material may be reproduced in its entirety, without modification, and freely distributed in written or electronic form without requesting formal permission. Permission is required for any other external and/or commercial use. Requests for permission should be directed to the Software Engineering Institute at permission@sei.cmu.edu.

* These restrictions do not apply to U.S. government entities.

Carnegie Mellon® is registered in the U.S. Patent and Trademark Office by Carnegie Mellon University.

DM-0001494

Table of Contents

Acknowledgments **vii**

Abstract **ix**

1 Introduction **1**

2 Eliminative Argumentation **3**

2.1 Induction 3
2.2 Confidence Maps and Eliminative Argumentation 6
2.3 Evaluating Confidence 12
2.3.1 Baconian Confidence 13
2.3.2 Probabilistic Confidence 14
2.3.3 Other Ways to Calculate Confidence 15
2.4 Rules for Composing Eliminative Arguments and Confidence Maps 15
2.4.1 Fundamental Elements of an Eliminative Argument 15
2.4.2 Elements of a Confidence Map 16
2.4.3 Rules for Well-Formed Elements of an Argument 16
2.4.4 Summary 21
2.5 A More Realistic Example 21
2.5.1 The Statistical Testing Argument 22
2.5.2 The Static Analysis Argument 28
2.5.3 Extending the Statistical Argument with a Static Analysis Argument 32
2.6 Summary 35

3 Issues and Concerns **38**

3.1 Unidentified Defeaters 38
3.2 Relative Importance of Defeaters 38
3.3 Eliminative Argumentation for Real Systems 39
3.4 Why Use Eliminative Induction? 39
3.5 True Defeaters (Counterevidence) 40
3.5.1 Making Counterevidence Irrelevant 41
3.5.2 Accepting Counterevidence 41

4 Connections to Other Work **43**

4.1 Comparison with Assurance Case Concepts 43
4.1.1 Claims 43
4.1.2 Context 43
4.1.3 Evidence 44
4.1.4 Inference Rule 45
4.1.5 Assumption 45
4.1.6 Rebutting Defeaters 46
4.1.7 Undercutting Defeaters 46
4.1.8 Undermining Defeaters 46
4.2 Eliminative Argumentation and Notions of Assurance Case Confidence 46
4.3 Argumentation Literature 47
4.3.1 Defeasible Reasoning 47
4.3.2 Convergent and Linked Argument Structures 48
4.4 Philosophy 52

5 Conclusion **53**

Appendix Calculating the Probability of Failure **54**

References **55**

List of Figures

Figure 1:	Notional Assurance Case	1
Figure 2:	Enumerative Induction	3
Figure 3:	Eliminative Induction	4
Figure 4:	Partial Confidence	5
Figure 5:	Assurance Case for “Light Turns On”	6
Figure 6:	Confidence Map for "Light Turns On"	8
Figure 7:	Complete Visualization of the Inference Rule for an Undercutting Defeater	10
Figure 8:	Complete Visualization for Defeaters and Evidence	11
Figure 9:	Confidence Evaluation of Defeater R2.1	12
Figure 10:	Confidence Evaluation for Top-Level Claim	13
Figure 11:	Probabilistic Confidence Assessment	14
Figure 12:	An Assurance Case Supported by Statistical Testing Evidence	22
Figure 13:	Top-Level Rebutting Defeater	23
Figure 14:	Analyzing an Imperfect Inference Rule	24
Figure 15:	An Alternative Argument Eliminating R2.1	25
Figure 16:	Splitting a Defeater	25
Figure 17:	An Example of Defeater Refinement	26
Figure 18:	Reasoning with Undermining Defeaters	27
Figure 19:	Implied Inference Rule with Undermining Defeaters	27
Figure 20:	A Confidence Map for Static Analysis Evidence	29
Figure 21:	A More Complete Argument Involving Static Analysis	30
Figure 22:	Defeaters for Static Analysis Evidence	31
Figure 23:	A Multi-legged Assurance Case	33
Figure 24:	Multi-legged Confidence Map	36
Figure 25:	Using Static Analysis Evidence in the Statistical Testing Leg	37
Figure 26:	A Convergent Argument Structure	50
Figure 27:	A Linked Argument Structure	51

List of Tables

Table 1: Types of Defeaters 7

Table 2: Confidence Map Symbols 7

Table 3: Upper Bound on pfd for Successful Test Runs 54

Acknowledgments

We are grateful to Robert Stoddard for developing Table 3, which gives upper bounds on the probability of failure on demand when a certain number of operationally random tests have executed successfully. We are also grateful to John Hudak, Ipek Ozkaya, and Paul Jones for helpful comments on earlier drafts.

Abstract

Assurance cases provide a structured method of explaining why a system has some desired property, for example, that the system is safe. But there is no agreed approach for explaining what degree of confidence one should have in the conclusions of such a case. This report defines a new concept, *eliminative argumentation*, that provides a philosophically grounded basis for assessing how much confidence one should have in an assurance case argument. This report will be of interest mainly to those familiar with assurance case concepts and who want to know why one argument rather than another provides more confidence in a claim. The report is also potentially of value to those interested more generally in argumentation theory.

1 Introduction

An assurance case is a structured argument showing why a claim is believed to be true given certain evidence. By providing explicit claims and reasoning for why evidence is believed to support the claims, an assurance case makes explicit the reasoning that is otherwise often implicit in arguments intended to show that a system is acceptably safe or secure [Kelly 1998]. Assurance cases are being used increasingly today to justify claims about properties of systems [GSN 2011, Hawkins 2013, ISO/IEC 2011, OMG 2013].

Although an assurance case is intended to provide a basis for deciding whether a system is acceptably safe or secure, what makes one case more convincing than another? Is there some conceptual basis for determining how much assurance a particular case provides? Is there a philosophical foundation for explaining why we should have a certain level of confidence in an assurance case claim? The *eliminative argumentation* formulation described in this report provides an answer to these questions.

Consider the notional assurance case shown in Figure 1.[1] This example can be read as saying, "The system is acceptably safe *because* Hazards A and B have been adequately mitigated. Moreover, evidence Ev1 and Ev2 show that Hazard A has been mitigated, and similarly, evidence Ev3 shows that Hazard B has been mitigated." A real case is more complicated, of course,[2] but how should we assess the adequacy of this particular argument?

- How confident should we be in claim C1? Why should we be confident?
- What does it mean to have confidence in the claim?
- What could be done to improve confidence?

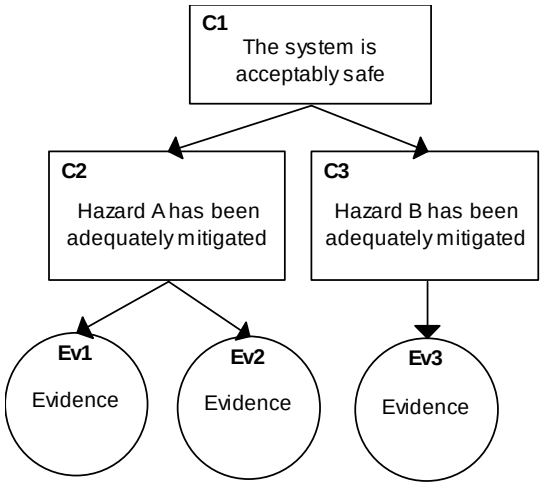


Figure 1: Notional Assurance Case

[1] This example uses Goal Structuring Notation (GSN) [GSN 2011]. Claims are enclosed in rectangles, evidence in a circle.

[2] In particular, there should be significantly more argumentation showing why the evidence implies that some hazard has been adequately mitigated (or eliminated).

How evidence provides confidence in a hypothesis is a long-standing philosophical and practical problem. Sir Francis Bacon first suggested in 1620 that confidence in a hypothesis (or claim) increases as reasons for doubting its truth are identified and eliminated [Schum 2001]. This basis for evaluating confidence in a claim is called *eliminative induction.* In this report, we show how eliminative induction can be a useful basis for justifying belief in a claim.

In an earlier report, we explored the notion of eliminative induction as the basis for assurance [Goodenough 2012]. The current report supersedes our earlier report by providing a revised notation, explicit rules for using the notation correctly, a deeper discussion of basic concepts, and a more realistic example of how our approach could be used. Because we are focused on the use of eliminative induction in developing and assessing an argument, we now call our approach *eliminative argumentation*.

An eliminative argument is visualized in a *confidence map*, a graphical structure that explicitly shows reasons for doubting the validity of the claims, evidence, and reasoning comprising an argument. In particular, the map shows why these doubts are either eliminated or remain as reasons for reduced confidence.

We introduce eliminative argumentation concepts and notation in the next section together with examples showing how to develop and evaluate an eliminative argument. In Section 3, we discuss various concerns and objections to our approach. Section 4 contains a detailed discussion of how our approach is related to assurance case concepts as well as concepts drawn from argumentation literature.

2 Eliminative Argumentation

In this section, we introduce the fundamental concepts of eliminative argumentation, introduce our method of visualizing an eliminative argument (confidence maps), and discuss two examples to illustrate the issues and benefits of our approach.

In Section 2.1, we discuss two fundamentally different ways of developing confidence in system properties: enumerative induction and eliminative induction. In Section 2.2, we introduce confidence map notation and illustrate the principles of eliminative argumentation with a simple, rather artificial, example. In Section 2.3, we discuss how a confidence map can be used to evaluate confidence in a claim, and in particular, we discuss how to evaluate confidence when our belief in the elimination of a doubt is a judgment expressed as a probability. In Section 2.4, we summarize and formalize the eliminative argumentation concepts and confidence map notations introduced in prior sections. In Section 2.5, we develop a more realistic confidence map to explain how statistical testing and static analysis could be combined to increase confidence in a claim of system reliability.

2.1 Induction

In this section, we introduce and illustrate fundamental concepts of eliminative argumentation using a simple example of a system for turning on a light. The assurance questions are "how confident are we that a light will turn on when a switch is clicked?" and "what is the basis for our confidence?"

In a deductive argument, the truth of the premises *guarantees* the truth of the conclusion. In an inductive argument, the truth of the premises only *indicates*, with some degree of strength, that the conclusion is true [Hawthorne 2012]. We use inductive arguments all the time to justify confidence in system properties. For example, as more tests run correctly, we become more confident that a system has the properties we desire.

In evaluating systems, two kinds of induction are commonly used: *enumerative* induction and *eliminative* induction. The difference is illustrated by the example shown in Figures 2 and 3.

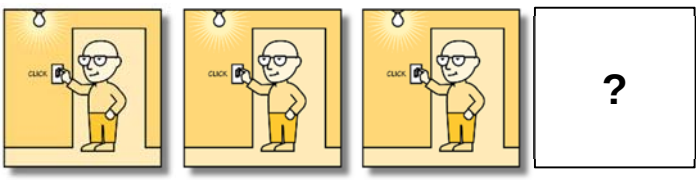


Figure 2: Enumerative Induction

When we go into a room, such as an office, and click a light switch, we expect a light to go on. What is the basis for our confidence? One basis could be past experience. If we have gone into this room many times in the past, and every time we clicked the switch, the light turned on (see Figure 2), then when we go into the room tomorrow and use the switch, we will not be surprised when the light turns on (that is, we will be very confident that the light will turn on). If asked why

we are not surprised, we might say that the basis for our confidence is our past experience of success in turning on the light. This is an example of confidence based on enumerative induction. In enumerative induction, confidence increases as confirming examples are found.

In eliminative induction (Figure 3), confidence increases as reasons for doubt are eliminated. In the lighting example, let's consider reasons for doubting that the light would go on. For example, we might say, "The light will not go on if the light bulb is defective." Or, "The light will not go on if there is no power at the switch." Or, "The light will not go on if the switch is not wired to the light."

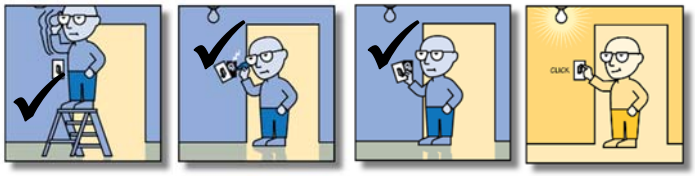


Figure 3: Eliminative Induction

In eliminative induction, we investigate each of these reasons for doubt (before clicking the switch) to see if the doubt can be eliminated. As we eliminate these doubts, our confidence that the light will go on increases (see Figure 3). If these were indeed the only reasons the light would not turn on and if we have indeed correctly eliminated them, then we would have no basis for doubting that the light will turn on. In such a case, we could say we have *complete* confidence that the light will turn on. But such confidence is, as we said, contingent on having confidence that we have identified all reasons for failure and that our evidence for their elimination is reliable. These concerns are doubts that reduce our confidence that the light will turn on even when all three identified sources of failure are thought to be eliminated, but these are doubts about the validity of the *reasoning* and *evidence* we are using to support our claim rather than doubts about why the system can fail. These doubts also need to be eliminated if we are to have "complete" eliminative confidence in the claim.

This simple example raises some of the fundamental concerns one might have about using eliminative induction as a basis for confidence:

- *Is complete confidence (in the sense of eliminating all doubts) possible for real-world systems?* In practice, there is always some uneliminated (residual) doubt in an argument. Deciding its importance to confidence in a claim is a matter we will discuss as an aspect of our argumentation approach.
- *Can all doubts ever be identified?* Since we wish to deal with real-world systems, it is always possible that new doubts can be suggested or discovered. In practice, overlooking some doubts is inevitable; our method doesn't guarantee that a complete argument will be developed—it just makes it less likely that argument deficiencies are overlooked. Our eliminative argumentation approach specifically addresses the possibility of overlooking some doubts.
- *How are doubts identified?* Our approach says nothing about how to identify doubts such as those shown in Figure 3. Our approach provides a basis for evaluating the amount of confidence one has in a claim given that certain doubts have been identified and not fully eliminated.

- *Are all doubts of equal importance?* The approach does not require that doubts be of equal importance, although for purposes of explaining the basic concepts, it is convenient to give equal weight to doubts.

Given that eliminating all doubt is, for practical purposes, impossible, what is the point of the eliminative approach? The point is to identify both sources of doubt and the arguments for eliminating them. This approach provides a concrete, reviewable form showing the extent to which it is reasonable to have a certain level of confidence in a claim.

Returning to the example, if we have eliminated no doubts, we will have no confidence that the light will turn on. Note that "no confidence" does not mean the light won't turn on—it means only that we don't have enough information to support a justified belief that the light will turn on.

When we have eliminated m out of n doubts, we write $m|n$ as a measure of confidence.[3] For Figure 3, we could write 3|3 to show that three out of three reasons for doubt have been eliminated. If we eliminate only some reasons for doubt (e.g., 1|3 as shown in Figure 4), we have *partial* confidence in a claim. Alternatively, we say that there are two *residual* doubts remaining. In general, there are $n - m$ residual doubts when only m out of n doubts have been eliminated.

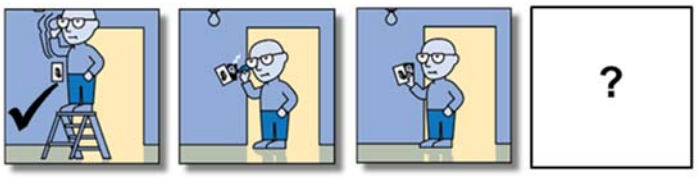


Figure 4: Partial Confidence

The example with the light switch is, of course, trivial. But when we consider a system that launches a missile, for example, we want to be as sure as reasonably possible that the system will work before we actually try it. In such a case, we look for possible reasons the system might not work (doubts) and develop evidence and arguments that eliminate (insofar as possible) these doubts.

We use eliminative and enumerative induction all the time to support our beliefs in system behavior. Each has an appropriate role. In particular, enumerative induction (in the form of Bayesian analysis [Cowles 2009]) is more powerful in many applications. But the eliminative approach is quite natural in some situations, for example, when seeking to develop confidence in the safety, security, or critical functionality of systems before they are put into service. In such analyses, it is natural to consider possible causes or symptoms of failure and then develop evidence and arguments intended to show that such failures cannot occur (or at least are very unlikely). Such an analysis is a form of eliminative induction. We seek to put this informal approach to use developing confidence on a more formal basis. We seek to show that the eliminative approach can be quite helpful as a way of thinking about why one should have confidence in system properties. Nonetheless, we will also show how some kinds of enumerative induction can complement our eliminative approach.

[3] The notation is an adaptation of notation proposed by L. Jonathan Cohen [Cohen 1989]. We discuss Cohen's influence on our work in Section 4.3.

In the introduction to this report, we asked questions about the assurance case in Figure 1. We now can provide some answers based on the eliminative induction approach:

- How confident should we be in claim C1? Why? *Answer:* We are confident to the extent that relevant doubts have been identified and eliminated. (A doubt about a claim, for example, is relevant if its truth means the claim is false. We consider only relevant doubts since eliminating an irrelevant doubt cannot, of course, be a basis for increased confidence.)
- What does it mean to have confidence in the claim? *Answer:* Lack of doubt. More specifically, confidence is our degree of belief in a claim based on lack of doubt.
- What could be done to improve confidence? *Answer:* Eliminate more doubts.

We will elaborate on these answers in the remainder of this report.

2.2 Confidence Maps and Eliminative Argumentation

Let's consider a possible assurance case for the lighting example (see Figure 5).[4] This case argues that the light turns on because the light bulb is functional, power is available at the switch, and the switch is connected to the light. In addition, we know the light bulb is functional because it doesn't rattle when it is shaken.[5] How much confidence should we have in the truth of claim C1.1?

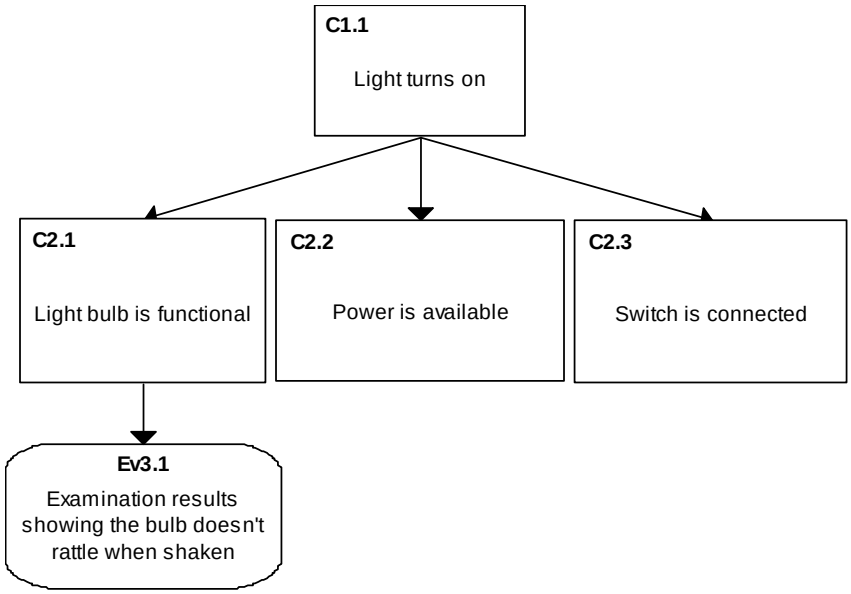


Figure 5: Assurance Case for "Light Turns On"

In this assurance case, each subclaim implicitly refutes a doubt about why the light might not turn on. In eliminative argumentation, we want to make doubts explicit, and these doubts include not

[4] The diagram uses a variant of GSN [GSN 2011]. Differences from GSN are small. In this example, the shape of the evidence node is different and we speak of claims and evidence rather than goals and solutions. Section 4.1 provides a complete discussion of differences.

[5] This test method is not necessarily valid; that is, we may have doubts about the relevance or reliability of such a test result. We deal with such doubts later as we develop an eliminative argument.

just doubts about why the light might not turn on but also doubts about the soundness of the evidence and the reasoning. In an eliminative argument, we make doubts explicit and provide arguments and evidence showing the extent to which doubts are eliminated.

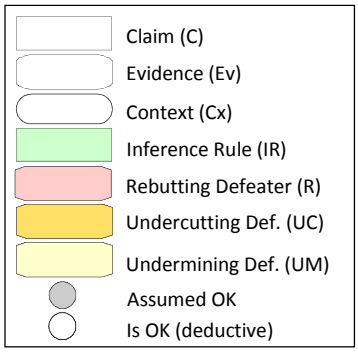
Table 1: Types of Defeaters

Defeater	Attacks	Meaning	Identification Process
Rebutting (R)	Claim	If R is true, the claim is false	Look for failure modes or possible counterexamples
Undermining (UM)	Evidence	If UM is true, the evidence is invalid	Look for reasons the evidence might be invalid
Undercutting (UC)	Inference rule	If UC is true, then the truth of the rule's premises is insufficient to imply the truth *or falsity* of the rule's conclusion	Look for conditions under which the inference rule is not adequate

An eliminative argument has five elements: claims, evidence, inference rules, argument terminators, and *defeaters*. Defeaters express doubts about the validity of the claims, evidence, and inference rules (see Table 1). There are (only) three kinds of defeaters, determined by the element to which the defeater applies: *rebutting* defeaters state reasons why a claim could be false, *undermining* defeaters state reasons why evidence could be invalid, and *undercutting* defeaters identify deficiencies in an inference rule such that the truth of the rule's premises is insufficient to determine whether its conclusion is true or not.[6] Undermining and undercutting defeaters express doubts about the soundness of the argument (in particular, the validity of evidence and inference rules).

