

# A Taxonomy of Fallacies in System Safety Arguments

William S. Greenwell<sup>1</sup> C. Michael Holloway<sup>2</sup> John C. Knight<sup>1</sup>

<sup>1</sup>*Department of Computer Science, University of Virginia  
151 Engineer's Way, Charlottesville, VA 22904-4740, USA  
Tel: +1 (434) 982-2216 Fax: +1 (434) 982-2214  
{greenwell | knight}@cs.virginia.edu*

<sup>2</sup>*NASA Langley Research Center  
MS 130 / 100 NASA Road, Hampton, VA 23681-2199, USA  
Tel: +1 (757) 864-1701 Fax: +1 (757) 864-4234  
c.m.holloway@nasa.gov*

## Abstract

A system's safety argument is intended to show that the system is acceptably safe to operate in a given environment. If that argument is fallacious, the system may be prone to hazardous modes of operation that could contribute to accidents. We conducted a case study of three industrial safety cases to determine the frequency and nature of fallacious reasoning in system safety arguments. Our results suggest that the frequency of logical fallacies in these arguments is significant and that they follow common themes. To avoid these fallacies, developers must be aware of them when they create safety arguments, and regulators and investigators must know how to discover them when they review those arguments. We present a taxonomy of logical fallacies tailored to system safety cases to assist developers and regulators in these tasks and then demonstrate the taxonomy by applying it to the three safety cases from our case study.

## 1. Introduction

Protection of the public interest requires that safety-critical systems operate at acceptable levels of risk. Safety cases have evolved as a valuable approach to structuring the argument that a safety-critical system is acceptably safe to use, and they have been developed in several application domains [1]. Failure of a safety-critical system indicates that the risk of operating the system might be higher than previously thought. In the most general sense, a failure is either the result of an event that was *anticipated* but which was predicted to have a probability of occurrence below some critical threshold, or it was the result of an *unanticipated* event. If an unanticipated event occurred, then there must have been a defect in the safety case since the total risk exposure for operation of the system would have been based entirely on the random occurrences of anticipated events.

Our prior analyses of accidents involving safety-critical systems, including the minimum safe altitude warning (MSAW) system whose failure contributed to a major commercial aviation accident, suggested to us that system safety arguments sometimes invoke incomplete or inherently faulty reasoning [6]. If undetected, these fallacies could lead to overconfidence in a system and the false belief that the system's design has obviated or will tolerate certain faults. That a safety case might contain a flaw is to be expected, but it is important to bring attention to the problem and to remove defects to the extent possible.

In this paper, we discuss a very specific source of possible defects in safety cases, namely *flawed arguments*. Informal review of the safety cases built for a set of important safety-critical applications showed that flawed arguments were present in each case. Based upon our observations, we adapted existing taxonomies of fallacious inferences to produce one that is tailored to system safety arguments. We then demonstrate the taxonomy by applying it to the safety cases we reviewed earlier.

The paper is organized as follows. Section 2 motivates the work by describing our case study to determine the frequency and nature of logical fallacies in system safety arguments. Section 3 then presents the fallacy taxonomy. Section 4 demonstrates the taxonomy by applying it to the safety cases discussed earlier. Finally, Section 5 concludes the work.

## **2. Fallacies in System Safety Arguments**

To test our hypothesis that there exist common types of logical fallacies in system safety cases and to determine what those fallacies are, we conducted a case study of publicly available safety cases from various industries. We were surprised at the rarity with which safety cases are made available for public scrutiny, but we were able to obtain eight safety cases on the topics of air traffic management, automotive engineering, commuter rail transit, electrical engineering, nuclear engineering, and radioactive waste storage [3]. Of these, we selected three safety arguments for our case study: the EUROCONTROL (EUR) Reduced Vertical Separation Minimums (RVSM) Pre-Implementation Safety Case, the Opalinus Clay geological waste repository safety case, and the EUR Whole Airspace Air Traffic Management (ATM) Safety Case. We selected these three safety cases from the eight we considered because their organization made it easier for us to assess the merits of their arguments without possessing expert knowledge of the systems in question or the relevant domains. The EUR RVSM and whole airspace safety cases are preliminary and do not necessarily reflect final engineering standards; however is it still appropriate to examine the arguments for fallacies so that those fallacies can be addressed prior to implementation.