We visualize an eliminative argument in a *confidence map*. A summary of confidence map notation is given in Table 2. The notation is explained in the next example.

Table 2: Confidence Map Symbols



A confidence map for the light-turns-on claim is shown in Figure 6. Figure 6 is structurally similar to Figure 5. The top-level claim (C1.1: "Light turns on") and the evidence (Ev3.1: "Examination results showing the bulb doesn't rattle when shaken") are the same. Rebutting defeaters R2.1, R2.2, and R2.3 (expressing reasons why C1.1 would not be true) are just the inverse of assurance case claims C2.1, C2.2, and C2.3 for this example. However, as compared to the assurance case, the confidence map explicitly specifies the inference rules used in the argument (the green rectangles labeled IR2.4 and IR3.2) so we can identify conditions under which the rules are *invalid*, that

[6] The terminology for defeaters is derived from the *defeasible reasoning* literature. We discuss these connections in Section 4.3.1.

is, insufficient to guarantee their conclusions when their premises are true. For example, inference rule IR3.2 ("If the bulb doesn't rattle when shaken, the bulb is not defective") is invalid if the bulb is not an incandescent bulb (e.g., if it is a LED bulb)—in such a case, the lack of a rattle would provide no information about whether the bulb was defective or not. This deficiency is captured in undercutting defeater UC4.2, "Unless the bulb is not an incandescent type." Similarly, if an incandescent bulb can fail without releasing a filament fragment (UC4.3), the lack of a rattle would also be uninformative. We need to show that these deficiencies in the rule are not relevant for the particular light bulb we are using by eliminating UC4.2 and UC4.3 (that is, by arguing that these defeaters are false). In this example we don't bother to provide evidence that the bulb is an incandescent bulb. We simply assert (with the shaded circle) that it is. Of course, a reviewer could challenge this assertion, in which case further argument and evidence would be have to be provided. But at some point, both reviewers and argument developers have to decide that no useful increase in confidence will be gained by providing additional evidence to eliminate some defeater; the shaded circle expresses this decision.

We leave UC4.3 ("Unless a bulb can fail without releasing a filament fragment") uneliminated; it is a source of residual doubt that the light will turn on. We will discuss in the next section (Section 2.3) how to evaluate the impact of this uneliminated defeater.

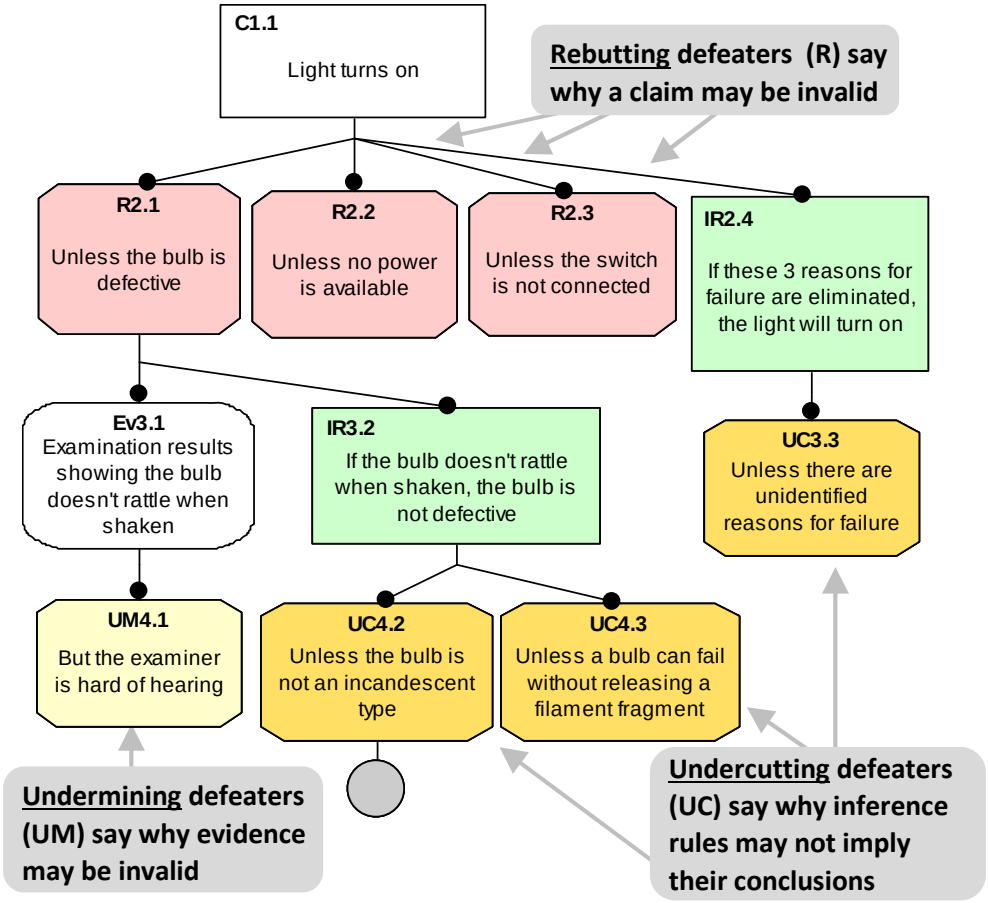


Figure 6: Confidence Map for "Light Turns On"

UC3.3 says that if we haven't identified all the reasons the light might not turn on, then eliminating the three identified reasons will not provide (complete) confidence that the light will turn on. To eliminate UC3.3, we would need to provide additional argument and evidence to show that these three reasons for failure are the only ones that need to be considered. For example, we could provide a hazard analysis, with further argument showing that the hazard analysis was carried out appropriately. However, the example argument does not provide such reasons, so UC3.3 is an uneliminated doubt, thereby reducing our confidence in the validity of the top-level claim.

Having considered doubts about the inferences in the example argument, we turn to the evidence—the results of shaking the bulb. The confidence map gives a reason for doubting the validity of the proffered evidence, namely, if the examiner shaking the light bulb is hard of hearing, a report stating that "No rattle was heard" is not worth much. No evidence is provided to eliminate this doubt, so it also remains as a reason for reduced confidence in claim C1.1.

It is common when justifying conclusions about the properties of a system to speak of evidence as *supporting* a claim. In eliminative induction, we never speak of evidence as directly supporting a claim; it supports a claim only to the extent that it eliminates doubts associated with the claim. This change in what constitutes "support" for a claim is the essence of the eliminative approach.

Before considering how the confidence map provides the basis for determining a degree of confidence in the top-level claim, we discuss the different types of defeaters in a little more detail and how an argument is terminated.

- *Rebutting* defeaters: These are doubts that contradict a claim. A rebutting defeater is a predicate. In a confidence map, the predicate is preceded with the word "Unless" for readability and to emphasize that the predicate is stating a doubt.
- *Undermining* defeaters: These are doubts about evidence. Various errors can be made when collecting and interpreting evidence. The impact of these errors is to cast doubt on the validity of the evidence. Invalid evidence is, of course, irrelevant to an argument. In our example, we looked for possible reasons why the report "Bulb doesn't rattle when shaken" might not be valid.

  An undermining defeater is a predicate. In a confidence map, the predicate is preceded with the word "But" to emphasize that the predicate is stating a doubt.
- *Undercutting* defeaters: These are doubts about an inference rule. An undercutting defeater is a predicate. In a confidence map, the predicate is preceded with the word "Unless" to emphasize that the predicate is stating a doubt.

  Undercutting defeaters are subtle. The truth of an undercutting defeater does not mean that a rule is wrong in the sense that it implies a false conclusion. An undercutting defeater instead points out conditions under which the rule does not necessarily apply and, hence, conditions under which its conclusion is not necessarily true *or* false. The rule (and its conclusion) can be trusted only if we show by further argument and evidence that the undercutting defeater does not apply for the system being considered.
- "Assumed OK" argument terminator: When the Assumed OK element is attached to a defeater, it asserts that elimination of the defeater is considered to be obvious—further evidence or argument is not needed.

In any argument, we can keep proposing new doubts, but in practice, we reach a point where positing a new doubt seems unproductive. The Assumed OK terminator symbol documents this decision. Someone reviewing the argument can always challenge the use of an Assumed OK terminator in a particular instance.

- "Is OK" argument terminator: When attached to an inference rule, it asserts that the rule has no undercutting defeaters because it is a tautology—its premise is deductively equivalent to its conclusion. We'll give an example below.

The confidence map in Figure 6 does not visualize a complete eliminative argument because there are some defeaters whose inference rules are still implicit. (Every defeater in an eliminative argument must be a premise of some inference rule since such an inference rule says why the elimination of certain defeaters increases confidence.) For example, consider IR2.4 ("If these 3 reasons for failure are eliminated, the light will turn on") and its undercutting defeater, UC3.3 ("Unless there are unidentified reasons for failure"). The implicit inference rule associated with UC3.3 is shown in Figure 7 as IR3.4: "If there are no unidentified reasons for failure, IR2.4 is valid." (An inference rule is valid if its premise always implies its conclusion.)

IR3.4 has no undercutting defeaters; it is a tautology because its conclusion (that IR2.4 is valid) is deductively true if its premise is true (i.e., if UC3.3 is eliminated). IR3.4 is said to be *indefeasible.* In a confidence map, we attach a clear circle to an inference rule when no undercutting defeaters are possible for the rule, that is, when no further information can ever invalidate the rule. In contrast, other inference rules may be invalidated by the discovery of additional information. Such rules are said to be *defeasible* [MRL 2005].

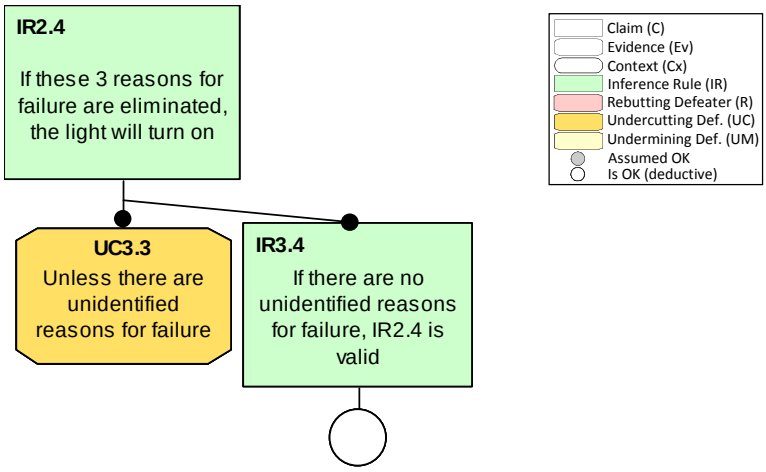


Figure 7: Complete Visualization of the Inference Rule for an Undercutting Defeater

Another implicit inference rule and argument is associated with UM4.1 ("But the examiner is hard of hearing"). The implicit argument in Figure 6 is that evidence Ev3.1 (the "bulb rattling" report) is valid if the examiner is *not* hard of hearing (i.e., if we eliminate UM4.1), but this inference rule associating UM4.1 with Ev3.1 is not shown in the diagram. Nor have we considered whether UM4.1 is the only reason for doubting the validity of the evidence.

Figure 8 shows the complete visualization of the eliminative argument for the validity of Ev3.1. In this figure, we have added two obvious inference rules and an undercutting defeater:

- IR4.2: “If the examiner is not hard of hearing, then the examination results are valid.”
- UC5.2: IR4.2 has an undercutting defeater (UC5.2) because its conclusion would be in doubt if there is some additional reason the examiner could be mistaken in reporting lack of a rattle. In our example, the argument asserts an assumption that UC5.2 is eliminated because UM4.1 is the only reason the evidence might not be valid.
- IR5.1: Since every defeater has an inference rule explaining its significance in the argument, IR5.1 is the inference rule using the negation of UC5.2 as its premise: “If there are no unidentified reasons for doubting the validity of the examination results, then IR4.2 is valid.” The argument then asserts that IR5.1 has no undercutting defeaters because it is the same kind of logical tautology as IR3.4: it says, in effect, “If there are no reasons for doubting the validity of IR4.2 (because UC5.2 is eliminated), then IR4.2 is valid, meaning that there are no reasons for doubting its validity.”

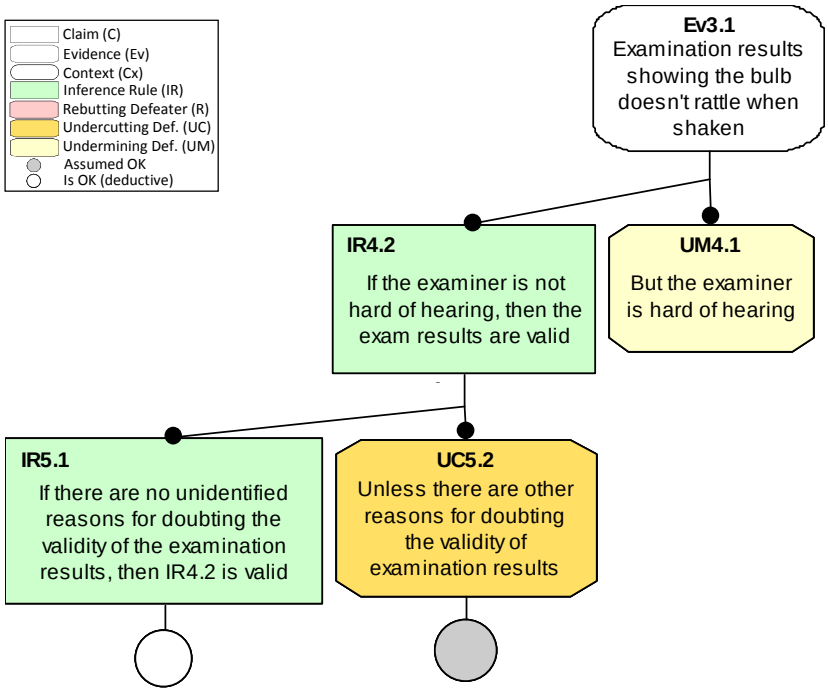


Figure 8: Complete Visualization for Defeaters and Evidence

The complete visualization of an eliminative argument is more complex than necessary for practical use. In Section 2.4, we define conventions for omitting some argument elements from a confidence map visualization. For example, the kinds of inference rules we introduced in Figure 8 are necessarily part of the eliminative argument even if they are not explicitly visualized in the confidence map.

To summarize, in this subsection, we introduced the basic concepts of eliminative argumentation (claims, evidence, inference rules, defeaters, and argument terminators) and showed how an eliminative argument can be visualized in a confidence map. In the next section, we discuss how to use a confidence map as the basis for calculating a confidence metric.

2.3 Evaluating Confidence

The foundational premise of eliminative argumentation is that confidence increases as reasons for doubt are eliminated. Our discussion in this section of possible ways of combining confidence assessments is intended to show, first, that a confidence map (and eliminative argumentation) provides the foundation for various ways of evaluating the significance of uneliminated doubts and, second, that useful information about confidence can be developed even when doubts can be eliminated only incompletely.

We first discuss *Baconian*[7] confidence in which defeaters at the leaves of the confidence map are assumed to be either fully eliminated or not. Then we consider a probabilistic approach in which we calculate confidence based on the probability that such defeaters are eliminated. Finally, we discuss various other approaches that could be taken to evaluate confidence and the additional research that is needed to understand which approach is most useful for different purposes.

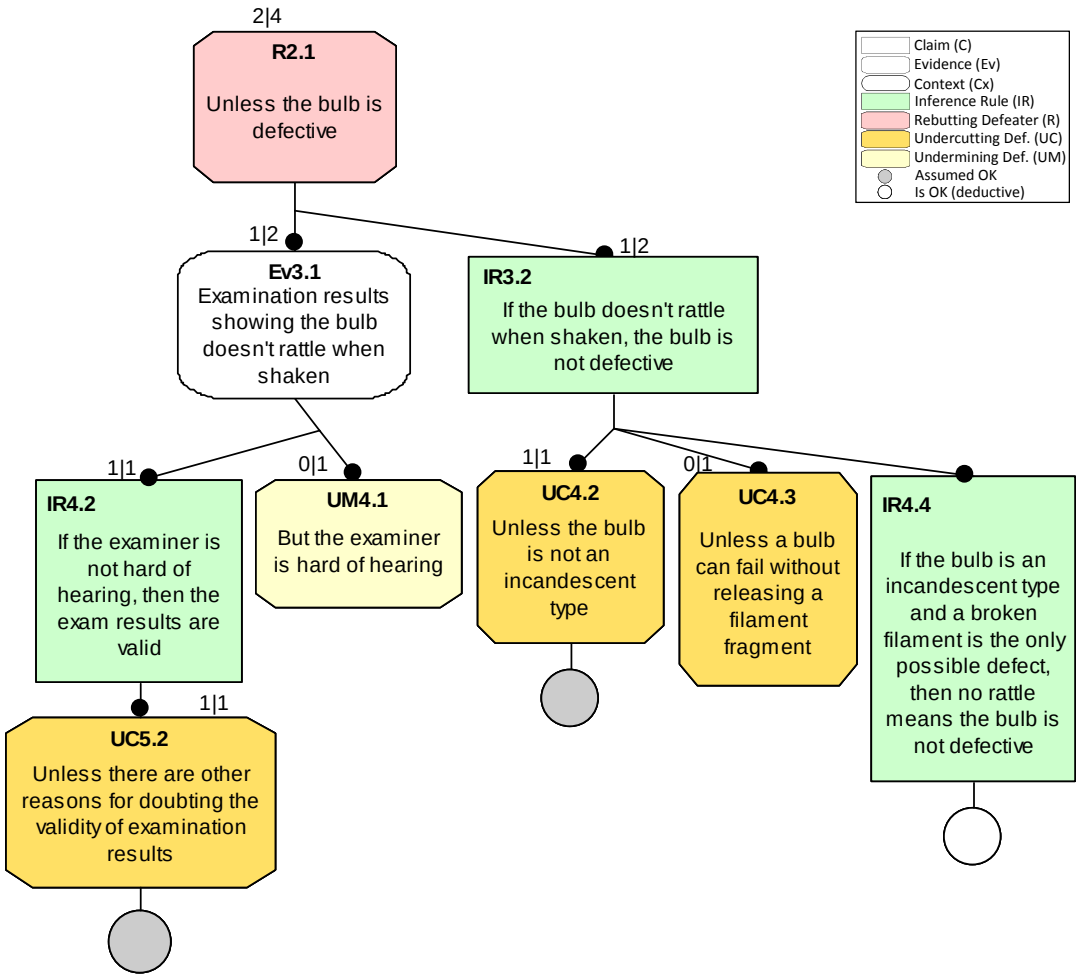


Figure 9: Confidence Evaluation of Defeater R2.1

[7] We call this approach *Baconian* in part because it follows Bacon's essential principle of eliminative induction and in part because of Cohen's notion of Baconian probability [Cohen 1989]. This relationship is discussed further in Section 4.4.

2.3.1 Baconian Confidence

In the Baconian view of eliminative argumentation, defeaters are either fully eliminated or not. If we apply this simple view to the lighting case, we arrive at the confidence evaluation shown in Figure 9. Starting with doubts about the validity of the evidence Ev3.1, we see there are two defeaters at or closest to the leaves of its subtree—UC5.2 and UM4.1. UC5.2 has been completely eliminated by assumption and UM4.1 is not yet eliminated, so of the two defeaters, only one has been eliminated. Therefore, our Baconian confidence in the evidence is 1|2. Similarly, if we look at defeaters applicable to inference rule IR3.2, we see that one of the two defeaters has been eliminated (by assumption), so our Baconian confidence in the validity of the rule is 1|2.

Confidence in the elimination of defeater R2.1 depends on how much confidence we have in evidence Ev3.1 and inference rule IR3.2. Since the defeaters for the evidence and for the inference rules are different, we take the union of the defeaters and their state of elimination as the basis for confidence in the elimination of R2.1. Looking at the subtree rooted at node R2.1, we see four defeaters at or nearest the leaves of the tree (UC5.2, UM4.1, UC4.2, and UC4.3), of which only two have been eliminated, so we say the confidence level associated with R2.1 is two out of four (2|4).

Moving up the tree and using the same calculation rule (see Figure 10), we arrive at C1.1 and assign it a confidence level of 2|7, meaning that seven doubts have been identified, of which only two have been eliminated. It is best to focus not on the number of eliminated doubts or the fraction of eliminated doubts but instead to focus on the *residual* doubt—that is, the number of uneliminated doubts—because this represents the additional assurance work that is required to develop complete confidence in the top-level claim. In this case, the residual doubt associated with C1.1 is 5—five doubts remain to be eliminated.

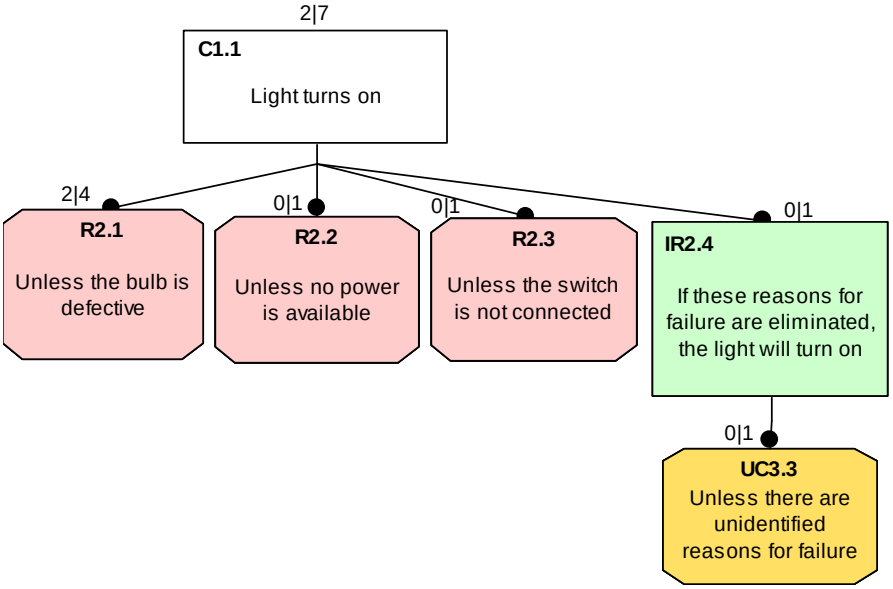


Figure 10: Confidence Evaluation for Top-Level Claim

2.3.2 Probabilistic Confidence

The Baconian approach is one method for calculating confidence (see Section 3.2 for further discussion). Considering the probability that a defeater is eliminated is another. For example, suppose we decide (or experimentally determine) that UC4.3 (about bulb failure modes) is true 10% of the time.[8] This means UC4.3 has a 90% probability of being eliminated (see Figure 11). Since the confidence map argument asserts that UC4.2 is eliminated with complete confidence (i.e., it has 100% probability of being false for the system being considered), and since UC4.2 and UC4.3 are logically independent, the joint probability that both are eliminated is 1.00 * 0.90 = 0.90. The probability that inference rule IR3.2 is valid is therefore determined by the probability that both undercutting defeaters are eliminated.

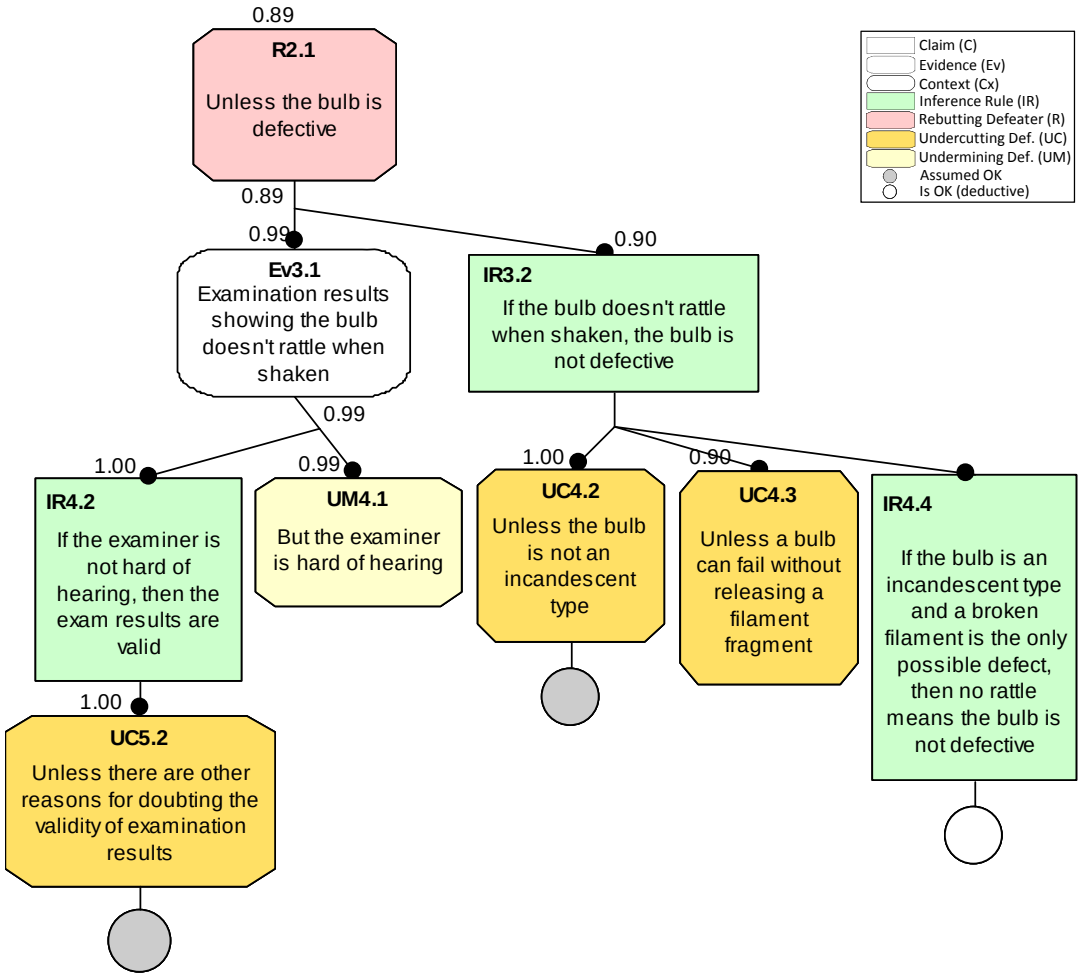


Figure 11: Probabilistic Confidence Assessment

Continuing in the same vein, how likely is UM4.1 ("The examiner is hard of hearing") to be false? For this example, suppose we assume the examiner is very unlikely to be hard of hearing; in this example, the probability that UM4.1 is eliminated is 0.99. Since we have assumed there are

[8] This is an example of enumerative induction because the probability is based on observing how often some effect occurs.

no other reasons for invalidating the examination results (UC5.2 is eliminated by assumption), the probability that Ev3.1 is valid is (1.00)(0.99) = 0.99. Finally, the probability that R2.1 is eliminated is the product of the probabilities that the evidence (Ev3.1) is valid and that the inference rule (IR3.2) is valid; that is, (0.99)(0.90) = 0.89.

2.3.3 Other Ways to Calculate Confidence

Other ways of evaluating and calculating confidence are certainly possible. For example, one might object that values such as 0.99 or 0.90 are overly precise and instead just use a confidence scale such as the following:

- CC = completely confident
- VC = very confident
- MC = moderately confident
- SC = somewhat confident
- NC = not confident

One would then need to develop rules for combining these confidence levels. One could adopt a fuzzy logic approach and say that when combining several confidence levels to determine a joint confidence level, the joint confidence is the minimum of all the confidence levels being combined. For example, *Joint*(VC, CC) = VC. If in our example we are (at least) "very confident" that every defeater is eliminated, we would be "very confident" in the top-level claim.

We have discussed how to determine the extent to which a defeater is eliminated based on the amount of confidence we have in the argument used to eliminate it. There is another aspect that could be considered as well, namely, the relative importance of eliminating one member of a set of defeaters. For example, we have not considered whether more confidence is gained from eliminating R2.1 ("Unless the bulb is defective") than from eliminating either of the other two defeaters.

We have investigated various ways of evaluating relative importance among a set of defeaters and using this information to arrive at an overall estimate of confidence, but more research is needed and we are not yet ready to present our findings. To explain the notions of eliminative argumentation, however, we treat a simple case in which in which we consider all defeaters to be of equal importance.

2.4 Rules for Composing Eliminative Arguments and Confidence Maps

We have introduced the elements of eliminative argumentation by example. In this section, we discuss and give precise specification of constraints on defeaters and what constitutes a well-formed eliminative argument. We also introduce notational conventions that are used in confidence maps. In Section 2.5 we provide a more realistic example and consider some of the issues that arise when constructing an eliminative argument.

2.4.1 Fundamental Elements of an Eliminative Argument

An eliminative argument has five fundamental elements: doubts (expressed as defeaters), claims (C), evidence (Ev), inference rules (IR), and argument terminators. There are (only) three kinds of defeaters, determined by the element to which the defeater applies. Rebutting defeaters (R) imply

that a claim is false. Undermining defeaters (UM) imply that evidence is invalid. Undercutting defeaters (UC) identify deficiencies in an inference rule such that even when its premises are true, the rule's conclusion could be true or false.

2.4.2 Elements of a Confidence Map

A confidence map is a visualization of an eliminative argument. Hence, a confidence map contains all five fundamental elements of an eliminative argument. As Table 2 on page 7 shows, claims and evidence are represented with uncolored graphical elements (rectangles and rectangles with rounded corners, respectively). Defeaters are expressed in rectangles with chopped-off corners. We distinguish the different kinds of defeaters by color: red for rebutting defeaters, yellow for undermining defeaters, and orange for undercutting defeaters. Inference rules are specified in green rectangles.

A "context" element, which gives additional information about the content of a fundamental element, is optional in a confidence map. A context element is represented with a clear rectangle whose ends are rounded. In our examples, a context element expresses a rewriting rule for some term in a claim (or other element). For example, when a claim states that "The system is acceptably reliable," a context element can specify that "acceptably reliable" means "the probability of failure on demand is less than 10^{-3} with probability 99%." Everywhere the phrase "acceptably reliable" is used in the argument, it is considered to be shorthand for the more precise specification of failure probability. Similarly, in the lighting example, we could have specified in a context element that "bulb" means "an incandescent bulb," in which case, IR3.2 in Figure 9 would be short for saying that "If an incandescent bulb doesn't rattle when shaken, the incandescent bulb is not defective." Consequently there would be no need to write UC4.2 expressing a doubt about whether the bulb is incandescent or not. If the bulb is not an incandescent bulb, we will need to construct a different argument.

2.4.3 Rules for Well-Formed Elements of an Argument

2.4.3.1 Rules for Claims

A claim is stated as a predicate, that is, a true or false statement. (Claims should be stated simply, with context elements being used to define general terms such as "acceptably reliable.")

2.4.3.2 General Rules for Defeaters

Defeaters have three possible states: true, eliminated (i.e., false), and uneliminated. In principle, an eliminative argument never contains a true defeater because, at best, such an argument would not provide confidence in the top-level claim and, at worst, it would falsify the top-level claim. If, in the course of developing an eliminative argument, we discover that a defeater is true, then we may need to change something in the system or argument so the defeater becomes irrelevant or so the defeater is no longer true (see Section 3.5 for a fuller discussion).

If a defeater is eliminated, it is not true. If a defeater is not eliminated, then it is unknown whether it is true. All our reasoning about levels of confidence is based on the degree of belief that a defeater is eliminated. For this reason, if D is a defeater, we define the predicate *elim*(D) such that *elim*(D) is true if D is false and otherwise its truth value is undetermined. We tend to speak of the probability that a defeater is eliminated or not eliminated rather than its being true or false.

2.4.3.3 Rules for Rebutting Defeaters

A rebutting defeater is a predicate associated with a claim. In confidence maps, the predicate is preceded by the word *Unless*. If P is the predicate, then the defeater appears in a confidence map as "Unless P."

The key requirement for a valid rebutting defeater is that it express a *relevant* doubt about the validity of the associated claim, that is, the doubt, if true, should imply that the claim is false. Eliminating such a doubt is then a reason for having more confidence in the claim.

More precisely, a rebutting defeater asserts a sufficient condition for its associated claim to be false; that is, P expresses a reason that is believed to indefeasibly imply the falsity of C. For example, in an eliminative argument, asserting that unconnected switch wiring is sufficient reason for believing a light will not turn on is equivalent to asserting that the wiring provides the only electrical path to the light. When reviewing an eliminative argument, one needs to decide whether a rebutting defeater really does state a sufficient condition to falsify the claim. If it does not, either the defeater needs to be reframed or its elimination will leave some doubt that the claim is true (that is, one cannot have complete confidence in the inference rule linking the falsification of P to the validity of the claim; an example is given in Section 4.3.2).

Ideally, a rebutting defeater should express a reason why the claim is false—P should not simply be ~C. For example, although "Unless the light does not turn on" satisfies the requirement that a rebutting defeater contradict a claim that the light turns on, it does not give any reason for believing that the light will not turn on and it does not suggest what further argument and evidence will increase confidence in the claim. Articulating a proper rebutting defeater can sometimes be difficult (see Section 2.5.2).

If more than one rebutting defeater is associated with a claim, it is desirable that each rebutting defeater state an independent reason that the claim might not be true.[9] Two defeaters are independent if the elimination of one defeater does not eliminate the other defeater. In the lighting case, the rebutting defeaters are independent because showing that the bulb is good, for example, says nothing about whether the switch has power or is wired to the light, and similarly, for the other rebutting defeaters. We desire defeater independence because if the elimination of one defeater implies the elimination of others, then the number of uneliminated defeaters is an overestimate of the amount of residual doubt and, therefore, an overestimate of our lack of confidence. We do not *require* defeater independence because it is more important to have a clear argument, and analyzing dependences among defeaters can sometimes be a difficult and unrewarding exercise.

2.4.3.4 Rules for Undercutting Defeaters

An undercutting defeater is a predicate, UC, expressing a doubt about the validity of an inference rule (P → Q). In confidence maps, the predicate is preceded by the word *Unless*.

[9] This requirement only holds for an argument in which the elimination of all defeaters associated with a claim is necessary to have complete confidence in the claim (see the discussion of linked arguments in Section 4.3.2).

The effect of a true undercutting defeater UC is to place Q in doubt when P is true. If UC is true, the rule's conclusion does not necessarily follow when its premise is true; that is, if both UC and P are true, we do not know whether Q is true or not.

Since UC indicates uncertainty about Q when P is true, it cannot be the case that UC → ~Q or that UC → Q. UC cannot imply that Q is true or false because its role is to identify reasons for not knowing whether Q is true or false. We discuss an example in Section 2.5.1.1 in which it is difficult to formulate an undercutting defeater that meets this condition.

If more than one undercutting defeater is associated with an inference rule, it is desirable that each such defeater state an independent reason the rule might not be valid. Two undercutting defeaters are independent if the elimination of one defeater does not eliminate the other.

2.4.3.5 Rules for Evidence

Evidence has the form "[Noun phrase] showing P," where the "Noun phrase" describes the data comprising the evidence (e.g., "Examination results") and P is a predicate (e.g., "bulb doesn't rattle when shaken"). P asserts an interpretation of the data that is relevant to the argument (the *evidence assertion*[10]), whereas the noun phrase serves only to identify the information whose interpretation is relevant.

2.4.3.6 Rules for Undermining Defeaters

An undermining defeater is a predicate, UM, associated with evidence. In confidence maps, the predicate is preceded by the word *But*.

Undermining defeaters express doubts about the validity of evidence, with the idea that invalid evidence cannot eliminate any defeater and therefore is irrelevant to an argument. Like undercutting defeaters, a true undermining defeater does not mean that the evidence assertion is false.

Evidence can be attacked in two ways:

1. by challenging the validity of the data comprising the evidence (e.g., in the light-turns-on example, the examination results would be invalid if the light bulb that was tested is not the bulb that is actually used in the system)
2. by challenging the validity of the interpretation of the data (e.g., by asserting a reason why the report of "No rattle when bulb is shaken" could be incorrect)

If more than one undermining defeater is associated with an item of evidence, it is desirable that each such defeater state an independent reason that the evidence might not be valid. Two undermining defeaters are independent if the elimination of one defeater does not eliminate the other. (We will see some examples in Section 0.)

2.4.3.7 Rules for Inference Rules

An inference rule is a predicate of the form (P → Q) where either P or Q (but not both) is an eliminated defeater.

---

[10] We use the notion of an evidence assertion as defined by Sun and Kelly [Sun 2013] and the Structured Assurance Case Metamodel, or SACM [OMG 2013], but we combine it with the evidential data in a single node. We discuss this difference further in Section 4.1.

In eliminative argumentation, confidence in claims, evidence, or inference rules is increased by identifying and eliminating defeaters. Consequently, there are two types of inference rules. The first explains why a defeater is eliminated, and the second explains why a defeater's elimination increases confidence in the validity of a claim, evidence, or inference rule:

1. The first type of inference rule has the form ($\wedge_i$ *valid*(EC_i) → *elim*(D)), where EC_i is a member of a set of claims or evidence associated with defeater D.
2. In the second type of inference rule, the premise, P, is a conjunction of eliminated defeaters and the conclusion, Q, is an assertion that a claim, evidence, or inference rule is valid:
   - ($\wedge_i$ *elim*(R_i) → *valid*(C)), where *valid*(C) means that C is true and R_i is a set of rebutting defeaters associated with C
   - ($\wedge_i$ *elim*(UM_i) → *valid*(Ev)), where *valid*(Ev) means that the evidence assertion is true and UM_i is a set of undermining defeaters associated with Ev
   - ($\wedge_i$ *elim*(UC_i) → *valid*(IR)), where UC_i is a set of undercutting defeaters associated with IR and *valid*(IR) means that the conclusion of the inference rule is valid when P holds

2.4.3.8 Rules for Terminators

An Assumed OK terminator is an assertion that some defeater is (assumed to be) false. (In confidence maps, such terminators are represented as small grey circles.)

An Is OK terminator applies to inference rules and indicates that the rule has no undercutting defeaters because it is a tautology. (In confidence maps, such terminators are represented as small clear circles.)

In any real-life argument, there are always possibilities for doubt, even if these possibilities are thought to be remote and not worth considering. Such doubts are best expressed as defeaters and eliminated explicitly with a terminator. But sometimes we have evidence or a claim that is so obviously valid we can't think of any reasons to doubt its validity (or more likely, we don't want to make the effort to think of remotely possible doubts that we are then going to immediately eliminate). In such cases, we allow an Assumed OK terminator to be associated directly with a claim, evidence, or an inference rule.

2.4.3.9 Eliminative Argument Structural Rules

The leaves of an eliminative argument are, in principle, terminators associated with an eliminated defeater,[11] uneliminated defeaters, or inference rules that are asserted to be a tautology. This constraint on the leaves of an argument reflects our view that confidence is dependent on identifying and eliminating doubts.

Confidence in a claim, evidence, or inference rule can only be supported, in principle, by identifying and eliminating defeaters associated with the element. In the simplest case, confidence is supplied by eliminating all doubts associated with the claim, evidence, or inference. We express this

[11] If a defeater D is at a leaf, it can be eliminated only by assumption. For example, if it is eliminated with evidence, the evidence and inference rule linking the evidence to the defeater will have undermining and undercutting defeaters, so D cannot be at a leaf.

as a rule of the form ($\wedge$*elim*(D_i) $\rightarrow$ *valid*(X)), where D_i is a member of the set of defeaters associated with X and X may be a claim, evidence, or inference rule. In argumentation theory, such a rule forms a *linked* argument (see Section 4.3.2).

In other cases, different groups of defeaters provide alternative arguments supporting X (we will see an example in Section 2.5.3); for example, ($\wedge_i$ *elim*(D_i) $\rightarrow$ *valid*(X)) $\vee$ ($\wedge_j$ *elim*(D_j) $\rightarrow$ *valid*(X)), where $D_i \neq D_j$ for all $i \neq j$. Such structures have been called "multi-legged arguments" in the assurance case literature [Bloomfield 2003] or "convergent" arguments in the argumentation literature [Beardsley 1950] (see further discussion in Section 4.3).

Multi-legged arguments seem to occur only for claims. We have not encountered examples in which alternative arguments are provided to show that evidence is valid or that an inference rule is valid. But such argument structures are possible in principle.

2.4.3.10 Notational Conventions in Confidence Maps

If an inference rule is indefeasible, it need not be represented explicitly in a confidence map. For example, if a defeater says "Unless statically detectable errors exist" and the evidence serving to eliminate the defeater is "Static analysis showing no statically detectable errors," the inference rule connecting these elements is "If there are no statically detectable errors, then no statically detectable errors exist." This inference rule is a tautology—that is, indefeasible—and need not be written explicitly since it is both indefeasible and obvious. In other cases, where the relationship between the premise and conclusion is not obviously indefeasible, an explicit inference rule should be written and explained (see Section 2.5.1).

When a set of undermining defeaters, UM_i, is associated with an item of evidence, the intent is usually to argue that all relevant undermining defeaters have been identified. In such a case, the associated inference rule ($\wedge_i$ *elim*(UM_i) $\rightarrow$ *valid*(Ev)) has the undercutting defeater "Unless all reasons for evidence invalidity have not been identified," and this defeater is assumed to be false. For example, see Figure 8. In this case, the inference rule, its defeater, and the defeater's "assumed eliminated" terminator can be omitted from the confidence map. (See the discussion associated with Figure 19 for an example.)

In eliminative argumentation, shaded terminators should only be associated with defeaters to show that a doubt is being eliminated by assumption. As shorthand, we sometimes might attach a terminator directly to a rule, a claim, or evidence to indicate we have no doubts about the validity of the rule, claim, or evidence. Methodologically we believe it gives more insight into an argument to take the effort to identify defeaters that are then eliminated by assumption, since this assumption can always be questioned. If confidence in a claim, for example, is asserted directly, one may be overlooking an important deficiency in the argument. Nonetheless, it is sometimes more practical to assert an assumption of no doubts.