Two of the authors read and independently evaluated each of the three safety cases selected for the study. The purpose of this evaluation was two-fold: (1) to determine whether fallacies appear in these safety arguments with any significant frequency; and (2) to identify the types of fallacies committed. Both of the reviewers had at least a basic understanding of fallacious reasoning from classical philosophical literature and drew from that knowledge in performing their evaluations. When a reviewer came across what he believed to be faulty reasoning in an argument, he highlighted the relevant section and recorded a brief note explaining why the reasoning was problematic. Upon completing their evaluations, the reviewers compiled their results into a comprehensive set of fallacies for each of the safety cases and then achieved a consensus as to which of the comments they made were indicative of fallacious reasoning in the arguments. The following sections summarize those results for each safety case. Note that the goal of this study was to examine the form of the safety argument, not to evaluate the safety of the associated systems.

### **2.1 EUR Reduced Vertical Separation Minimums (RVSM) Safety Case**

The EUR Reduced Vertical Separation Minimums (RVSM) Pre-Implementation Safety Case concerns a proposed reduction in the minimum amount of vertical distance that must be present between any two aircraft operating in EUR airspace. RVSM would accommodate a greater density of air traffic and would thus enable EUR to meet an expected increase in demand for air travel over the next several years. The RVSM safety case is organized as a natural language document but provides a graphical version of the argument in Goal Structuring Notation (GSN) in an appendix [7]. To make the study tractable, we chose to limit our evaluation to the GSN portion of the safety case.

Table 1 summarizes the fallacies the reviewers identified in the EUR RVSM safety case with one column for each reviewer. Together, the two reviewers identified 29 fallacies in the nine pages of the argument they evaluated, or about three fallacies per page of GSN. Relevance fallacies were the most prevalent in the argument and accounted for two-thirds of the fallacies identified.

Although the purpose of employing two reviewers in the case study was to assemble a more complete set of fallacies and not to examine the consistency between reviewers, the disparity between the results of each reviewer is significant. Differences are present both in the quantities and types of fallacies that each reviewer identified, the most notable of which concerns the using-the-wrong-reasons fallacy. All but one of the instances of using the wrong reasons concerned a fallacious inference that the argument

**Table 1: Summary of Fallacies in the EUR RVSM Safety Case**

Fallacy	Reviewer A	Reviewer B	Total <sup>a</sup>
Using the Wrong Reasons	5	15	16
Drawing the Wrong Conclusion	3		3
Red Herring	1		1
Fallacious Use of Language	2	2	4
Hasty Inductive Generalization	4		4
Omission of Key Evidence		1	1
Total	15	18	29

a. If both reviewers identified the same instance of a fallacy, that instance is only reflected once in the total.

repeatedly invoked. Reviewer A flagged only the first few of these instances before choosing to ignore them while reviewer B flagged them all. Moreover, both reviewers identified two instances of fallacious use of language due to ambiguity in the argument; however the specific instances they identified did not overlap, suggesting that they had trouble agreeing upon which language was ambiguous or misleading. Section 4.1 discusses the particular inferences that exhibited these fallacies in greater detail. Finally, reviewer A identified a greater variety of fallacies than did reviewer B, which may be due to A's more extensive background and experience with logic and argumentation.

## 2.2 Opalinus Clay Geological Repository Safety Case

The Opalinus Clay safety case concerns the feasibility of constructing a long-term radioactive waste storage facility within the Opalinus Clay—a geological formation in the Zürcher Weinland of Switzerland. The argument claims that the Clay is sufficiently stable to enable the facility to meet its safety requirements for at least the next ten million years [10]. The safety case is written in bulleted natural language with major safety claims enumerated as subsections accompanied by their corresponding arguments. It includes a variety of arguments for the feasibility and safety of the facility, including:

*“...multiple arguments for safety that:*

- *demonstrate safety and compliance with regulatory protection objectives;*
- *use indicators of safety that are complementary to those of dose and risk and that show that radionuclide releases and concentrations due to the repository are well below those due to natural radionuclides in the environment;*
- *indicate that the actual performance of the disposal system will, in reality, be more favorable than that evaluated in quantitative analyses; and*
- *no issues have been identified that have the potential to compromise safety” [10].*

Both reviewers agreed that the Opalinus Clay safety case was the most compelling of the three arguments they reviewed. Indeed, reviewer B did not identify any fallacies in the argument while reviewer A identified only three, one of which was later determined to be valid reasoning. Table 2 shows the fallacies identified by reviewer A. At 22 pages, the Opalinus Clay argument had the lowest rate of the arguments we considered with only one fallacy every 11 pages. Section 4.2 discusses the specific instances of fallacious reasoning and the falsely identified instance of arguing from ignorance in greater detail.