Sometimes the elimination of a defeater is presented more clearly by breaking a defeater into subcases that can be more clearly addressed individually. In such a case, the defeater can be considered a conjunction of independent subdefeaters, that is, $D = D_1 \wedge D_2 \wedge D_3 \ldots$; we call this *defeater refinement*. It is the only situation in which a defeater can be directly connected to another defeater. In such a case, D is eliminated only if all the subdefeaters are eliminated. We will give an example in Section 2.5.1.1.

2.4.4 Summary

In this section, we have specified the requirements that a well-formed eliminative argument must satisfy and, correspondingly, the requirements for a well-formed confidence map. We will see in the next section how building an argument that conforms to these requirements provides guidance as to how an argument should be structured. In Section 4.1 we discuss similarities and differences between confidence map notation and various assurance case notations.

Since a confidence map is a visualization of an eliminative argument, we have specified notational conventions that make confidence maps more compact than they would otherwise be. As we gain more experience using confidence maps, we may develop different conventions for making maps more compact.

2.5 A More Realistic Example

In the preceding sections, we introduced the basic concepts of eliminative argumentation using an artificial example and specified the structural and notational rules for a well-formed confidence map. In this section, we consider a more realistic example and show how to develop a well-formed confidence map. In particular, we consider alternative argumentation structures.

Bloomfield and Littlewood consider how to provide confidence in a claim that a system is acceptably reliable if its probability of failure on demand (*pfd*) is less than 10^{-3} and we are 99% certain that 10^{-3} is an upper bound on the failure rate [Bloomfield 2003]. Such a claim is statistically justified if 4,603 operationally random tests are executed successfully [Littlewood 1997]. Bloomfield and Littlewood consider the case where it has not been possible to execute the required number of tests. They write, "[If] it is infeasible to test for sufficiently long to substantiate the claim at the required level of confidence, we might require a second argument involving extensive static analysis." The addition of a static analysis argument is expected to provide "more (justifiable) confidence in the dependability of our system … than is provided by either one alone" [Bloomfield 2003, p. 27].

We use eliminative argumentation to analyze this reasoning and, in particular, to show the possible impact of unsatisfied doubts on our degree of belief (confidence) in the reliability claim. In addition, we use this example to further explicate the subtleties of formulating defeaters and different argument structures.

We first develop a confidence map for the statistical testing argument, assuming (for concreteness) that it has only been possible to execute 4,100 successful tests. We then consider an additional argument that uses static analysis evidence. Finally, we discuss how confidence in the top-level reliability claim is increased by combining the two arguments.

2.5.1 The Statistical Testing Argument

Bloomfield and Littlewood are not explicit about how to structure their argument, but Figure 12 is a possible representation of the statistical testing argument leg. We represent the argument using our version[12] of GSN [GSN 2011].

We now develop a confidence map (Figure 13) showing the reasoning that might underlie their argument. First, we consider the claim "The system is acceptably reliable." Since we plan to reason from statistical testing evidence, we have to decide what doubt would be eliminated by statistical evidence. In this example, we would doubt that the system is acceptably reliable if a set of randomly selected test cases failed one or more times in 4,603 cases (see R2.1 in Figure 13). (This rebutting defeater is a counterexample rather than a failure mode and stands in contrast to the rebutting defeaters used in the light-turns-on example.)

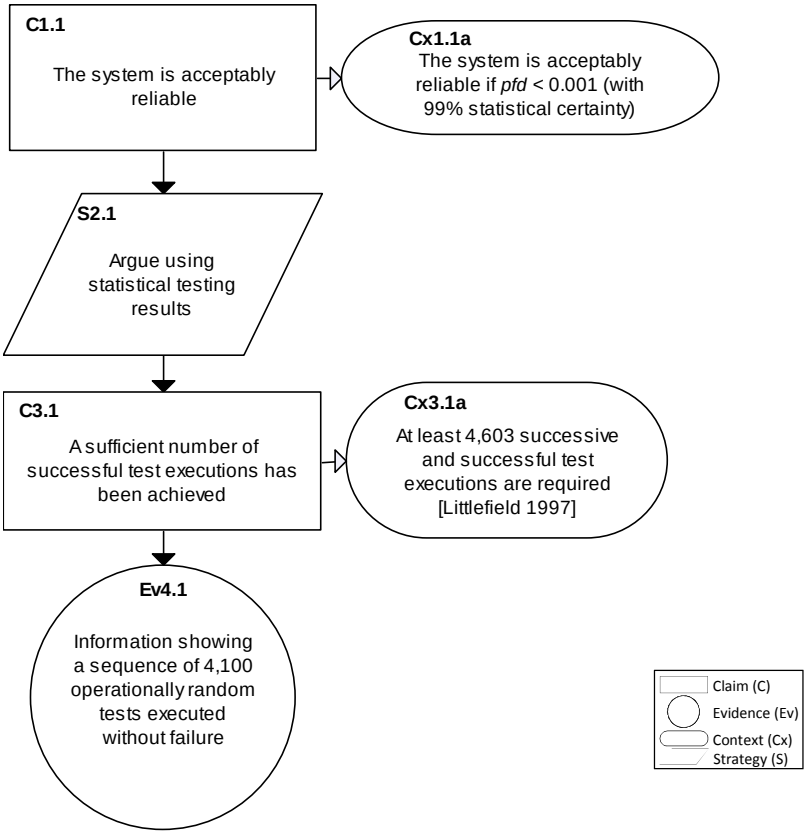


Figure 12: An Assurance Case Supported by Statistical Testing Evidence

[12] The differences in this example are mostly small: Evidence (Ev) instead of Solution (Sn), Claim (C) instead of Goal (G), the use of "Cx" for a context element, and the content and shape of the evidence node. S2.1 is a "Strategy" node, that is, a node explaining the argumentation approach. See Section 4.1 for a discussion of differences.

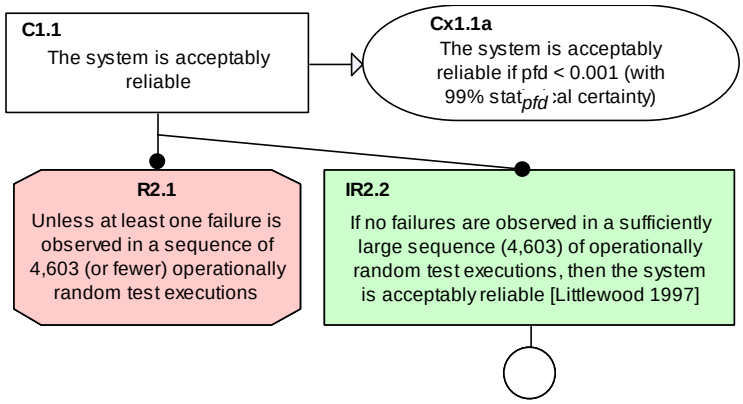


Figure 13: Top-Level Rebutting Defeater

The argument presented in Figure 13 is that elimination of the doubt expressed in R2.1 is sufficient to conclude that the system is acceptably reliable. This argument is expressed both by the structure of the confidence map (the choice of the rebutting defeater and its connection to claim C1.1) and explicitly in the inference rule (IR2.2), which says, in effect, that elimination of the rebutting defeater is sufficient to support the claim.

The premise of the inference rule is that 4,603 test executions are a random sample of the operational usage profile for the system and no failures are observed. When such a premise holds, the rule's conclusion necessarily follows [Littlewood 1997]. Simply put, *elim*(R2.1) indefeasibly implies C1.1. We indicate this with a reference to the Littlewood and Wright article and put a white Is OK terminator symbol under the inference rule to assert that no undercutting defeaters logically exist for the rule. Of course, doubts about adequacy of the operational profile, randomness of the test selection process, and test oracle reliability need to be part of the argument. However, they are not doubts relevant to this inference rule because any weaknesses of an inference rule (i.e., its undercutting defeaters) concern the validity of its conclusion *when its premises are true.* Sources of doubt about the premises of IR2.2 are captured in doubts about whether R2.1 has really been fully eliminated.

2.5.1.1 Analyzing an Inference Rule

We develop the confidence map further by considering how to eliminate defeater R2.1. In this case, the only evidence we have is successful results from 4,100 operationally random test executions (Ev4.1; see Figure 14).

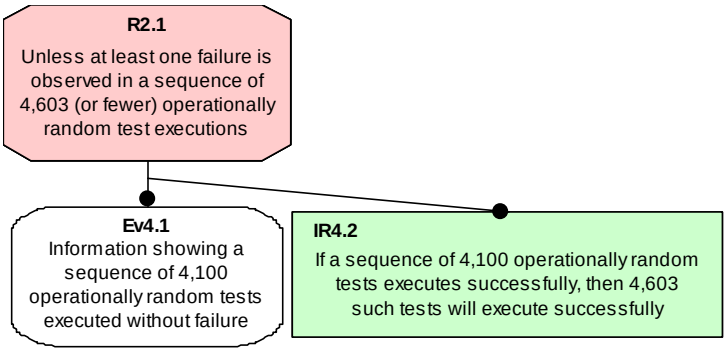


Figure 14: Analyzing an Imperfect Inference Rule

The inference proposed in Figure 14 is that the successful execution of 4,100 operationally random tests is sufficient to ensure that 4,603 tests would execute successfully. Of course, IR4.2 is invalid—satisfying its premise is *never* sufficient to guarantee its conclusion. The inadequacy of the rule could be captured in an undercutting defeater saying explicitly that the premise is insufficient—that is, "4,100 successful tests are insufficient"—but this defeater could never be eliminated, so R2.1 could never, in principle, be eliminated. This structure does not capture our intuition that we have learned something from the 4,100 successful test executions.

If we are to give an argument that eliminates R2.1, we need a different structure in which information about 503 additional tests is used. A possible argument is shown in Figure 15. In this argument, we insert a claim[13] about the desired result of executing 503 additional tests (claim C4.2). This claim is inserted in parallel with Ev4.1. The combination of the claim and Ev4.1 indefeasibly eliminate R2.1, so we don't need to show the corresponding inference rule.

Next we state a rebutting defeater for C4.2, namely, a reason why this claim would be false. In this case, if there are some errors in the system that cause[14] at least one of the next 503 tests to fail (R5.1), the claim will be invalid.

[13] We stated C4.2 as a claim because the 503 test results do not actually exist.

[14] A technical point: We do not say "could cause" a failure because such an assertion would not guarantee that a test fails, and a rebutting defeater must falsify a claim.

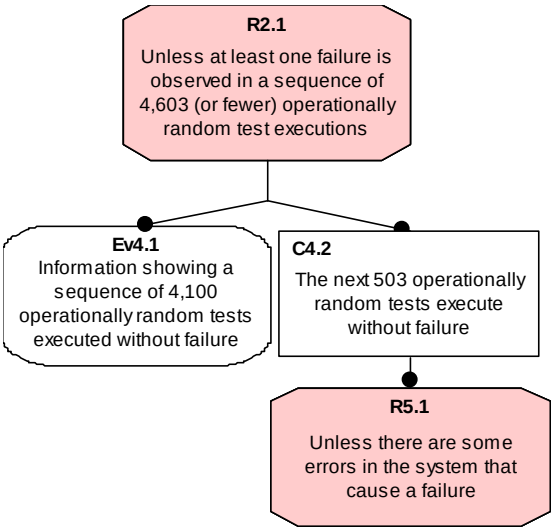


Figure 15: An Alternative Argument Eliminating R2.1

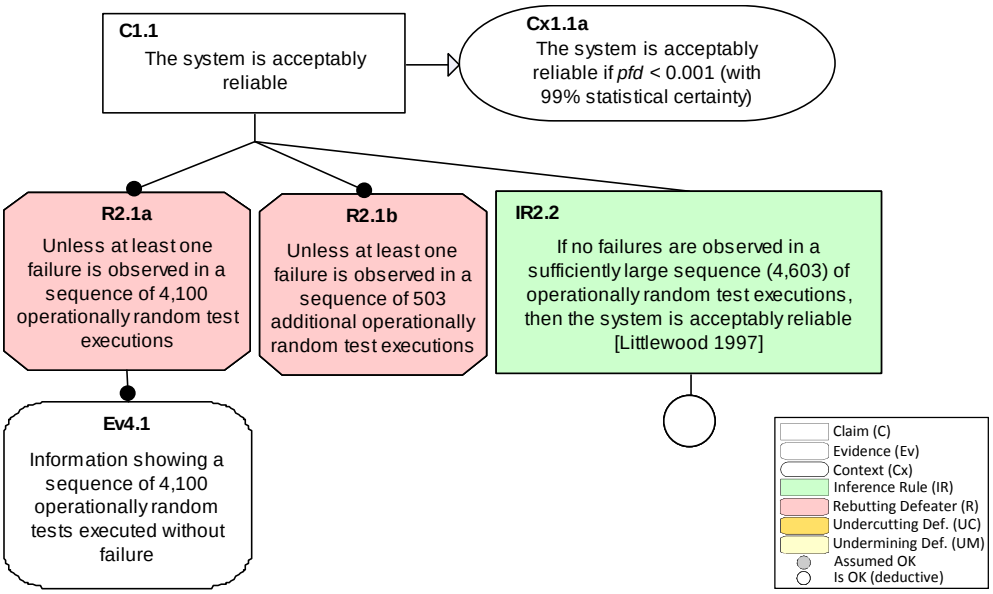


Figure 16: Splitting a Defeater

An alternative argument structure could split R2.1 into two rebutting defeaters: one for the evidence that we have and the second for the evidence we would like to have (see Figure 16).

The inference from Ev4.1 to *elim*(R2.1a) is a tautology, so we do not show the inference rule explicitly. The inference *elim*(R2.1) ∧ *elim*(R2.2) → C1.1 is indefeasible by virtue of Littlewood and Wright [Littlewood 1997] and by virtue of the deductive inference that if both defeaters are eliminated, no failures will have been observed.

Which of these arguments is preferred given that they are logically equivalent? The argument in Figure 16 is in some ways the simplest and most intuitive, although it fails to capture in a simple way the key notion that 4,603 successful test executions are required. This could be fixed by using the notion of *defeater refinement* (Section 2.4.3.10), in which a defeater is subdivided into a set of

mutually exclusive and collectively exhaustive defeaters that are logically equivalent to the defeater being refined (see Figure 17; the numbers in the figure will be used later).

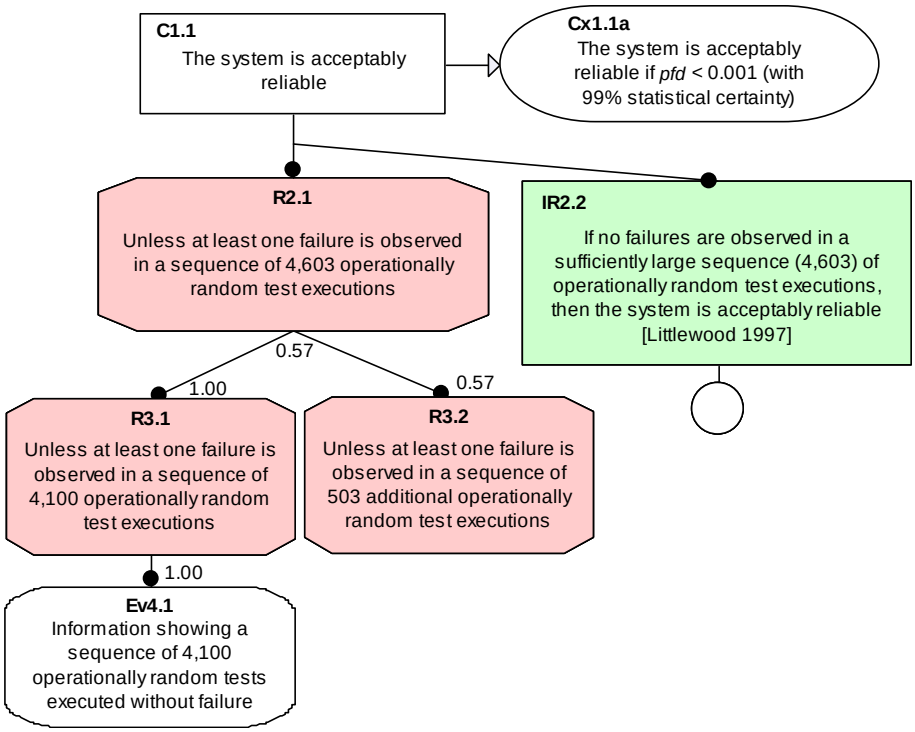


Figure 17: An Example of Defeater Refinement

2.5.1.2 Formulating Undermining Defeaters

Now let's consider why the evidence (Ev4.1) might be invalid. The evidence assertion consists of four independent assertions: (1) 4,100 tests were executed, (2) the tests were randomly generated or selected from (3) an accurate operational profile of system usage, and (4) the test executions were successful. Since information about how the tests were selected is relevant to the significance of their successful execution, we characterize the evidence as consisting of "information," not just "test results."

As shown in Figure 18, the five undermining defeaters (UM5.1, UM5.2, UM5.3, UM5.4, and UM5.5) each express doubts about the evidence: the operational profile might be inaccurate, the test selection process might be biased in some way, the test oracle might sometimes misclassify test results (and in particular, might misclassify a test failure as a success), the number of executed tests might be wrong due to some failure of test management procedures, and the system configuration might have changed (since the testing process may take a long time). Inference rule IR5.6 says explicitly what the structure of the diagram implies, namely, that elimination of these five doubts is necessary for the evidence to be considered completely valid. This inference about evidence validity would, however, not be correct if there are other reasons the evidence might be invalid; that is, the inference rule has an undercutting defeater to the effect "Unless all reasons for evidence invalidity have not been identified."

For this example, we do not want to argue further about whether these undermining defeaters are the only reasons for evidence invalidity. We do so by attaching an Assumed OK terminator symbol to defeater UC6.1. Of course, this assumption could be challenged by someone reviewing the argument. Similarly, we do not choose to argue further about whether each of the doubts about evidence validity is eliminated (although in practice, we would need to do so). To show that we are assuming all of these undermining defeaters are eliminated, we attach an Assumed OK terminator to each undermining defeater.

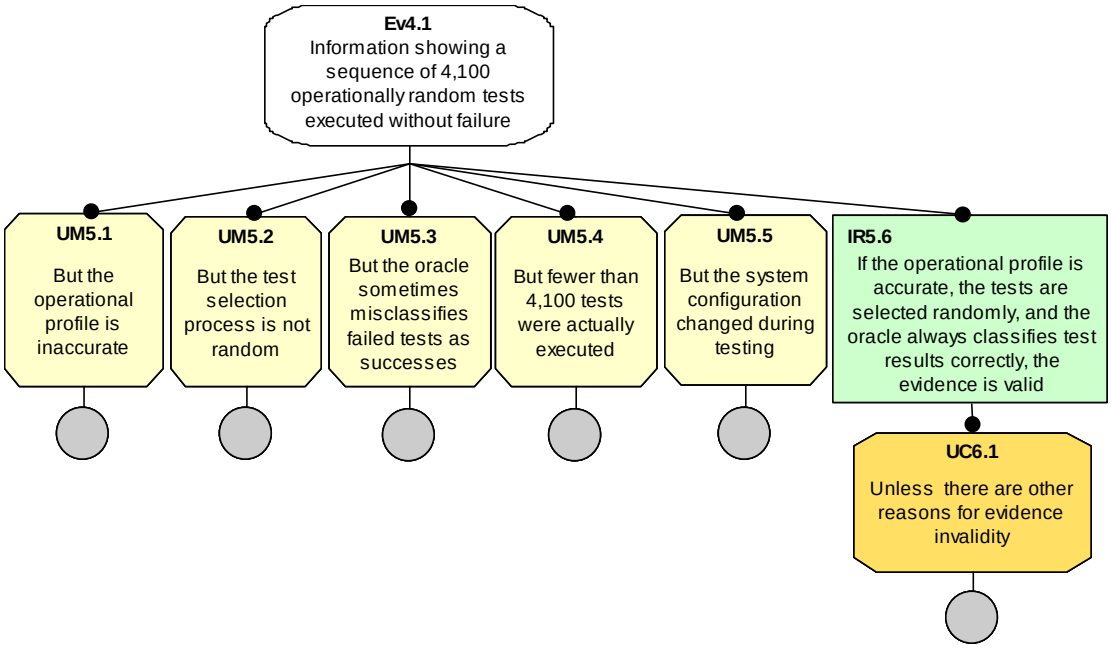


Figure 18: Reasoning with Undermining Defeaters

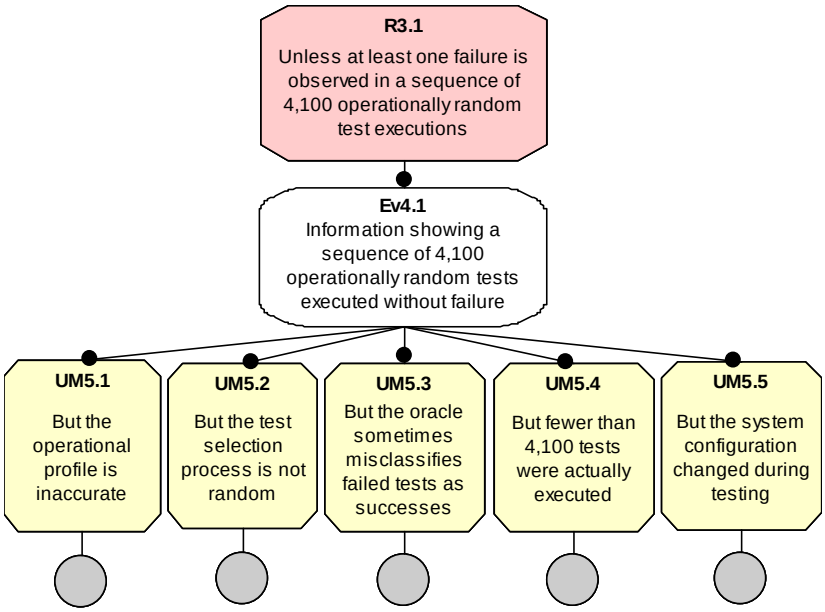


Figure 19: Implied Inference Rule with Undermining Defeaters

It is quite common to have a set of undermining defeaters associated with an item of evidence and to assume that these are the only relevant undermining defeaters (as in Figure 18). In such cases (as we noted in Section 2.4.3.10), we can elide the inference rule and its associated defeater from the diagram and tie the undermining defeaters directly to the evidence node, as shown in Figure 19. When reviewing such an argument, remember to ask whether you believe that the set of undermining defeaters is complete. When it is not obvious why the set is complete, you will need to expand the confidence map to show an argument eliminating the undercutting defeater.