**Table 2: Summary of Fallacies in Opalinus Clay Repository Safety Case**

Fallacy	Reviewer A	Reviewer B <sup>a</sup>	Total
Arguing from Ignorance	1 <sup>b</sup>		0
Omission of Key Evidence	2		2
Total	3	0	2

- a. Reviewer B did not identify any fallacies in this safety case.
- b. Through later discussion, the reviewers agreed that the reasoning labeled as arguing from ignorance was not actually fallacious. Therefore, this fallacy is not reflected in the total.

### 2.3 EUR Whole Airspace Preliminary Safety Case

The EUR Whole Airspace Air Traffic Management (ATM) System Safety Case preliminary study was conducted to evaluate “the possibility of developing a whole airspace ATM System Safety Case for airspace belonging to EUROCONTROL member states” [8]. The study proposes arguments for preserving the current safety level of EUR airspace under a unified air traffic organization instead of the patchwork of organizations that comprise EUR today. We are aware that the arguments presented in the EUR safety case are preliminary; nevertheless, we think it is appropriate to examine them because operational arguments will likely be derived from them. Like the RVSM safety case, the report presents mostly natural language arguments but does make use of GSN in some places. Again, we chose to limit our review to the major GSN elements of the report. These included two separate arguments for the safety of the whole airspace: one based on arguing over individual geographic areas and one based on reasoning about the whole airspace ATM rules. Both arguments shared the same top-level goal that “the airspace is safe.”

Table 3 contains the results of the reviewers’ evaluations of the EUR Whole Airspace argument, which reflect the combined fallacies in both the argument over geographic regions and the argument for the safe implementation of whole airspace rules. Together, the arguments spanned two pages, giving a rate of seven fallacies per page. Acceptability and relevance fallacies comprised a majority of those discovered.

**Table 3: Summary of Fallacies in the EUR Whole Airspace Safety Case**

Fallacy	Reviewer A	Reviewer B	Total <sup>a</sup>
Red Herring	4		4
Fallacious Use of Language	2	6	6
Fallacy of Composition	2	2	2
Omission of Key Evidence	2		2
Total	10	8	14

- a. If both reviewers identified the same instance of a fallacy, that instance is only reflected once in the total.

The reviewers agreed that both arguments committed the fallacy of composition when they decomposed the whole airspace into geographic regions without considering interactions between regions or external influences. Both reviewers also identified ambiguous language in the argument; however reviewer B flagged more instances of this fallacy because he was unsatisfied with the argument’s use of the

term “safe.” Again, reviewer A’s results included a greater variety of fallacies as was the case with the EUR RVSM safety argument. Detailed discussion of this safety case is provided in Section 4.3.

## 2.4 Summary

All three of the case studies we reviewed exhibited common types of faulty reasoning to varying degrees, confirming our initial hypothesis. The wide spectrum of fallacies displayed overall by the safety cases suggests that safety-case developers are unaware of common pitfalls in argument. Moreover, despite the discrepancies between their results, the reviewers ultimately achieved agreement regarding the fallacies in the safety cases, suggesting that safety-case review, although easily influenced by the backgrounds of the individual reviewers, need not be as error-prone or as subjective as it might appear to be. These two findings demonstrate the need for a taxonomy of logical fallacies specifically tailored to system safety arguments in order to make developers aware of common logical pitfalls and ensure that those who review safety cases do so from a common perspective.

## 3. A Taxonomy of Safety Argument Fallacies

To aid system developers in avoiding fallacious reasoning and to assist reviewers in detecting it, we present a *taxonomy of logical fallacies*, which is summarized in Table 4. We have adapted this taxonomy from Damer and Govier’s taxonomies of fallacies in classical arguments [2, 4]. Specifically, we have reused their names but adjusted the descriptions of the fallacies so that they more accurately describe the manner in which the fallacies are likely to appear in safety arguments. We have also reorganized and removed some fallacies from their taxonomies, especially those pertaining to political and emotional appeals since they are unlikely to appear in written safety cases. The selection of fallacies below consists of those we observed in our case study described in Section 2 as well as some we did not observe but nevertheless pertain to safety arguments. The following sections describe each of the categories in the taxonomy and their associated fallacies.