2.5.1.3 Using Probabilities

Let's consider our degree of confidence that R2.1 in Figure 17 is eliminated. To do so, we need to consider how much confidence we have that each subdefeater is eliminated. Starting with R3.1, we note that the validity of Ev4.1 is not in doubt since we assumed that all its identified undermining defeaters are eliminated and that there are no other undermining defeaters. The inference (Ev4.1 → *elim*(R3.1)) is a tautology, and since we are completely confident of its premise (the validity of Ev4.1), we have no doubt about the conclusion that R3.1 is completely eliminated.

Our degree of confidence in the elimination of R2.1 depends on the joint probability that R3.1 and R3.2 have both been eliminated. We have no doubt about the elimination of R3.1. As for the elimination of R3.2, we have no direct evidence of its elimination, but we can estimate the probability that 503 additional tests will execute successfully because the successful execution of 4,100 tests puts an upper bound on the probability of failure on demand. Statistical analysis (see the appendix) tells us that this upper bound is $1.123*10^{-3}$ with 99% confidence. Therefore, the probability that 503 additional tests will execute without failure is $(1 – 0.001123)^{503} = 0.57$. This then is our degree of confidence that R3.2 is eliminated after 4,100 successful executions. The degree of confidence that R2.1 is eliminated can then be calculated as the joint probability that all its subdefeaters are eliminated, that is, (1.00)(0.57) = 0.57.

Our confidence that C1.1 is valid (see Figure 17) can be calculated as the joint probability that R2.1 has been eliminated and the probability that IR2.2 is valid, that is, (0.57)(1.00) = 0.57. Similar calculations can be done for the other argument structures, arriving at the same results for confidence in C1.1.

2.5.2 The Static Analysis Argument

Since we have not tested for sufficiently long to have complete confidence that $pfd < 10^{-3}$, we need to find other information that will help eliminate doubts relevant to the probability of failure. One such doubt would be the presence of a statically detectable coding error. It is probably for this reason that Bloomfield and Littlewood suggest the addition of "extensive" static analysis [Bloomfield 2003].

Bloomfield and Littlewood do not define the kind of static analysis they are considering. For purposes of concreteness in this report, we consider only source code analysis based on the program language's semantics. The object of the analysis is to identify surprising uses of the language that are likely to be errors, such as applying the `sizeof` function to a pointer instead of to the object referenced by the pointer [see PVS 2014]. The definition of a statically detectable error depends in part on the amount of analysis that can be done and whether the amount of effort needed to sort out false positives is acceptable.

Strictly speaking, a statically detectable coding error might not decrease system reliability; for example, if the error is in unreachable code, it can't cause a failure so it doesn't have to be removed.[15] But in practice, if the goal is to have a highly reliable system and static analysis detects a coding error, the error will be fixed. The justification for using static analysis evidence is that in practice, the system will not be considered acceptably reliable if a statically detectable error is known to exist (see IR2.4 in Figure 20).

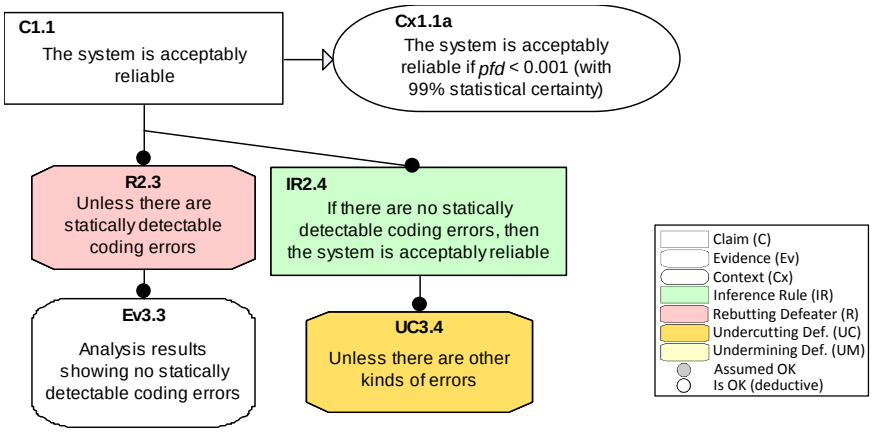


Figure 20: A Confidence Map for Static Analysis Evidence

Of course, the absence of statically detectable errors is not, by itself, a sufficient reason to believe that a system's *pfd* is $< 10^{-3}$. A variety of errors are not detectable by static analysis, such as coding errors not detectable by static code analysis (e.g., use of an incorrect constant in a formula), design errors (e.g., architectural decisions leading to occasional timing errors), specification errors (e.g., errors causing or allowing interactions among system elements that cause a system failure), requirements errors (of omission or commission), and unexpected environmental effects (e.g., operator errors). Given these other possible sources of failure, it is clear why reasoning about system reliability from the absence of statically detectable errors alone is insufficient.

[15] The statically detectable error also could occur in a section of code that is rarely executed. Or it might be executed under conditions in which the intended result is serendipitously produced despite the error, such as if the size of an object referenced by a pointer is equal to the size of the pointer and the `sizeof` function is incorrectly applied to the pointer instead of to the referenced object.

Strictly speaking, a system can be adequately reliable despite the presence of bugs as long as such bugs are not encountered too often in actual system operation. But in the absence of knowing how often some failure situation may occur during system operation, we try to eliminate as many such reasons for failure as possible.

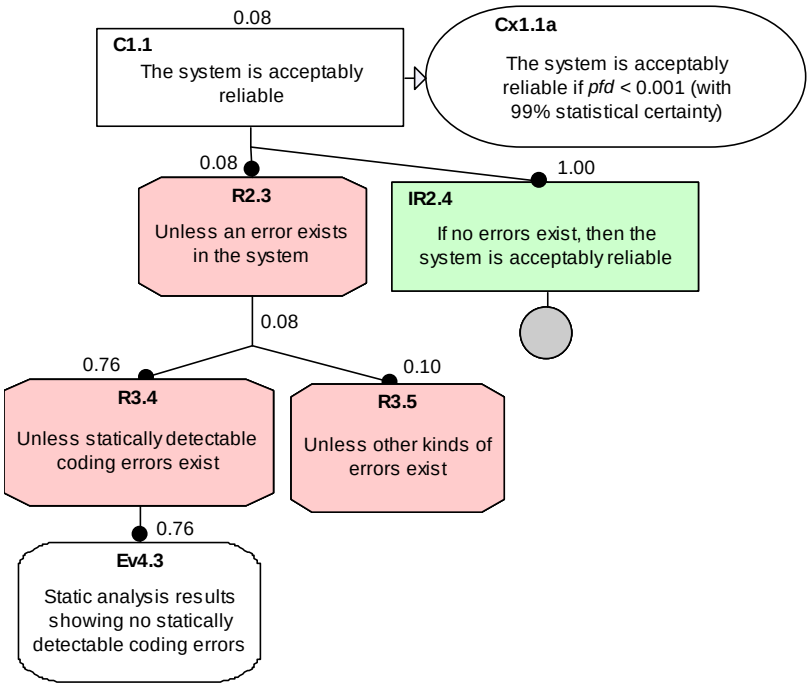


Figure 21: A More Complete Argument Involving Static Analysis

The undercutting defeater proposed in Figure 20 ("Unless there are other kinds of errors") indicates why the inference from lack of statically detectable errors to acceptable system reliability is unsound. But this is not an acceptable undercutting defeater; it does not put the conclusion about system reliability in doubt. Instead, it contradicts the conclusion—knowing about some error that is not statically detectable would eliminate any confidence we might have in system reliability just as much as knowing about a statically detected error would. To show the inadequacy of reasoning just from a lack of statically detectable errors, we must treat the proposed undercutting defeater as the rebutting defeater it actually is. In essence, we are arguing that the existence of *any* kind of error (not just statically detectable coding errors) causes us to lack confidence in the system's reliability. This means the confidence map argument must be structured something like that shown in Figure 21, where the fundamental doubt about the existence of an error is captured in R2.3, which, in turn, is decomposed into doubts about different types of errors. Since we are primarily interested in statically detectable errors, we single out this error type in R3.4 and lump the other types of errors into R3.5. With this structure, inference rule IR2.4 is clearly indefeasible, although stronger than necessary to have full confidence in the claim.[16]

The map in Figure 21 indefeasibly asserts that evidence Ev4.3 eliminates R3.4 (because the evidence assertion, "no statically detectable coding errors," directly contradicts the defeater). Hence no inference rule is shown for this relationship. All doubts about the elimination of R3.4 stem from doubts about the validity of Ev4.3.

[16] A weaker, but sufficient, rule would require that the number of failures caused by all types of errors not exceed the number allowed by the reliability claim. This would be a different argument and would require the formulation of different rebutting defeaters that would be harder to eliminate with static analysis evidence.

2.5.2.1 Doubts About the Evidence

The evidence eliminating R3.4 is the result of running a variety of static analysis tools on the code. We know that such tools don't detect all statically detectable coding errors and that not all errors are necessarily detected even by a combination of analysis tools. The evidence assertion states that because no statically detectable errors were found by the tools, no statically detectable errors are present. But we should have doubts about the validity of this assertion. We need further information to eliminate these doubts, such as information showing that the analysis tools are sufficiently powerful to detect static errors, that the tools have been used correctly in their search for errors, that human review of possible errors reported by the tools has not overlooked an actual error, and that all code has been analyzed. These reasons for doubting the validity of the evidence are expressed in UM5.6 and its subdefeaters, UM6.1–6.4, in Figure 22. To the extent that any of these are true, the static analysis effort could have overlooked some detectable errors, and the validity of Ev4.3 would be in doubt.

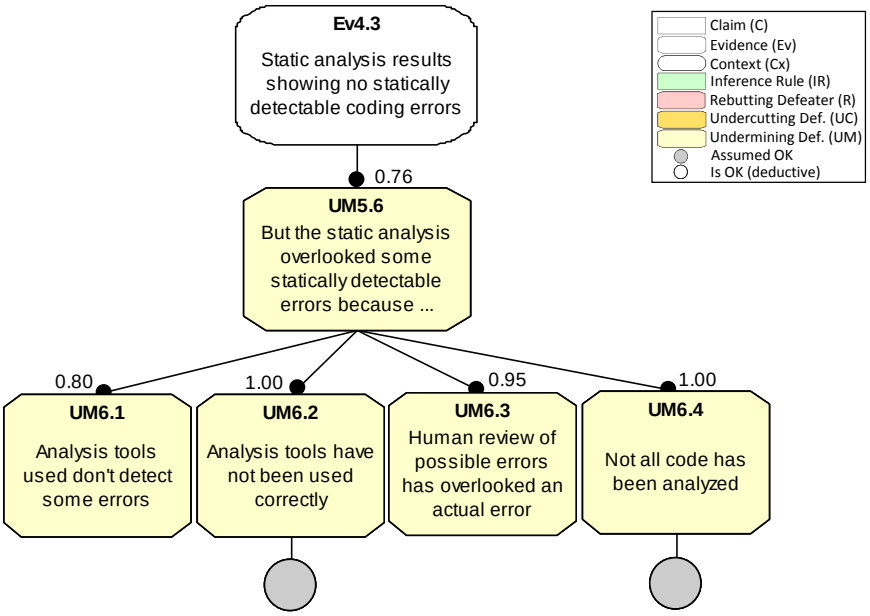


Figure 22: Defeaters for Static Analysis Evidence

How likely is it that UM5.6 is eliminated? For purposes of example, we might suppose that we are very confident (100%) that all code has been analyzed and that the tools were used correctly (UM6.4 and UM6.2). But we could be less confident that human review of possible errors (UM6.3) has reached the correct conclusions (say, only 95% confident) and even less confident in the ability of the tools to detect all static errors (UM6.1) (say, 80% confident).[17] Since each undermining defeater is independent of the others, the joint probability that UM5.4 is eliminated would be (0.80)(1.00)(0.95)(1.00) = 0.76. The inference (*elim*(UM5.4) → *valid*(Ev4.3)) is a tautology. Consequently, the probability that Ev4.3 is valid is also 0.76.

[17] These confidence estimates could be validated by collecting data from previous projects. This would be a use of enumerative induction since our confidence in the elimination of, say, UM6.1 would be based on how often statically detectable errors have been overlooked by the tools.

Concerning the notation in Figure 22, the inference rule associating the undermining defeaters with Ev4.3 is $\bigwedge_i$ *elim*(UM6.i) → *valid*(Ev4.3). Such a rule would not be valid if we had failed to identify a possible undermining defeater, but we assume here that all relevant undermining defeaters have been specified. By convention, we do not need to write such a rule and undercutting defeater explicitly; the diagram therefore implies an assumption that the listed defeaters are collectively exhaustive.

Confidence that R3.4 has been eliminated is based solely on confidence in evidence Ev4.3. At the end of Section 2.3.3, however, we discussed the difference between *confidence* in eliminating a defeater and the *importance* of eliminating a defeater. The importance of knowing that statically detectable errors are not present depends on how often such errors are the cause of failure. If they are not very significant, then having high confidence in their elimination should not significantly increase our confidence in the elimination of R2.3 and our confidence in C1.1. However, integrating assessments of importance into confidence calculations is a subject of future research.

2.5.2.2 Eliminating R3.5 and R2.3

Now let's consider the probability that there are errors other than those detectable by static analysis. Let's consider the probability that R3.5 is not eliminated. For this example, given no information about how the system has been vetted, there is probably only a small degree of confidence in the elimination of R3.5, say, 10%. Our confidence in the elimination of R2.3 is a function of our confidence in the elimination of R3.4 (76%) and R3.5 (10%), namely, (0.76)(0.10) = 0.08, so confidence in C1.1 is 8% given "extensive" static analysis evidence. This calculation shows how confidence gained from evidence about statically detectable coding errors is diminished by lack of confidence in the absence of other types of errors.

2.5.3 Extending the Statistical Argument with a Static Analysis Argument

We started with the goal of developing evidence to show that a system is acceptably reliable, but given that we can execute only 4,100 tests, we were faced with the question of how to increase confidence in the dependability of the system. Bloomfield and Littlewood suggested that static analysis evidence, in particular, could be combined with statistical testing evidence to form a "multi-legged" argument justifying increased confidence [Bloomfield 2003]. See Figure 23 for how such an argument about claim C1.1 might be structured as an assurance case.

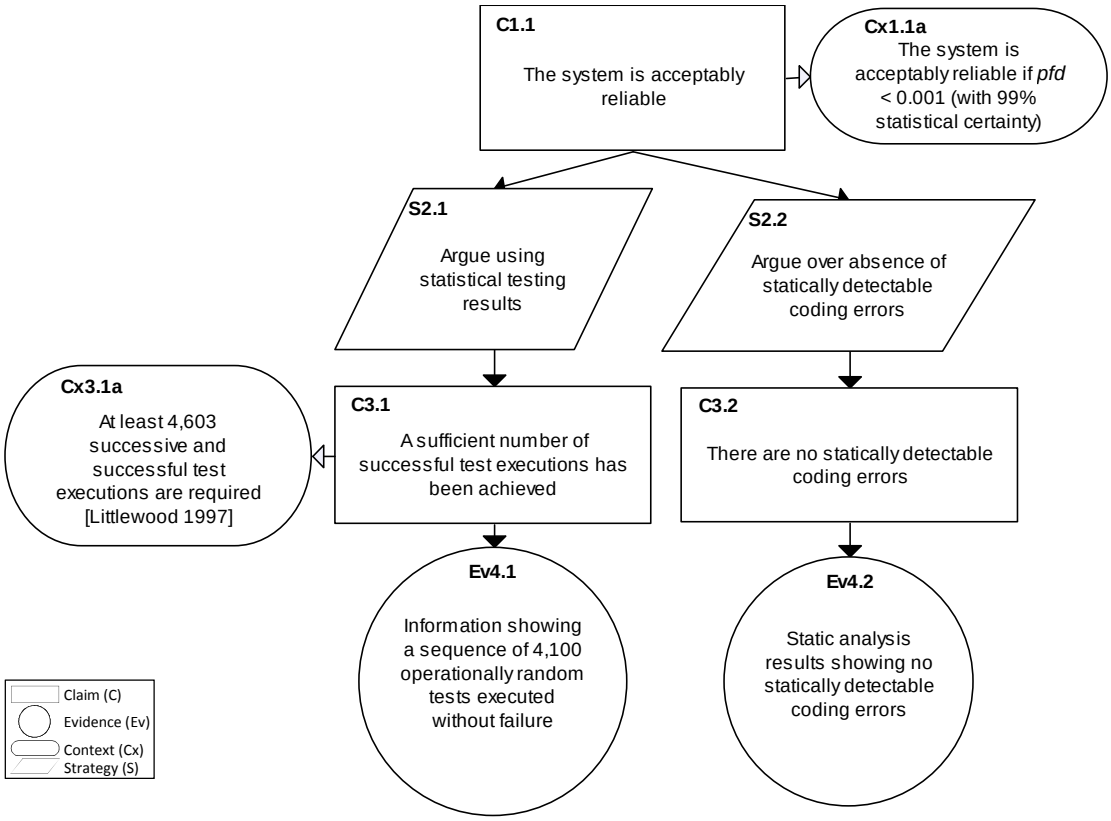


Figure 23: A Multi-legged Assurance Case

The idea of a multi-legged argument is that if the same top-level claim is supported by independently developed evidence, confidence in the validity of the claim should increase [Kelly 1998].

The statistical testing eliminative argument and the static analysis argument present independently developed evidence eliminating doubts about the reliability claim. In particular, if we look at the confidence maps for the evidence in each argument (Figure 19 and Figure 22), the undermining defeaters for the evidence in each leg are independent. Consequently, invalidation of evidence in one leg (due, perhaps, to further information being developed) does not affect evidence in the other leg.

Combining the two eliminative arguments leads to the structure shown in Figure 24. In this figure, rebutting defeater R2.3 represents an additional reason for doubting claim C1.1. However, if either one of these defeaters is completely eliminated, we would be completely confident in C1.1; we would not need the evidence associated with the other defeater. In particular, if we were to execute 4,603 tests successfully, we would not care if there are coding, design, specification, or other errors in the system. We would not need to check for other doubts that the system is acceptably reliable. In a multi-legged argument, having total confidence in one leg makes confidence in the other legs irrelevant.

In confidence maps, a multi-legged argument is indicated structurally by associating the conclusion of two or more inference rules with the same element of the argument, in this case, with a claim. In Figure 24, IR2.2 and IR2.4 both support the same claim. According to this argument

structure, either inference rule implies complete confidence in the claim when its premise is satisfied, and any undercutting defeaters are eliminated.

Bloomfield and Littlewood explain the increase in confidence from static analysis by supposing that the probability of each leg's being valid is independent [Bloomfield 2003]. Thus, if one leg supports the claim with 90% probability and the other with 95% probability, the probability that both legs will fail is $(1 – 0.90)(1 – 0.95) = (0.10)(0.05) = 0.005$. Therefore, given that both legs are independent, confidence in the top-level claim is the probability that at least one leg is valid, that is, $1 – 0.005 = 0.995$. This calculation is proposed as the rationale underlying our intuitive feeling that confidence is increased when independently developed evidence is offered in support of a claim.

If we apply this reasoning to our analysis of the statistical testing and static analysis arguments, we arrive at a confidence of $1 – (1 – 0.57)(1 – 0.08) = 0.60$. That is, confidence in the top-level claim is increased when both legs are considered together. This calculation depends on the assumption that each leg's contribution to belief in the top-level claim is statistically independent. But this is not clearly the case for our example. The *truth* of some rebutting defeater in the statistical testing leg will tend to increase the likelihood that a rebutting defeater in the analysis leg is true, and vice versa. If, for example, one of the 4,100 tests failed (or one of the additional 503 tests), it will be because there is an error in the system—one of the error types that is considered in the static analysis argument—and that argument should fail as well. Similarly, knowing that there is a statically detectable error in the system could reduce our confidence that 503 additional tests would succeed.

Equally well, the *elimination* of rebutting defeaters in one leg can increase the likelihood that certain defeaters in the other leg are eliminated. For example, the fact that 4,100 tests have executed successfully is information increasing our confidence that all errors in the system have been removed and therefore could be viewed as raising our confidence in the static analysis leg without providing any static analysis evidence at all.

One might choose to argue that as long as any evidence in each leg is independent in the sense that the evidence assertions and associated undermining defeaters in each leg are independent, information in one of the legs should not be used to determine the probability that a defeater in the other leg is eliminated. In short, confidence in one leg might best be argued without using any information from the other leg(s).

On the other hand, one might decide that the increase in confidence due to the development of independent evidence is best explained by incorporating the additional evidence into the original argument. For example, in the statistical testing leg, our doubt about the top-level claim arises from not knowing whether an additional 503 tests would execute successfully. We used the fact that 4,100 tests executed successfully to estimate how likely this doubt is to be true. But in addition, we could use static analysis evidence to argue that the probability of 503 successful tests is greater than 0.57 (see Ev4.4 in Figure 25).

Either argument structure provides the same degree of confidence in the top-level claim. A choice between them depends on which structure is thought to be more understandable to reviewers.

Figure 25 shows how the additional information provided by independent evidence can increase confidence by increasing the probability that one or more defeaters in the first leg are eliminated. This source of increased confidence is not so clear in the assurance case formulation of a multi-legged argument because the role of defeaters is not obvious.

We offer a few methodological observations on the confidence map in Figure 25:

- Undercutting defeater UC5.6 meets our requirements for an undercutting defeater because the existence of errors leaves it uncertain whether the 503 test executions will all succeed—a failure will occur if an error is encountered and otherwise the executions will succeed.
- Rebutting defeater R5.4, in contrast, accounts for errors being "encountered" while executing tests, since this is the only condition guaranteeing that the next 503 test executions will not all succeed. The phrasing of the defeater allows for the fact that errors may exist and not be encountered by the particular tests that are executed. The defeater is also appropriately phrased because it indicates a (not very interesting) reason why the test executions will fail, namely, because errors exist. An inappropriate rebutting defeater would have been something like "Unless at least one test fails," since this is just a rewording of the claim being attacked.
- In Figure 17, we evaluated the probability that R3.2 was eliminated without writing a claim equivalent to C4.3 in Figure 25. But in Figure 25, we needed to create a multi-legged argument structure so we could incorporate the static analysis evidence. To do so, we needed to provide a claim that would be parallel to Ev4.5.
- We needed to insert IR4.4 into the map to indicate visually that R3.2 is being eliminated by different inference rules; that is, a multi-legged argument is being used. But since IR4.4 is a tautology, we didn't bother to write out the inference in the confidence map.

2.6 Summary

The essential principle of eliminative argumentation is simple—identify doubts and show, by further argument and evidence, why certain doubts are eliminated and why some doubts remain. As doubts are eliminated, confidence increases. The eliminative argument approach provides both a basis for evaluating confidence in a claim and a method for developing a sound argument, namely, by identifying and eliminating doubts about system properties (rebutting defeaters), doubts about the reasoning (undercutting defeaters), and doubts about the evidence (undermining defeaters). At the minimum, the concepts provide a mental model for thinking about and developing confidence in system properties.

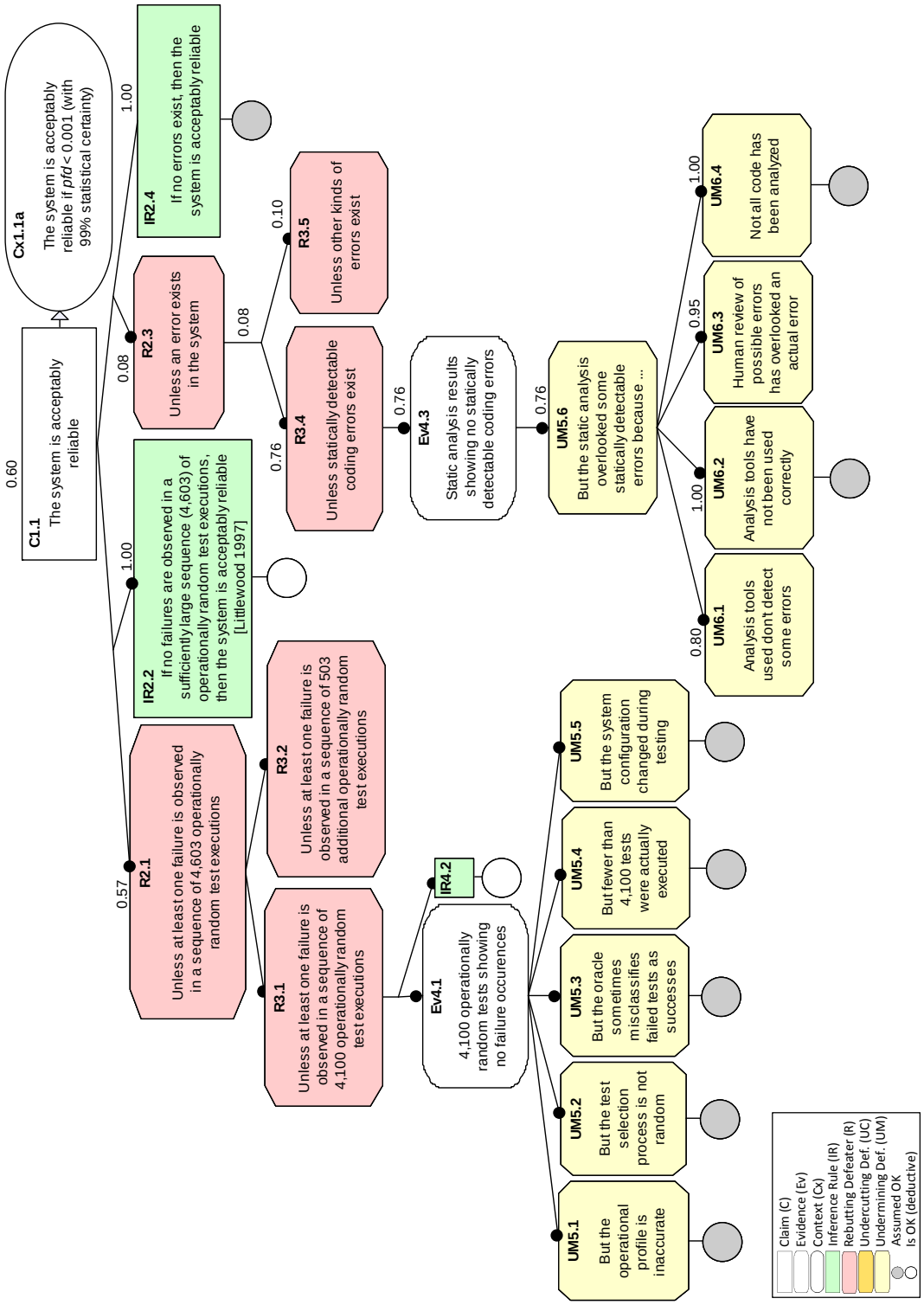


Figure 24: Multi-legged Confidence Map

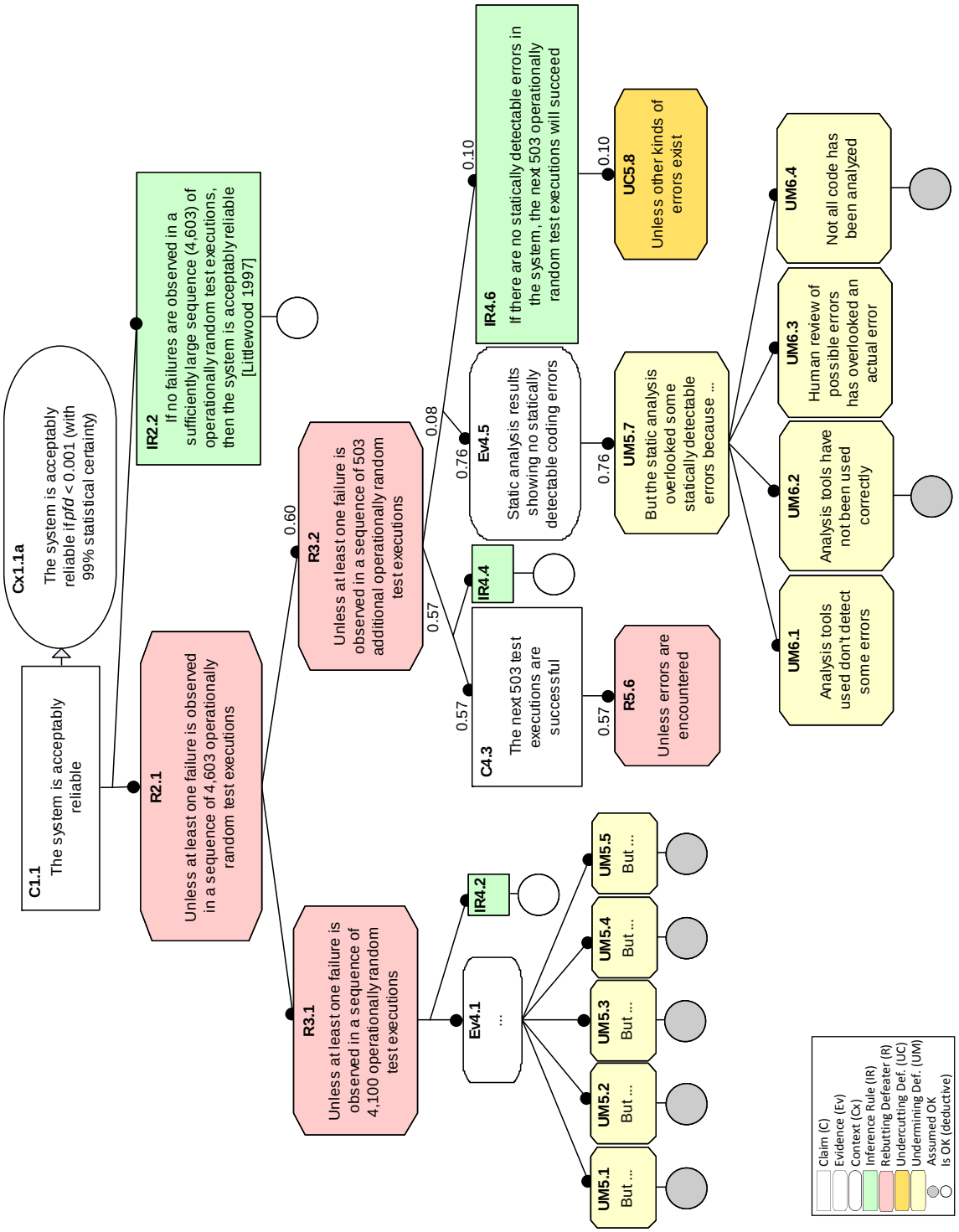


Figure 25: Using Static Analysis Evidence in the Statistical Testing Leg

3 Issues and Concerns

Various concerns have been raised about the practicality and utility of eliminative argumentation:

- What if a defeater has not been identified?
- What if all defeaters are not of equal importance?
- Isn't the number of defeaters too large to be practical for real systems?
- Why use eliminative induction rather than enumerative induction, such as Bayesian reasoning?
- What if a defeater is true?

Although we have discussed some of these issues so far in this report and in an earlier paper [Weinstock 2013], we discuss them definitively here.

3.1 Unidentified Defeaters

Since eliminative argumentation is based on identifying and eliminating defeaters, failure to identify relevant[18] defeaters would seem to give an inflated sense of confidence. But "failure to identify defeaters" is itself a doubt that is explicitly considered in an eliminative argument, since the possibility of unidentified defeaters is a standard undercutting defeater for inference rules that have eliminated defeaters as premises. For example, in the light-turns-on case, we explicitly stated that eliminating the three identified reasons for failure is insufficient if we have not identified all reasons for failure. Arguing that all defeaters have been identified is part of an eliminative argument and can be a source of residual doubt in the top-level claim.

An assessment process for reviewing an argument will need to take into account the possibility that some defeaters have not been identified. Our approach offers a consistent and theoretically exhaustive method for identifying sources of doubt, namely, by examining every inference rule for missing undercutting defeaters, every item of evidence for missing undermining defeaters, and every claim for possible counterexamples or failure modes. Nonetheless, assurance cases (and confidence maps) are inherently defeasible, which means that there is always the possibility that something has been omitted. A confidence map can reflect only what is known at a particular time. Eliminative argumentation is a way of thinking about and explaining why one should have confidence in a case or a claim. The concepts help in developing sound and complete arguments but do not guarantee that such arguments have been produced. We expect that this framework will help trained individuals to, in practice, reach agreement that all usefully important defeaters have been identified; of course, this remains to be demonstrated.

3.2 Relative Importance of Defeaters

In any set of defeaters, it is unlikely that they all seem equally important. Intuitively, it can seem that the elimination of one defeater (e.g., the failure of a system that controls the braking in an automobile) may have higher implications for confidence in safety than the elimination of another

[18] A *relevant* defeater is one whose elimination increases confidence in the validity of an associated claim, evidence, or inference rule.

(e.g., the failure of a system that controls the backup lights in that same automobile). If we are able to eliminate the first defeater and not the second, shouldn't we have higher confidence in a claim of system safety than if we are able to eliminate the second defeater and not the first?

Although incorporating a notion of relative importance of defeaters into our proposed grading of confidence may seem intuitively desirable, it is not essential. To understand why, consider the role of hazard analysis in the process of assuring system safety. Hazard analysis helps identify situations that may need to be mitigated for a system to be considered acceptably safe. However, not every potential hazard makes the cut to be represented in a safety case because mitigating potential hazards that are conceivable yet extremely unlikely and minimally impactful on system safety would contribute negligible increases to safety and unnecessarily increase the cost of developing the safety case or the system. So even though the hazards that are addressed in a safety case have different likelihoods and impacts, the point of representing them in the safety case is that they *all* must be demonstrably mitigated in order to establish sufficient confidence that the system is acceptably safe.[19] Just as a system developer would not represent extremely unlikely and minimally impactful safety hazards in a safety case as a way of justifying an increase in confidence, under our framework a system developer would not use the elimination of low-impact defeaters to justify an increase in confidence.

Although for practical purposes, the notion of relative importance of defeaters is not essential for using our framework to evaluate confidence, additional research on this issue is needed.

3.3 Eliminative Argumentation for Real Systems

Although the number of defeaters relevant to an argument seems to be quite large for a real system, the amount of relevant argument and evidence for a real system is also quite large. The question is whether the approach of identifying defeaters allows one to develop a more thorough and cost-effective basis for developing confidence in system behavior than current methods. This question cannot be answered until we have obtained practical experience in applying the approach, but our initial interactions with systems developers have been promising.

3.4 Why Use Eliminative Induction?

In enumerative induction, the number of confirming test results provides evidence of the statistical likelihood that future test results will also be confirming. In software, statistical testing is an example of this use of enumerative induction. But given a body of confirming evidence, enumerative induction, by itself, gives us no insight into possible conditions for system failure. In contrast, when a rebutting defeater is eliminated by test or analysis evidence, we have added to our knowledge about why a system works (cf. Popper's critical rationalism [Popper 1963]). In short, with eliminative induction, we learn something concrete about why a system works, and with enumerative induction, we at best only learn something statistical about the system (although statistical knowledge can be valuable). Moreover, from a psychological viewpoint, the search for defeaters helps to avoid confirmation bias—the tendency to overlook possible problems with a system and an assurance argument.

[19] The hazards that don't make the cut fall in the category of unidentified defeaters in terms of their effect on confidence.

An eliminative argument is useful to evaluate a system prior to its operational use; we need to think about what could go wrong before it does go wrong. For example, before a safety-critical system is put into operation, we must be able to reason about possible ways the system could be unsafe and why the system design eliminates or mitigates these possibilities. This is eliminative induction. An assurance case structure can help to structure such reasoning, but our addition of defeaters and inference rules provides a framework for explaining *why we believe* the system is safe, namely, because potential problems (rebutting defeaters) have been identified and eliminated *and* because possible problems with the argument (undercutting and undermining defeaters) have also been identified and eliminated. Much useful evidence and argumentation can be developed significantly in advance of having an actual operational system. Of course, once a system is operational, we can collect statistics on observed defects (enumerative induction) to predict its future operational reliability and safety.

Finally, we use eliminative induction informally all the time as a way of convincing ourselves, or others, of a claim's validity. More formal or structured uses also exist: logical proof by contradiction is one example. Others include Clarke's counterexample guided abstraction refinement for model checking [Clarke 2000] or resilience engineering [Limoncelli 2012].

None of this is to discount the very real importance of enumerative approaches to assurance, including Bayesian methods. In fact, we believe they coexist: one informs the other. If one wants to make probabilistic claims about system reliability, make a probabilistic claim and then elucidate the possibilities that would make you doubt the validity of the claim (as in our *pfd* claim earlier). On the other hand, if you want to determine what the actual operational reliability of a system is, then take repeated samples.

3.5 True Defeaters (Counterevidence)

Evidence that contradicts a top-level claim is called *counterevidence* in the Structured Assurance Case Metamodel (SACM) [OMG 2013], in the U.K. Ministry of Defence (MOD) safety management standard [U.K. MOD 2007], and by Hawkins and colleagues [Hawkins 2011]. The existence of counterevidence means that some defeater in a well-formed eliminative argument is true. Since any eliminative argument will, in practice, have some incompletely eliminated defeaters (i.e., residual doubt) or defeaters eliminated by assumption, counterevidence is always possible.

The search for counterevidence is an inherent part of eliminative argumentation because of our focus on identifying and eliminating defeaters. An eliminative argument can be thought of as a way of positing various reasons why counterevidence might exist and then showing that such counterevidence is unlikely or not possible. For example, consider an eliminative argument showing that a system is secure. To build confidence in the system's security, one might mount an extensive penetration testing exercise in which experts try to crack the system. In doing so, they might hypothesize various security weaknesses and attempt to exploit them. Each weakness could be represented as a rebutting defeater in an eliminative argument; the unsuccessful attempts to exploit that weakness (or a combination of such weaknesses) could be considered as evidence eliminating such defeaters, thereby increasing confidence in the system's security.

Given an eliminative argument for a system property, when counterevidence is found, there are basically two ways of dealing with it: make the counterevidence irrelevant (e.g., by changing the

system) or accept the counterevidence (as a reason for lack of confidence [residual doubt] that should be represented in the argument).

3.5.1 Making Counterevidence Irrelevant

There are three ways of making counterevidence irrelevant:

1. **Change the system** (so the revised system no longer behaves unacceptably). If the counterevidence (e.g., a usage failure) is due to a flaw in the system's implementation, one might remove the flaw, thereby eliminating the counterevidence (the system now behaves as specified).
2. **Change the specification (and argument)** (so the counterevidence is consistent with the revised specification). For example, suppose a system specification (i.e., a top-level claim) says that a system can be operated at below-freezing temperatures, but the system is found to fail at −40°. One might decide that instead of fixing the system, the specification should be changed, for example, to say that the system is only to be used in above-freezing conditions. Changing the specification means that some claim in the associated eliminative argument must also change.[20] After the change, any results obtained at below-freezing temperatures are irrelevant because such evidence eliminates no defeaters in the revised argument. In this approach, we are not challenging the validity of the counterevidence, that is, the fact that the system behaved in a certain way. We are changing the definition of failure.
3. **Attack the counterevidence.** Like any evidence, counterevidence has undermining defeaters. Counterevidence can be made irrelevant by showing that one or more of its undermining *defeaters* are true. For example:
   - If the counterevidence was gathered using an inappropriate or inaccurate evidence-collection technique, the counterevidence is invalid and therefore irrelevant. For example, if a test shows that the system fails at 20° but the thermometer being used is so inaccurate that the actual test condition was −10°, the failure evidence would be invalid and can be ignored.
   - The counterevidence might use a different definition of failure than the definition used when constructing the eliminative argument. This can occur, for example, when a system behaves in a way that surprises a user but is actually the specified behavior. In such a case, the initial interpretation that the counterevidence was valid is superseded by a more careful examination of the evidence.

3.5.2 Accepting Counterevidence

Counterevidence may be consistent with the residual doubt in a current argument, or it may require changes to properly reflect a new understanding of sources of doubt. There are three situations to consider, depending on the state of the defeater associated with the counterevidence:

1. **The defeater is uneliminated:** Consider the "light-turns-on" argument in Figure 6 and suppose the defeater R2.3, "Unless the switch is not connected," has not been eliminated. Not eliminating this defeater means we don't have full confidence in the top-level claim, so counterevidence that occurs because the switch is not connected would be consistent with the

[20] Such a change will usually require changes to subordinate defeaters and inference rules as well.

existence of the uneliminated defeater; that is, the failure is consistent with the residual doubt allowed by the argument. No change is needed to the argument structure.