**Table 4: Taxonomy of Safety Argument Fallacies**

Relevance Fallacies	Acceptability Fallacies	Sufficiency Fallacies
Appeal to Improper Authority	Fallacious Use of Language	Hasty Inductive Generalization
Red Herring	Arguing in a Circle	Arguing from Ignorance
Drawing the Wrong Conclusion	Fallacy of Composition	Omission of Key Evidence
Using the Wrong Reasons	Fallacy of Division	Ignoring the Counter-Evidence
	False Dichotomy	Confusion of Necessary & Sufficient Conditions
	Faulty Analogy	Gambler’s Fallacy
	Distinction without a Difference	
	Pseudo-precision	

### 3.1 Relevance Fallacies

The fallacies presented in this section describe common types of irrelevant premises. Having no bearing on the truth or falsehood of a claim, such premises add no value to an argument, but they may distract the reader from other, substantive premises. They may also deceive the creator of an argument into believing that he has provided ample support for a claim when in fact very few of the premises offered have any merit. It is important to note that the premises supporting these lines of reasoning are not necessarily irrelevant; in some cases the premises may be relevant if accompanied by additional evidence. Developers of system safety arguments should avoid these fallacious lines of reasoning or offer the evi-

dence required to make the reasoning relevant. Likewise, reviewers of safety arguments should strike premises of these forms when they appear unless they are accompanied by the required additional evidence.

***Appeal to Improper Authority.*** This fallacy occurs when an argument supports a claim by referencing an authority that is in some way inappropriate or inadequate. Authorities that might be cited in a safety case include individuals, committees or other groups, standard documents, “best practices,” and system pedigree. Appealing to such authorities may be proper or improper depending on the specifics of the claim made, the system in which the claim is made, and the authority itself. As a general rule, any appeal to an authority within a safety case should be accompanied by an argument as to why this authority is relevant to the system being considered. Appeals to authority without such arguments should be considered suspect. As an example, consider a safety case in which a claim about the adequacy of the software within the system is supported by an appeal to having followed a particular software development standard. Arguments and supporting evidence should be included to justify why this particular standard is *relevant* to the software produced, and why following the standard sufficiently guarantees software adequacy.

***Red Herring.*** This fallacy occurs when an argument includes sub-arguments that are entirely unrelated to the claim that the argument is intended to support, alongside sub-arguments that are related to the claim. Suppose, for example, that the claim is made: “The software cannot cause the valve to stick.” Red herrings for this claim might include sub-arguments about the existence of software requirements, functions of the software unrelated to the valve, and the educational pedigree of the software developers. None of these have any real bearing on the truth of the asserted claim and thus should not be included in the safety case. Including red herrings wastes time and money, both for developers and reviewers.

***Drawing the Wrong Conclusion.*** This fallacy occurs when an argument asserts a claim different from what is actually supported by the evidence given within the argument. Suppose for example, evidence exists concerning the *reliability* of various parts of a particular system and about the interactions within this system. If this evidence alone is used within a safety case to support a claim about the *safety* of the system, the wrong conclusion would have been drawn. At most, the evidence adequately supports a claim about the reliability of the system; it does not support a claim about the system’s safety.

***Using the Wrong Reasons.*** This fallacy may be considered to be the reverse of drawing the wrong conclusion, with the difference between them resting in chronology. Using the wrong reasons occurs when the claim is developed first, followed by the development of (wrong) reasons supporting the claim. So, for example, if rather than having evidence about reliability already existing, as in the previous example, reliability evidence was sought as the sole support for a safety claim, then the fallacy of using the wrong reasons would have been committed.

### 3.2 Acceptability Fallacies

A premise is acceptable if there is reason to believe it may be true. An *unacceptable* premise, therefore, is one that is known to be false or is otherwise unbelievable. Unlike irrelevant premises, which are inappropriate for the particular claims with which they are associated, an unacceptable premise is inappropriate as support for *any* claim because it invokes inherently faulty reasoning. The fallacies in this section describe various ways in which unacceptable premises may appear in an argument. Methods to remove the fallacies are also explained, when such methods exist.

***Fallacious Use of Language.*** Language can be used fallaciously within a safety case argument in at least three ways. An argument may use a key term in multiple senses, without distinguishing them, so that the reader might be led to an unwarranted conclusion unless the ambiguity is detected. Alternatively, an argument may use a term or reference that has multiple senses consistently, but without indicating which particular sense is intended, possibly leading the reader to an unwarranted conclusion. Finally, an argument may define terms circularly, or contain premises that are nothing more than restatements of claims they are intended to support. Within safety cases, terms that seem particularly likely to be used fallaciously are

those that describe desirable system properties, such as “safety,” “reliability,” and “dependability.” Whenever these terms are used, they should be accompanied by careful, precise, and non-circular definitions. Any safety case that uses such terms without the requisite definitions is unacceptable.