2. **The defeater is completely eliminated:** The counterevidence could mean that an eliminative argument has incorrectly eliminated a defeater. In this case the argument needs to be modified to account for the new source of residual doubt represented by the counterevidence. For example, consider the "light-turns-on" argument. Suppose we are "completely confident" that the light will turn on because we have eliminated all reasons for doubt: the switch is connected, power is available, and the bulb has passed the "shake" test. The failure to turn on means that the argument has underestimated the amount of residual doubt and must be revised. Some defeater has been incorrectly eliminated (or is missing).

   Suppose that examination of the system shows that the bulb did not turn on because the glass is cracked; that is, the bulb is defective even though its filament is intact. This constitutes a new way in which a bulb can be defective. The recognition of this new mode of bulb defect is equivalent to saying that undercutting defeater UC4.3 in Figure 6 is true ("Unless the bulb can fail without releasing a filament fragment"). In the absence of any information about the bulb's physical state, the counterevidence shows that UC4.3 cannot be fully eliminated. The argument must be modified to show the need to assess whether the bulb is intact. Until new evidence is provided showing that the bulb is physically intact, this source of doubt reduces our confidence in the top-level claim.

3. **The defeater is partially eliminated:** If we are very confident that a defeater has been eliminated, counterevidence supporting the defeater would probably be considered surprising and would, at the minimum, require us to provide a different estimate of confidence in the defeater's elimination. On the other hand, if we didn't have much confidence in the defeater's elimination, the existence of the counterevidence could be viewed just as reflecting our understanding of sources of doubt.

If the counterevidence reflects a flaw in the system's design or implementation and it is not economical (or feasible) to correct the system immediately, then one might decide to live with the flaw by accepting the decreased confidence caused by the system flaw. For example, if a system misbehaves because of a timing defect that doesn't occur very often, one might accept the residual risk from not fully eliminating the defect. In eliminative argumentation, this is equivalent to concluding that an incompletely eliminated defeater does not significantly reduce overall confidence in some claim. Here we are using counterevidence as a basis for arguing that although a defeater is not fully eliminated, it is not a significant source of residual doubt. In eliminative argumentation, we focus on our inability to eliminate a defeater rather than whether a defeater is known to be true. Not eliminating a defeater can have the same impact on confidence as that of the defeater being true, namely, it decreases confidence in the top-level claim.

4 Connections to Other Work

Eliminative argumentation uses concepts from assurance cases [GSN 2011, ISO/IEC 2011, OMG 2013] (see the next section for a detailed discussion), eliminative induction [Cohen 1989] (see Section 4.4), and defeasible reasoning [Pollock 2008, Prakken 2010] (see Section 4.3.1). We discuss these (and other) connections in this section.

4.1 Comparison with Assurance Case Concepts

In this section, we briefly compare our notations and concepts with those of GSN [GSN 2011], claims-argument-evidence (CAE) [Adelard 2014], the ISO Assurance Case Standard [ISO/IEC 2011], the SACM [OMG 2013], and the concepts and requirements mentioned by the MOD guidance for structured safety cases [U.K. MOD 2007].

4.1.1 Claims

Our concept of a claim is the same as the GSN notion of a goal. The same symbology (a rectangle) is used in confidence maps and in GSN, but we label the element with a "C" instead of a "G." We make the same requirement that claims be stated as predicates.

The term *claim* is used in the CAE, ISO, SACM, and MOD documents with essentially the same meaning as for a GSN goal. But the ISO and MOD documents require that the top-level claim in a case be justified. For example, if the top-level claim defines "acceptable reliability" as $pfd < 10^{-3}$, these documents require that an argument be given to explain why this level of reliability is sufficient for the intended use of the system.

We impose no such requirement. The goal of eliminative argumentation is to justify a specific degree of confidence in a particular claim given certain evidence and inferences. Eliminative argumentation, as a framework for evaluating confidence, is not concerned with whether the top-level claim is useful or appropriate in some real-world context. If we wanted to know whether $pfd < 10^{-3}$ was an appropriate claim for some system usage context, we could construct a separate argument in which the top-level claim would perhaps be "$pfd < 10^{-3}$ is an appropriate reliability requirement for system X," and then go on to construct an eliminative argument supporting this claim.

4.1.2 Context

Context nodes are labeled "C" in GSN; the label "Cx" is used in confidence maps.

In the GSN standard, a context element provides additional information needed to understand an element of a case. In particular, it might specify the source of additional information (e.g., engineering documents) in which some aspects of the element are defined in detail. We use a context element more specifically as a rewriting rule. In this way, the definition provided by a context element is carried throughout the argument wherever the corresponding term or phrase is used.

4.1.3 Evidence

In GSN, evidence is called a solution and is written in a circle labeled with "Sn." In confidence maps, we use a rectangle with rounded corners as the graphical symbol and label these nodes with "Ev." In GSN, source data are represented in a solution node and the evidence assertion is captured in a separate claim supported by the solution node [Sun 2013]. In our notation, we combine these ideas using the form "[*Source data description*] showing [*evidence assertion*]."

Sun provides an extensive discussion of evidence in assurance cases [Sun 2012]. In addition, Hawkins and Kelly discuss ways of evaluating the sufficiency of evidence [Hawkins 2010]. Some of what they consider to be evidential deficiencies we capture as undermining defeaters, and some (those concerning the inferential use of evidence in an argument) we capture as undercutting defeaters. Hawkins and Kelly pose three questions to be considered in the process of evidence selection and justification:

1. Is the *type* of evidence capable of supporting the safety claim?
2. Is the particular *instance* of that type of evidence capable of supporting the safety claim?
3. Can the *instance* of that type of evidence be trusted to deliver the expected capability?

Among the types of evidence considered are testing, analysis, and review. They point out that understanding the role and limitations of a particular evidence type is essential to understanding its appropriate use in a case. From an eliminative argumentation perspective, the limitations of a particular type of evidence need to be captured in undercutting defeaters for inference rules using that type of evidence as premises. For example, the validity of conclusions based on models is typically limited by the accuracy of the model and assumptions about the system being modeled. These limitations would be captured in undercutting defeaters for inference rules using models as premises. In our earlier analysis of system reliability using static analysis evidence (Figure 25), we acknowledged a limitation of static analysis (namely, that it cannot detect all errors) with undercutting defeater UC5.8.

The doubts raised by Question 2 are specific doubts about the relevance of actual evidence of a given evidence type. For example, if the execution of operationally random tests is considered a type of evidence, the existence of 4,100 such successful tests is an instance of that evidence type. Whether 4,100 successful test executions is sufficient to imply acceptable reliability is a question about how much support such evidence contributes to a claim of system reliability. In eliminative argumentation, such questions can lead to the kind of argument restructuring we discussed in Section 2.5.1.

Question 3 also focuses on the actual evidence used in a case but raises doubts about whether it is what we have called *valid* evidence. In eliminative argumentation, deficiencies in the validity of evidence are captured in undermining defeaters. For example, in Section 0, we discussed undermining defeaters for the evidence described as consisting of 4,100 operationally random successful test executions. Such defeaters included questioning whether the test selection was truly random, whether the operational profile was accurate, and whether 4,100 tests actually were executed on the same system version and configuration.

Various terms are commonly used to characterize the role of evidence in an argument, including relevance, trustworthiness, reliability, and strength. In eliminative argumentation, *relevance* is captured both in the inference rule linking evidence to a defeater and in the inference rule giving

the reason why the elimination of the defeater is considered to support a claim. Weaknesses in these rules are typically captured in undercutting defeaters. Such weaknesses can be considered to explain doubts about the relevance of evidence.

Trustworthiness of evidence is defined by Hawkins and Kelly as "confidence that the item of evidence delivers its expected capability" [Hawkins 2010]. It is affected by many factors such as "'bugs' in the item of evidence presented, the rigor of review, the qualification of a tool adopted, the experience and competence of the personnel" [Sun 2012, p. 122–123]. Doubts about the trustworthiness and reliability of evidence are captured as undermining defeaters in an eliminative argument.

Strength of evidence sometimes refers to its inferential force; strong evidence presumably leaves no doubt about a claim. From an eliminative argument perspective, the inferential strength of evidence is captured by the inference rule explaining why a defeater is eliminated by the evidence and by whether this defeater is a significant source of doubt; only in this case does its elimination provide a significant increase in confidence in an associated claim. An inference rule has strong inferential force if it has no significant uneliminated undercutting defeaters.

Strength of evidence also sometimes refers to its validity; that is, highly believable evidence is considered strong evidence. From an eliminative argument perspective, such evidence has few or unlikely undermining defeaters. The OMG's SACM uses the term *strength* in this sense [OMG 2013].

The SACM provides many attributes for describing evidence in a common format. It provides many evidence attributes that could be considered in formulating undermining defeaters in a particular argument.

4.1.4 Inference Rule

Our notion of an inference rule is a more structured version of what GSN calls a "strategy" element and CAE calls an "argument" element. In GSN, a strategy element is used as an expository device to make explicit the argument approach being taken. For example, "Argue over identified hazards" is a possible strategy in an argument in which the mitigation of each hazard is considered to support a claim of safety. Such a strategy is equivalent to the inference rule "If all identified hazards are eliminated, the system is safe." The CAE notation uses its argument element in a similar fashion (see the comparison examples in SACM [OMG 2013]). The GSN standard explicitly talks about "undercutting" challenges to a strategy element. The Hawkins paper discusses the implied inference between a claim and supporting claims or evidence [Hawkins 2011].

4.1.5 Assumption

The way we deal with assumptions is different from what is required in the other notations. We view an assumption as a doubt that is eliminated without further evidence or argument. For convenience, we allow a claim, inference rule, or evidence to be assumed valid without having to identify associated defeaters that are then assumed to be eliminated. In principle, however, an assumption is always an eliminated doubt, and its elimination helps to strengthen confidence in a claim, inference rule, or evidence.

The MOD and SACM documents specify that an assumption is a *claim* whose validity is accepted without further evidence or argument. The ISO standard considers an assumption to be *evidence* that is used in supporting some claim. In GSN, an assumption is associated with a goal or a strategy element. It applies throughout the argument supporting that goal or strategy.

In eliminative argumentation as well as in these other notations, the idea is to identify elements of an argument that are accepted without further justification.

4.1.6 Rebutting Defeaters

No other notation uses the concept of a rebutting defeater, but claims in the other notations sometimes can be viewed as negated rebutting defeaters, that is, as something that has to be true if a supported claim is to be considered valid. For example, some claims in Figure 5 (e.g., "Light bulb is functional") are represented as rebutting defeaters in Figure 6 (e.g., "Unless the bulb is defective").

4.1.7 Undercutting Defeaters

The GSN standard explicitly talks about "undercutting" challenges to a strategy element. In a paper proposing the concept of "assurance claim points" as a means of determining confidence in an assurance case, Hawkins and his coauthors discuss assurance claim points associated with the implied inference between a claim and supporting claims or evidence [Hawkins 2011]. These assurance claim points are described as providing opportunities for identifying weaknesses in the inference. Such weakness are captured as undercutting defeaters in an eliminative argument.

The negation of an undercutting defeater is sometimes articulated as a claim in an assurance case, for example, a claim that "All hazards have been identified" can be combined with a series of claims indicating that various safety hazards have been eliminated or mitigated. In eliminative argumentation, the intent of such a claim is represented by the undercutting defeater "Unless all hazards have not been identified," which could be associated with the inference rule "If all hazards have been identified, the system is safe." In the assurance claim point approach, such a claim would be made in the confidence case associated with the "strategy" element stating that the approach to arguing safety is to argue over the elimination of hazards [Hawkins 2011].

4.1.8 Undermining Defeaters

We discussed the notion of undermining defeaters above when we discussed evidential concepts.

4.2 Eliminative Argumentation and Notions of Assurance Case Confidence

A safety case is evaluated to determine whether it sufficiently supports its safety claims. Such an evaluation characterizes the residual risk (likelihood of misbehavior) exposed by the safety case argument.

Eliminative argumentation, however, is not focused particularly on evaluating residual risk. We are focused on evaluating the justified degree of belief one can have in a top-level claim, given an argument containing particular evidence, inference rules, and uneliminated doubts. If the goal in a safety case is to argue that residual risk is less than some amount, then that should be part of the claim, just as our system reliability claim expressed what degree of statistical confidence we

wanted to have in the claim that *pfd* was less than 10^{-3}. We then use an eliminative argument to decide how much confidence we have in such a claim.

For us, confidence in a claim is a degree of belief. We want to know how much belief is justified by an offered supporting argument. An eliminative argument structure allows an analysis of the basis for an argument's credibility. It allows for an evaluation of confidence in the reasoning as well as confidence in the evidence. Of course, as our examples and discussion show, combining confidence evaluations to arrive at an overall confidence assessment in a top-level claim can be done in a variety of ways, and as yet, we have no particular basis for choosing one way rather than another. As we have mentioned, this is a subject of further research. The essential contribution of eliminative argumentation concepts is to provide a basis for arguing confidence.

Grigorova and Maibaum provide an overview of approaches to determining confidence in an assurance case [Grigorova 2014]. In one of these, Hawkins and colleagues [Hawkins 2011] separate the *assurance* case (focused on system properties and evidence derived from the system itself) from the *confidence* case. They associate *assurance claim points* (ACPs) with each relation between elements of the assurance case. For each ACP, they provide a separate argument asserting that the relation is well justified, for example, that the inference from evidence to claim is well justified (or as we would say, any doubts about the inference rule are identified and eliminated with further argument, evidence, or both). They argue that the separation is an important separation of concerns.

In eliminative argumentation, we do not provide a separate confidence argument; developing confidence is the purpose of the whole argument. The different kinds of defeaters account for different reasons for lacking confidence. An automated tool such as Trust-IT [Cyra 2008] could help in providing different views of a particular confidence map as well as in applying various algorithms for propagating confidence estimates up the argument tree.

4.3 Argumentation Literature

4.3.1 Defeasible Reasoning

While our use of defeaters is drawn from the concept and types of defeaters in the argumentation literature on defeasible reasoning, our reasoning approach—eliminative argumentation—and the function of defeaters in our approach deviates from the function of defeaters in defeasible reasoning [Pollock 1987]. In this section, we briefly discuss the major ways in which our use of defeaters is consistent with and deviates from defeasible reasoning.

Conceptually, a defeater in eliminative argumentation is the same as a defeater in the defeasible reasoning literature; defeaters are pieces of information that give us reason for doubting parts of an argument. Moreover, the defeaters in eliminative argumentation are taken from the three (and only three) kinds of defeaters in defeasible reasoning, which correspond to the three ways of attacking an argument: rebutting and undercutting defeaters [Pollock 1987] and undermining defeaters [Prakken 2010].

According to Pollock, reasoning is defeasible "in the sense that the premises taken by themselves may justify us in accepting the conclusion, but when additional information is added, that conclusion may no longer be justified" [Pollock 1987, p. 481]. In defeasible reasoning, an argument—a conclusion and its supporting premises—is formulated prior to and independent of the search for

additional information that could weaken or invalidate justification of the conclusion (i.e., defeaters).

In constructing an *eliminative* argument, however, the argument originates by identifying rebutting defeaters of a claim and providing subclaims or pieces of evidence that eliminate those defeaters. That is, the inferences in an eliminative argument are identified only after a defeater has been identified; in particular, a premise or piece of evidence can support a claim only if it has the potential to eliminate a rebutting defeater of the associated claim.

In defeasible reasoning, defeaters are not sought out and eliminated as a methodology for increasing confidence in an argument's conclusion; rather, defeaters are observed conditions in the world that force us to revise our justified belief in conclusions. The role of defeaters in defeasible reasoning is to weaken or invalidate our justification in accepting a conclusion, not to strengthen our justification by demonstrating their absence. In defeasible reasoning, defeaters are relevant to the original argument only if they are observed to be present—in other words, only if they appear to be true.

In contrast, defeaters in eliminative argumentation are, in principle, not actually observed in the world. We are not interested in how the presence of a defeater modifies our justification in accepting a conclusion. Instead, we are interested in the extent to which the absence of a defeater increases our confidence in a conclusion. We use defeaters as a conceptual, systematic framework for identifying *potential* conditions under which assurance claims are weakened or invalidated, and by eliminating defeaters (i.e., by demonstrating that potential defeaters cannot become actualized), we increase our confidence in the argument's conclusion.

4.3.2 Convergent and Linked Argument Structures

A *multi-legged* or *diverse* argument in the assurance case literature [Bloomfield 2003, Kelly 1998, Littlewood 2007, Weaver 2003] is conceptually similar to what has been called a *convergent* argument in the argumentation literature. In this literature, a distinction is made between *convergent* and *linked* argument structures [Govier 1987, Walton 1996].

In argumentation theory, a convergent argument [Beardsley 1950] consists of separate pieces of evidence for the same conclusion [Freeman 1991]. That is, in a convergent argument, each piece of evidence, or premise, is independently sufficient to support the conclusion; convergent premises do not require each other to imply the conclusion. These separate, independent arguments are "in principle, alternative defences of the same standpoint" [Van Eemeren 1992]. "Each is an independent evidential route for supporting the conclusion" [Walton 2006, p. 140].

In a convergent argument, it is not necessary for all premises to hold, but is such an argument stronger if more than a single premise holds? Some theorists say that if only one of two premises in a convergent argument holds, "we would have a weaker argument"; that is, having both premises hold could be viewed as making the argument stronger even though only one premise is, strictly speaking, sufficient [Groarke 1997]. This view is consistent with our probabilistic view that in a multi-legged argument, when existing legs do not provide complete confidence, an added

leg can provide additional confidence[21] (as long as the added leg provides an independent source of confidence, of course).

In a "linked" argument, two or more premises work together to imply the argument's conclusion [Thomas 1973]. "All the component single argumentations are, in principle, necessary for a conclusive defence of the standpoint" [Van Eemeren 1992, p. 77]. Linked arguments are sometimes conceptualized as consisting of a "set of premises" [Gordon 2006] rather than as single premises supporting a conclusion because the conclusions of linked arguments are "defensible only if all of their premises hold."

For example, the following three premises would form a linked pattern of support; they work together as a set to provide support for the conclusion that Side C of Triangle ABC has length 5: (1) Side A has length 3; (2) Side B has length 4; and (3) Sides A and B are at right angles to each other. Taken independently, none of these premises implies the conclusion. The conclusion follows only if all three premises hold.

Of course, if only some of the premises in a linked argument are known to hold (and the status of the others is unknown), one might still say there is some support for the conclusion. According to Walton, "In a linked argument [with two premises], if one premise is deleted, the other by itself offers much less evidential support for the conclusion than the two do together" [Walton, 2006]. In other words, the remaining premise can be viewed as offering some support for the conclusion. In eliminative argumentation, we take this approach to measuring support when we count the number of eliminated defeaters (in a linked argument) with the idea that eliminating two out of three defeaters (2|3) provides more support for a claim than 1|3 and less than 3|3.

Although the differences between linked and convergent argumentation structures are obvious in principle, in practice, the choice is not always obvious. For example, suppose that in developing a keypad for a medical device, we subject its design to an expert review to gain confidence that the design will not be a source of user input errors. Then, after the design has been implemented, suppose operational testing shows a negligible user error rate. Given each kind of evidence singly, or in combination, how much confidence should we have that the keypad's design is good, in the sense that the design is not a cause of user input errors?

The answer depends on whether we think both items of evidence are *required* if we are to have complete confidence in keypad design (a linked argument for confidence) or if we would be satisfied if just one of the items of evidence was provided (a convergent argument for confidence). We make these alternatives explicit in the example confidence maps shown in Figures 26 and 27.

Figure 26 shows a convergent argument (i.e., a multi-legged argument). We have assigned some confidence values to key elements of the map. We consider that the design review was done thoroughly, so we have 98% confidence that R2.1 is eliminated. However, a design review is not the most authoritative method for ensuring that the keypad design discourages user input errors, because even a well-conducted design review might underestimate certain keypad usage factors that are present in the operational environment. We reflect the inferential weakness associated with a design review by estimating that undercutting defeater UC3.2 is only 80% eliminated, meaning

[21] Of course, if one of the legs should happen to provide complete confidence in a claim, none of the other legs would be needed, so they would not be viewed as adding confidence.

we have only 80% confidence in the conclusion of the inference rule. Consequently, our confidence in the keypad design, based just on the result of the design review, is only (0.98)(0.80) = 0.78.