***Arguing in a Circle.*** This fallacy occurs when a premise of an argument asserts the claim that the argument is attempting to support. A circular argument reduces to nothing more than “claim *P* is true because claim *P* is true.” That is, the evidence cited for the truth of the claim is simply the assertion that the claim is true. Such arguments are unacceptable because they prove nothing. Circular arguments are unlikely to appear directly within a safety case; but developers and reviewers should scrutinize the arguments looking for implicit circularity.

***Fallacy of Composition.*** Fallacious composition occurs when an argument claims that, because a property holds over the parts of a system or process, it therefore holds for the larger entity, without considering possible interactions between parts or external influences. A prototypical example of composition within a safety case is an argument that claims that a whole system is safe because its subsystems *A*, *B*, and *C* are safe. Alone, such an argument is unacceptable because it fails to consider the effect on safety of interactions among the subsystems. Arguments of this form may be made acceptable by adding an argument for the safety of the interactions, either by demonstrating that no possible interactions exist or by showing how the interactions do not reduce safety.

***Fallacy of Division.*** The opposite of the fallacy of composition, this fallacy exists when an argument claims that, because a property holds for a system or process, it also holds for the individual parts of the system or process, respectively. A prototypical example of fallacious division within a safety case is an argument that uses the reliability of a system as evidence for the reliability of a component residing within that system. Arguments of this form may be made acceptable by adding a cogent argument that demonstrates why the subcomponents’ reliability may not be less than that of its system.

***False Dichotomy.*** A false dichotomy is created when an argument presents a limited set of alternatives for a decision and assumes that one of the proposed alternatives must be correct, without having demonstrated that the presented set of alternatives encompass all the possible alternatives. A false dichotomy within a safety case would likely take the form of an argument for the adequacy of, for example, a particular design choice, based on arguments why a limited set of other design choices are inadequate, omitting additional argument demonstrating that the design choices considered span the range of possible alternatives. If such an argument is included, and this argument is cogent, then the dichotomy is not false.

***Faulty Analogy.*** An argument with a faulty analogy identifies how two or more systems or processes are alike in *certain* respects, and extends this likeness to other respects without considering the significance of the similarities identified or important differences between them. An example of a faulty analogy within a safety case is an argument for the use of software developed for a previous system based solely on certain similarities between the previous system and the current one, without an accompanying argument demonstrating that no differences relevant to the software exist between the systems. The Ariane 5 accident in 1996 may be said to have resulted, at least in part, from a faulty analogy within the rocket’s safety case [9]. To avoid faulty analogy, any argument that is based on an analogy should be supplemented by an argument justifying the use of the analogy and demonstrating the absence of relevant differences that invalidate it.

***Distinction without a Difference.*** This fallacy is essentially the opposite of a faulty analogy. It occurs when an argument claims that a system, component, or process is different from another but provides no substantive evidence of a difference. Two likely scenarios in which this fallacy might occur within a safety case are when a new system has been developed to correct problems with a previous system; or when some particular technique is used that might be thought to be similar to another, discredited technique. A distinction-without-a-difference fallacy will exist in these scenarios unless arguments are supplied that demonstrate the important, relevant differences between the systems or techniques. As a general rule, any

argument based on differences between alternatives should be accompanied by an argument providing sufficient evidence for the existence of relevant differences.

***Pseudo-precision.*** This fallacy may occur in two ways: (1) an argument asserts a quantitative claim about a system, component, or process using greater precision than that which may be ascertained from the premises to the claim; or, (2) an argument asserts a quantitative claim that is supported only in qualitative terms by its premises. To avoid this fallacy, any safety case that contains quantitative claims should include cogent arguments that justify the precision used within those claims and should ensure that all quantitative claims are supported, at least in part, by quantitative evidence.

### **3.3 Sufficiency Fallacies**

To establish the truth of a claim, an argument must provide sufficient supporting evidence. For safety arguments, sufficiency is determined by the specifics of the claim and the details of the system being considered. Depending on the claim being asserted, an argument may have to provide multiple and possibly independent reasons why the reader should accept the truth of the claim. Moreover, some forms of evidence are more convincing than others, and developers and reviewers must consider the strength of each piece of evidence in evaluating the sufficiency of the evidence as a whole. The fallacies in this section describe ways in which arguments can fail to provide sufficient evidence to support their claims, either by providing no or little evidence, by omitting crucial types of evidence, or by providing biased or weak evidence. Sufficiency fallacies are more subtle than relevance and acceptability fallacies, and so those developing and evaluating safety arguments should ask themselves what the burden of proof is for each claim they consider and then judge whether the evidence provided convinces them that the burden has been met.

***Hasty Inductive Generalization.*** This form of fallacious reasoning occurs when an argument makes a claim about a system or process based on data obtained from a sample that is too small or unrepresentative of the population from which it was drawn. The ways in which hasty generalization might occur within a safety case include claiming software correctness based on limited testing, claiming an adequate human-machine interface based on trials with only a handful of subjects, or claiming adequate hazard mitigations based on considering only a few usage scenarios. This fallacy may be avoided by supplementing each generalization within the safety case by a cogent argument justifying the generalization.

***Arguing from Ignorance.*** An argument from ignorance asserts that a claim regarding a system or process is true solely because there is no evidence to suggest otherwise. In its most bald form, this fallacy would appear in safety case as a claim that a system is safe based on nothing more than the evidence that no one has proven that it is not safe. Evidence of lack of proof of un-safeness may play a legitimate role within a cogent safety argument, but it should not stand alone. Instead, it should always be accompanied by positive arguments for safety.

***Omission of Key Evidence.*** Key evidence is omitted when an argument makes a claim about a system or process that one would expect to be supported with a particular kind of evidence; however, no evidence of that kind is provided and no valid reasons are given for its omission. For many typical claims that appear within a safety case, there exist corresponding types of evidence that should be expected to justify the claim. For example, claims related to human-machine interfaces should be expected to include as evidence data from usability studies. As another example, claims about the sufficiency of hazard mitigation should be expected to include evidence related to the hazard identification techniques employed. A good safety case will either include the expected evidence for such claims, or it will justify the omission.

***Ignoring the Counter-Evidence.*** Counter-evidence is ignored when, as the name suggests, an argument makes a claim for which there exists contrary evidence; however the argument fails to acknowledge or consider that evidence, perhaps giving an impression that the evidence does not exist. As an example, consider a human-machine interface for which four different usability studies were conducted. Suppose that three of the studies suggested that the interface was suitable for its intended purpose, but the fourth study



indicated it was not. A good safety case would not ignore this fourth study, but rather include an argument about why the results of the three studies should prevail over those of the other study.

***Confusion of Necessary & Sufficient Conditions.*** This fallacy occurs when an argument asserts without evidence that a condition that is necessary for a claim to be true is also sufficient for it to be true. An example of this fallacy within a safety case is as follows: the claim is made that “hazards have been mitigated,” but the evidence that is presented demonstrates only that hazard *identification* has been completed adequately. Although hazard identification is certainly necessary for hazard mitigation, it is hardly sufficient. Avoiding this fallacy requires that careful attention be paid to what constitutes the sufficient conditions for the specific claims that are made within the safety case.

***Gambler’s Fallacy.*** This fallacy is committed when an argument claims that the probability of a chance event has been altered by its historical performance. It draws its name from gamblers who believe they are more likely to win a game of chance while they are on a winning streak or, if they are on a losing streak, that their luck is bound to improve. Such reasoning is fallacious because past performance alone cannot influence the likelihood of future occurrence. Note that the fallacy does not occur when an argument uses statistical data to estimate what was previously an unknown probability or to refute a prior estimate, because in these cases historical performance is used as an indicator, not a determiner, of probability. An example of this fallacy would be a claim that, because a system has operated without fail for an extended period of time, it is less likely to fail in the near future. Without additional argument that other factors such as maintenance, re-design, or degradation have altered the probability, such an argument is fallacious.

## **4. Examples of Fallacious Inferences in Safety Cases**

To demonstrate the taxonomy, we revisit the three safety cases from our case study and examine some of the fallacious inferences we observed. The following sections elaborate upon the tables we presented for each safety case in Section 2 and discuss specific fallacies in greater detail. Again, our goal is to examine the inferences made by the arguments and not to assess the safety of the associated systems.

### **4.1 EUR Reduced Vertical Separation Minimums (RVSM) Safety Case**

A recurring problem we observed in the RVSM safety case was the use of evidence that did not support the immediate claim the argument was attempting to make. As an example, consider the top-level RVSM safety argument presented in Figure 1. The structure of this argument, as expressed, makes G2 irrelevant because if the safety requirements are realized in the implementation of the Concept (G3), then it does not matter whether they are also realized in the Concept itself. Since G2 does not contribute to the truth of the top-level claim that the collision risk is tolerable (G0), it does not belong in the argument. Moreover, in support of claim G1, “Safety requirements for Vertical Separation (under RVSM) are complete and correct,” the argument cited as evidence the fact that the requirements have been specified and allocated. The existence of requirements and their allocation to elements of the system does not aid the reader in determining whether they are complete and correct, and so this evidence is irrelevant to the claim. In both cases, the argument commits the fallacy of using the wrong reasons because it makes appropriate claims but supports them with irrelevant evidence.