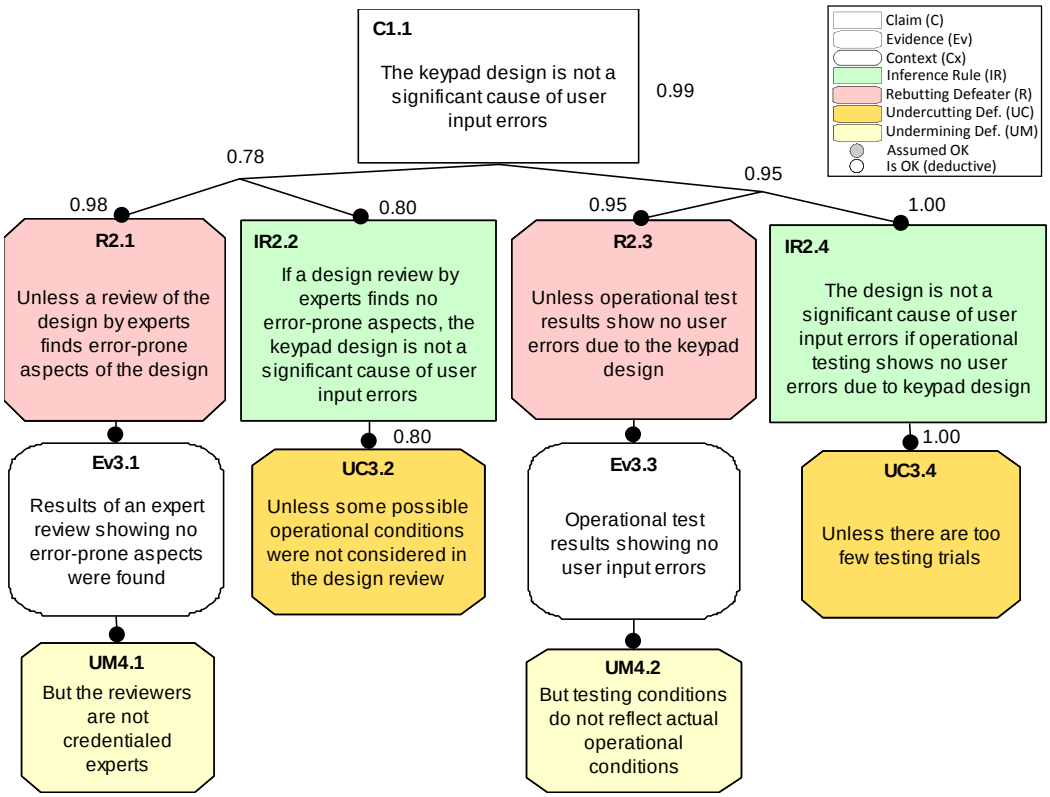


Figure 26: A Convergent Argument Structure

Operational test results (Ev3.3) are probably more indicative of the quality of the keypad design as long as the tests are conducted appropriately (UM4.2) and there are enough trials to provide an adequate sample of usage (UC3.4). For this example, we assume that enough trials were conducted and that we are very confident (95%) that the test conditions reflected actual operational conditions. So the successful operational testing results give us (0.95)(1.00) = 0.95 confidence in the claim.

Since independent evidence is associated with both "legs" of this argument, our confidence in the claim can be calculated as 1 – (1 – 0.76)(1 – 0.95) = 0.99. Note that in convergent argument structures, joint confidence in a claim is *at least* the *maximum* of the confidence associated with each leg individually.

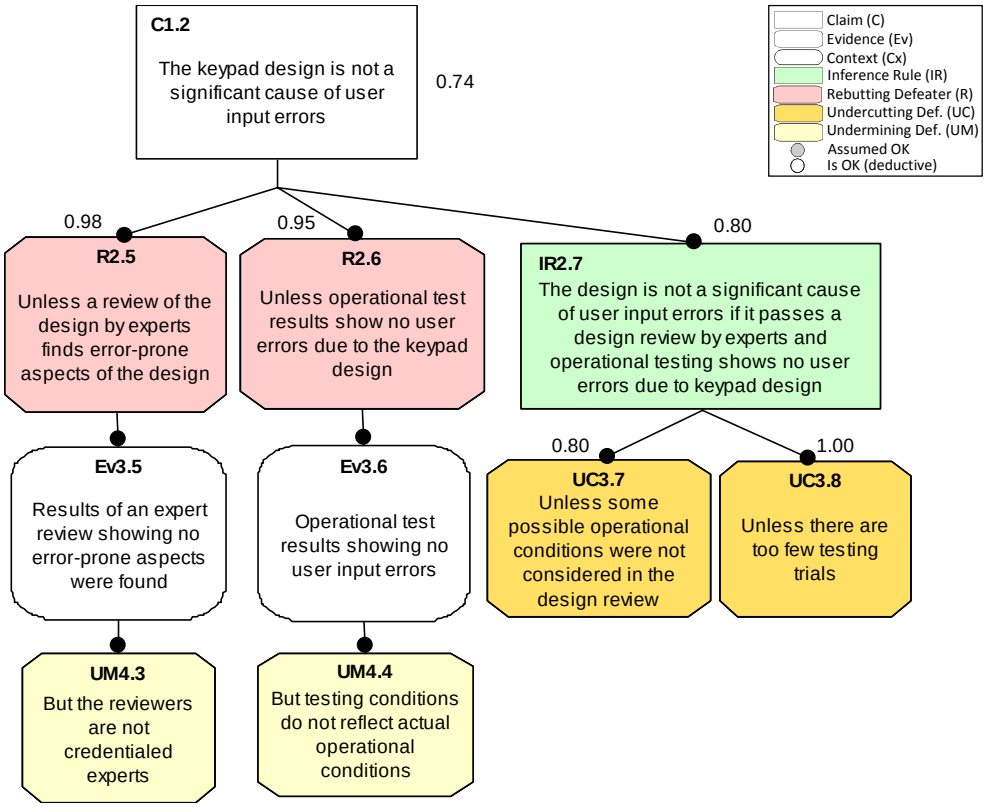


Figure 27: A Linked Argument Structure

If we feel that both rebutting defeaters must be eliminated to have complete confidence in the claim, we need to use a linked argument structure such as that shown in Figure 27. Here we have the same defeaters with the same probabilities of elimination, but our inference from elimination of both rebutting defeaters to a conclusion about the keypad design (IR2.7) is weakened by its two undercutting defeaters. The linked argument structure gives a confidence of (0.98)(0.95)(0.80) = 0.74. Note that in a linked argument structure, joint confidence in a claim is *at most* the *minimum* of confidence associated with each supporting element.

The choice of a linked or convergent argument structure is not entirely free. The light-turns-on example cannot be turned into a convergent argument. In a convergent structure, the elimination of each rebutting defeater is treated as an independently sufficient reason for believing the claim. This requires an inference rule such as "If the light bulb is not defective, the light will turn on," but this rule does not guarantee its conclusion if the switch is not connected or there is no power. Such conditions could be proposed as undercutting defeaters showing that the rule is insufficient, but they would be improper undercutting defeaters since their truth does not leave the conclusion uncertain. As we argued in Section 2.5.2, this means such conditions must be specified as rebutting defeaters. This problem does not arise for the inference rules and undercutting defeaters in our keypad design example.

In summary, the linked and convergent argumentation structures occur naturally in eliminative arguments. Deciding which structure to use is usually intuitively obvious (would absence of some evidence totally destroy confidence in a claim? If so, a linked argument structure is appropriate.)

But more research is needed to fully articulate why it is better to choose one structure or the other in some circumstances.

4.4 Philosophy

The notion of establishing confidence in a hypothesis is a long-standing problem in philosophy. Schum extensively discusses various approaches that have been developed [Schum 2001]. In particular, he discusses the use of eliminative induction for this purpose, citing and expanding upon L. Jonathan Cohen's work on Baconian probabilities [Cohen 1989]. Our use of defeaters as a basis for establishing confidence in a claim builds on this work but deviates from it in significant ways. In this section, we briefly discuss Cohen's notions and our deviations.

Cohen is primarily interested in using eliminative induction as a method for determining which hypothesis is most favorable among a set of possibilities. For Cohen, the notion of "evidence" refers to the result of examining whether a hypothesis is favored when evaluated under various *conditions* that have the potential to cast doubt on the hypotheses (e.g., variations in temperature, humidity, shock, electromagnetic interference). He defines Baconian probability as B(H, E) = i/n,[22] where E represents the number of tested conditions (n) as well as whether test results are deemed to favor or disfavor hypothesis H. In his formulation, results are available for all n test conditions, and i is the number of favorable results. B(H, E) = i/n represents the tendency of the evidence to favor the hypothesis.

Because all of the individual tests ("n") have been performed in Cohen and Schum's use of B(H, E) = i/n, each test will either favor or not favor the hypothesis. Any unfavorable test is implicitly represented in the form of $n - i$. For Cohen and Schum, unfavorable tests do not serve to diminish or cancel out the favorability of the evidence as determined by the number of successful observations (i). For their use in testing hypotheses relating to the real world, it is rarely the case that a hypothesis completely explains a particular phenomenon, so it is acceptable from their perspective to have some aspects of the evidence not favor a hypothesis while still retaining the amount of favorability that was acquired by the other aspects of the evidence.

In eliminative argumentation, defeaters are equivalent to what Schum calls test conditions. Whereas Cohen and Schum's notion of evidence encompasses both test conditions and the results of subjecting a hypothesis to these conditions, we treat defeaters separately from observed results.

In our use of the Baconian probability, B(H, E) = i/n, $n - i$ does not equal the number of unfavorable test results; rather, it equals the number of defeaters whose status is unknown. Whereas Cohen and Schum view test conditions producing unfavorable results as simply diminishing the hypothesis's favorability, in a strict view of eliminative argumentation, such test results can be viewed as invalidating the claim. In other words, Cohen and Schum's use of Baconian probabilities allows for having belief in both a hypothesis (H) and a competing hypothesis (H^C), whereas it is logically impossible in our use of Baconian probabilities to have belief in both a claim (H) and a counterclaim (~H).

[22] The notation i/n is read as "i out of n"; it does not represent division. In our use of these ideas in this report, we write this as "$i|n$" to avoid confusion with the division operation. In this section, since we are discussing Cohen's ideas, we use his notation.

5 Conclusion

Eliminative argumentation is a structured way of thinking about how to develop and evaluate arguments, particularly assurance arguments about properties of systems. Our framework provides a theoretical basis (eliminative induction) for deciding the extent to which the weaknesses in an assurance case (sources of doubt) are sufficiently small that we can have confidence in the main claim.

We are not the first to note sources of unsoundness in arguments, namely, questionable inference rules and weaknesses in proffered evidence. However, the notions of eliminative argumentation and, in particular, the different kinds of defeaters, provide a helpful way of thinking about how to formulate and evaluate arguments. The focus on eliminating defeaters (as opposed to searching for supporting evidence) provides a way of avoiding so-called "confirmation bias," in which people seek out support for their beliefs rather than seeking to defend their beliefs against possible attacks. Of course, our framework does not ensure that one will find all the relevant defeaters, but it is more likely that one will find them if one is looking for them.

We have suggested a variety of ways to synthesize measures of confidence from measures of constituent confidence. More work is needed to determine what methods of confidence calculation are most useful for given purposes and what practical effect various confidence numbers may have.

We continue to explore the practical implications of this approach. But at least as a mental model for use in constructing and evaluating arguments, we find the notions of eliminative argumentation helpful.

Appendix Calculating the Probability of Failure

Running a sequence of successful operationally random tests gives an estimate of the upper bound on the probability of failure on demand (*pfd*). The more tests that run successfully, the lower the estimate of *pfd* and the more confidence one has in the estimate. For the example used in this report, we wanted to know the upper bound on *pfd* (with 99% confidence) when a certain number of successful tests was run. The various values are shown in Table 3. We highlight the 4,100 row because this is the value we used in our statistical testing example (Section 2.5.1.3). This table was calculated with the Minitab statistical package (www.minitab.com), using the standard one-proportion hypothesis test without any assumption of normality and with alpha error set at 0.01 to reflect 99% confidence.

Table 3: Upper Bound on pfd for Successful Test Runs

Number of Failure-Free Test Runs	Upper Bound on *pfd*	Probability of Remaining Tests Succeeding	Number of Remaining Tests
1,000	0.004595	0.00%	3,603
2,000	0.002300	0.25%	2,603
3,000	0.001534	8.54%	1,603
3,500	0.001315	23.42%	1,103
3,700	0.001244	32.50%	903
3,800	0.001211	37.79%	803
3,900	0.001180	43.60%	703
4,000	0.001151	49.93%	603
4,100	0.001123	56.83%	503
4,200	0.001096	64.28%	403
4,300	0.001070	72.30%	303
4,400	0.001046	80.86%	203
4,500	0.001023	89.99%	103
4,600	0.001001	99.70%	3
4,603	0.001000		

References

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

[Adelard 2014]
Adelard. *Claims, Arguments and Evidence (CAE)*. Adelard, LLP, 2014. http://www.adelard.com/asce/choosing-asce/cae.html

[Beardsley 1950]
Beardsley, M. C. *Practical Logic*. Prentice-Hall, 1950.

[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.

[Clarke 2000]
Clarke, E., Grumberg, O., Jha, S., Lu, Y., & Veith, H. "Counter-example Guided Abstraction Refinement." *Computer-Aided Verification Conference, Lecture Notes in Computer Science 1855* (2000): 154–169.

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

[Cowles 2009]
Cowles, K., Kass, R., & O'Hagan, T. *What Is Bayesian Analysis?* International Society for Bayesian Analysis (ISBA), 2009. http://bayesian.org/Bayes-Explained

[Cyra 2008]
Cyra, L. & Górski, J. "Supporting Expert Assessment of Argument Structures in Trust Cases." Presented at the 9th International Probabilistic Safety Assessment and Management Conference PSAM. Hong Kong, China, May 2008. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.163.1409&rep=rep1&type=pdf

[Freeman 1991]
Freeman, J. B. *Dialectics and the Macrostructure of Arguments: A Theory of Argument Structure*. Foris Publications, 1991.

[Goodenough 2012]
Goodenough, J. B., Weinstock, C. B., & Klein, A. Z. *Toward a Theory of Assurance Case Confidence* (CMU/SEI-2012-TR-002). Software Engineering Institute, Carnegie Mellon University, 2012. http://resources.sei.cmu.edu/library/asset-view.cfm?assetid=28067

[Gordon 2006]
Gordon, T. F. & Walton, D. "The Carneades Argumentation Framework," 195–207. *Proceedings from COMMA '06: Computational Models of Argument*. Edited by P. E. Dunne & T. J. Bench-Capon. Liverpool, U.K., Sep. 2006. IOS Press, 2006.

[Govier 1987]
Govier, T. *A Practical Study of Argument*, 2nd ed. Wadsworth Publishing, 1987.

[Grigorova 2014]
Grigorova, S. & Maibaum, T. S. E. "Argument Evaluation in the Context of Assurance Case Confidence Modeling," 485–490. *2014 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW)*, Naples, Nov. 2014. IEEE Computer Society Press, 2014.

[Groarke 1997]
Groarke, L. A., Tindale, C. W., & Fisher, L. *Good Reasoning Matters! A Constructive Approach to Critical Thinking.* Oxford University Press, 1997.

[GSN 2011]
GSN Standardisation Committee. *GSN Community Standard Version 1.* Origin Consulting, 2011.

[Hawkins 2010]
Hawkins, R. & Kelly, T. "A Structured Approach to Selecting and Justifying Software Evidence," 1-6. In *5th IET International Conference on System Safety System Safety 2010.* Manchester, Oct. 2010. IEEE Computer Society, 2010.

[Hawkins 2011]
Hawkins, R., Kelly, T., Knight, J., & Graydon, P. "A New Approach to Creating Clear Safety Arguments," 3–23. *Advances in Systems Safety: Proceedings on the Nineteenth Safety-Critical Systems Symposium*. Bristol, U.K., Feb. 2010. Edited by C. Dale & T. Anderson. Springer, 2011.

[Hawkins 2013]
Hawkins, R., Habli, I., Kelly, T., & McDermid, J. "Assurance Cases and Prescriptive Software Safety Certification: A Comparative Study." *Safety Science 59* (Nov. 2013): 55–71.

[Hawthorne 2012]
Hawthorne, J. "Inductive Logic." *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta. Stanford University, 2012. http://plato.stanford.edu//entries/logic-inductive/

[ISO/IEC 2011]
ISO/IEC. *Systems and Software Engineering—Systems and Software Assurance—Part 2: Assurance Case*. ISO, 2011.

[Kelly 1998]
Kelly, T. *Arguing Safety: A Systematic Approach to Safety Case Management.* Department of Computer Science, University of York, 1998.

[Limoncelli 2012]
Limoncelli, T. "Resilience Engineering: Learning to Embrace Failure." *ACM Queue 10*, 9 (Sep. 2012). http://queue.acm.org/detail.cfm?id=2371297

[Littlewood 1997]
Littlewood, B. & Wright, D. "Some Conservative Stopping Rules for the Operational Testing of Safety-Critical Software." *IEEE Transactions on Software Engineering 23*, 11 (Nov. 1997): 673–683.

[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 System." *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 Encyclopedia of Philosophy: Defeasible Reasoning*. Stanford University, 2005. http://plato.stanford.edu/entries/reasoning-defeasible/

[OMG 2013]
Object Management Group. *Structured Assurance Case Metamodel (SACM).* Object Management Group, 2013. http://www.omg.org/spec/SACM

[Pollock 1987]
Pollock, J. L. "Defeasible Reasoning." *Cognitive Science 11*, 4 (Feb. 1987): 481–518.

[Pollock 2008]
Pollock, J. "Defeasible Reasoning," 451–469. *Reasoning: Studies of Human Inference and Its Foundations*. Edited by J. E. Adler & L. J. Rips. Cambridge University Press, 2008.

[Popper 1963]
Popper, K. *Conjectures and Refutations: The Growth of Scientific Knowledge.* Routledge, 1963.

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

[PVS 2014]
Program Verification Systems. *Examples of Errors Detected by the V511 Diagnostic*. PVS, June 2014.

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

[Sun 2012]
Sun, L. *Establishing Confidence in Safety Assessment Evidence.* University of York, 2012.

[Sun 2013]
Sun, L. & Kelly, T. "Elaborating the Concept of Evidence in Safety Cases," 111–126. *Assuring the Safety of Systems: Proceedings of the Twenty-first Safety-Critical Systems Symposium*. Bristol, U.K., Feb. 2013. Edited by C. Dale & T. Anderson. Safety Critical Systems Club, 2013.

[Thomas 1973]
Thomas, S. N. *Practical Reasoning in Natural Language.* Prentice-Hall, 1973.

[U.K. MOD 2007]
U.K. Ministry of Defence. *Defence Standard 00-56 Safety Management Requirements for Defence*. U.K. MOD, 2007.

[Van Eemeren 1992]
Van Eemeren, F. H. & Grootendorst, R. *Argumentation, Communication, and Fallacies: A Pragma-Dialectical Perspective.* Lawrence Erlbaum Associates, 1992.

[Walton 1996]
Walton, D. *Argument Structure: A Pragmatic Theory.* University of Toronto Press, 1996.

[Walton 2006]
Walton, D. *Fundamentals of Critical Argumentation.* Cambridge University Press, 2006.

[Weaver 2003]
Weaver, R., Fenn, J., & Kelly, T. "A Pragmatic Approach to Reasoning About the Assurance of Safety Arguments," 55–67. *Conferences in Research and Practice in Information Technology, Proceedings from SCS '03: The 8th Australian Workshop on Safety Critical Systems and Software.* Canberra, Australia, Oct. 2003. Edited by P. Lindsay, & T. Cant.: Australian Computer Society, 2003.

[Weinstock 2013]
Weinstock, C. B., Goodenough, J. B., & Klein, A. Z. "Measuring Assurance Case Confidence Using Baconian Probabilities," 7–11. *1st International Workshop on Assurance Cases for Software-Intensive Systems (ASSURE)*. San Francisco, CA, May 2013. IEEE, 2013.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE February 2015	3. REPORT TYPE AND DATES COVERED Final
4. TITLE AND SUBTITLE Eliminative Argumentation: A Basis for Arguing Confidence in System Properties		5. FUNDING NUMBERS FA8721-05-C-0003
6. AUTHOR(S) John B. Goodenough, Charles B. Weinstock, and Ari Z. Klein		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Software Engineering Institute Carnegie Mellon University Pittsburgh, PA 15213		8. PERFORMING ORGANIZATION REPORT NUMBER CMU/SEI-2015-TR-005
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) AFLCMC/PZE/Hanscom Enterprise Acquisition Division 20 Schilling Circle Building 1305 Hanscom AFB, MA 01731-2116		10. SPONSORING/MONITORING AGENCY REPORT NUMBER
11. SUPPLEMENTARY NOTES		
12A DISTRIBUTION/AVAILABILITY STATEMENT Unclassified/Unlimited, DTIC, NTIS		12B DISTRIBUTION CODE

13. ABSTRACT (MAXIMUM 200 WORDS)

Assurance cases provide a structured method of explaining why a system has some desired property, for example, that the system is safe. But there is no agreed approach for explaining what degree of confidence one should have in the conclusions of such a case. This report defines a new concept, *eliminative argumentation*, that provides a philosophically grounded basis for assessing how much confidence one should have in an assurance case argument. This report will be of interest mainly to those familiar with assurance case concepts and who want to know why one argument rather than another provides more confidence in a claim. The report is also potentially of value to those interested more generally in argumentation theory.

14. SUBJECT TERMS assurance case, confidence, defeasible reasoning, eliminative induction, software assurance	15. NUMBER OF PAGES 71
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