Figure 2, which shows a portion of the argument concerning the role of flight training in the RVSM safety requirements, presents a more severe example of this problem. Of the four subgoals supporting the claim that “there is sufficient direct evidence of [flight crew] training design validity,” (St2.3.1) only one, “Hazards and risks controlled and mitigated” (G2.3.1.4), has any bearing on the truth of the claim. The other three subgoals are irrelevant for the reasons cited earlier. The argument is a red herring because it mixes weak supporting evidence with unrelated premises, giving the reader the impression that the basis for the claim is stronger than it actually is.

Some of the inferences made by the RVSM safety argument did not follow from the evidence, usually because the argument jumped directly from a high-level goal into foundational evidence without pro-

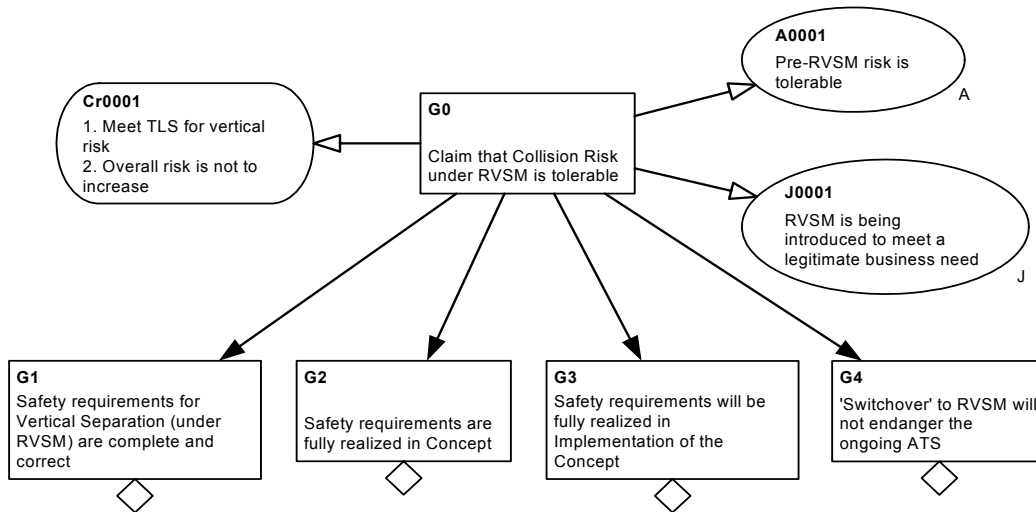


Figure 1: EUR RVSM Top-Level Safety Argument [5]

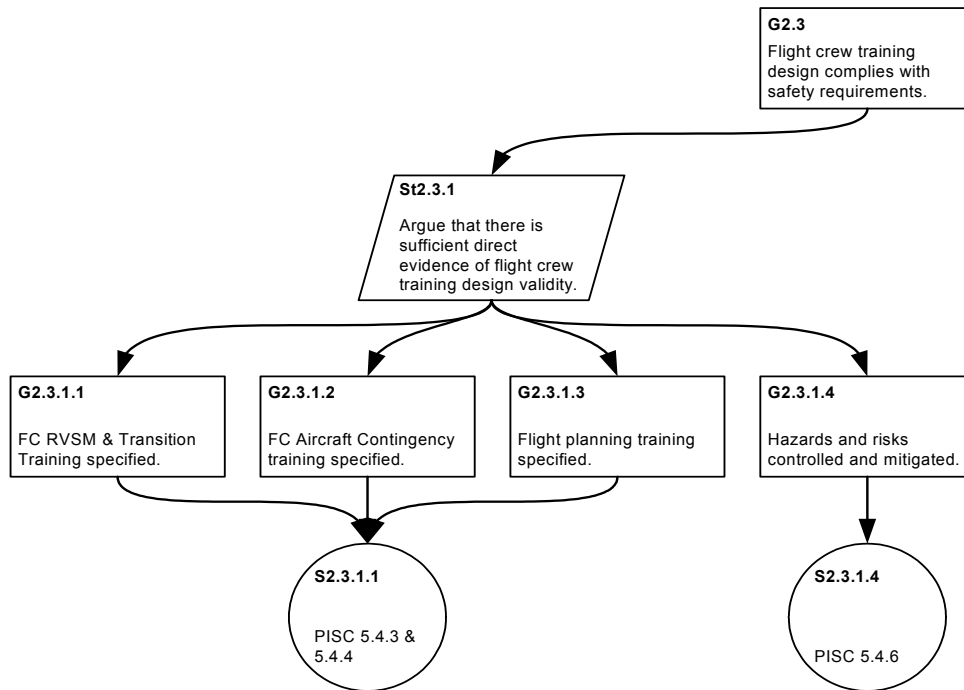


Figure 2: RVSM Flight Crew Training Validity Argument [5]

viding any justification as to why the evidence meets the goal. In its argument for the correctness and completeness of the RVSM safety requirements, one of the supporting goals states, “Analysis shows overall safety requirements to be complete and correct.” To support this claim, the argument points to a subsection in the natural language portion of the safety case describing the use of standard IEC 61508 in developing the safety requirements as well as an appendix listing each requirement and its corresponding safety objective. The argument makes a hasty inductive generalization because, although the evidence supplied is relevant and acceptable in a safety argument, it is not readily apparent from the evidence that the requirements are complete and correct. To make this connection, additional evidence supporting the com-

pleteness of the evidence needs to be given, justification should be provided as to why the evidence presented shows the requirements to be complete and correct, or the claim should be split into narrower claims to make the argument easier for the reader to follow.

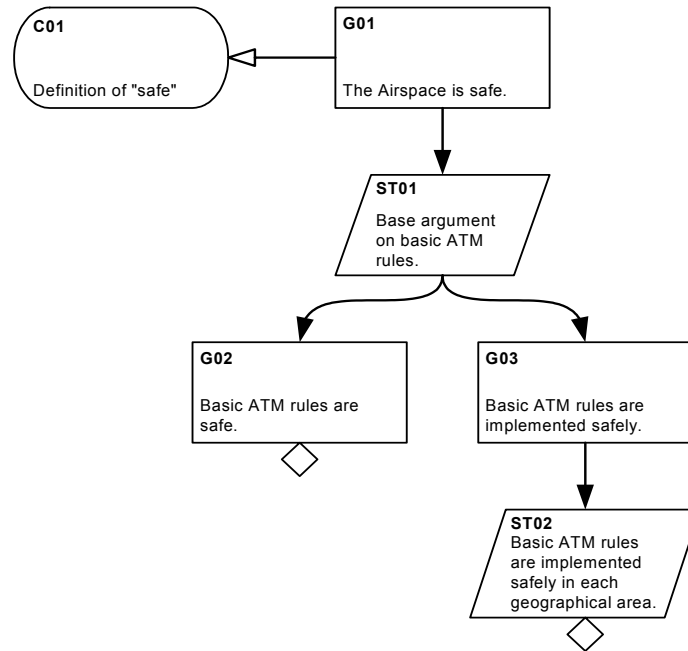
Finally, we observed a few instances in which the RVSM argument makes claims that are either so strong that they would be nearly impossible to support or uses imprecise language that makes them difficult for the reader to evaluate. In support of the claim in Figure 1 that the switchover to RVSM will not endanger current operations (G4), the argument states that “all possible failure modes were considered in the hazard and risk analyses.” Without proof that no possible failure mode could have been omitted from consideration, such a strong claim cannot be supported and should be tempered. Also in support of the switchover plan, the argument claims, “Failure modes [are] representative of [the] operational environment and workload.” The use of the term “representative” here is vague because the reader is not told what the operational environment and workload are and thus cannot evaluate the claim. Moreover, the restriction of the failure modes to those that are representative of the operational environment seems to contradict the earlier claim that *all* possible failure modes were considered by the hazard assessment.

## 4.2 Opalinus Clay Geological Repository Safety Case

Recall that the only genuine fallacies we identified in the Opalinus Clay argument were two instances in which key evidence was omitted. One of the arguments in the safety case discusses uncertainty in the characteristics of the chosen disposal system. The argument states that the choice of uncertainty scenarios to consider “remains a matter of expert judgement” and then proceeds to describe the process in which scenarios were developed and considered using a panel of experts. An obvious criticism of this approach would be that scenarios suggested by some experts were not selected for consideration but should have been. To avoid this criticism, the argument should mention some of the scenarios that were suggested but excluded from consideration by the panel along with its rationale for doing so. Elsewhere, the argument claims that “uncertainties [in the risk assessment cases] are treated using a pessimistic or conservative approach,” but no evidence is provided to support the claim of conservatism. Finally, in considering possible human intrusions, the argument assumes that “...possible future human actions that may affect the repository are constrained to those that are possible with present-day technology or moderate developments thereof” [10]. Although it is difficult to imagine how one would avoid making this assumption, it is possible that unforeseen future innovations will render the analysis moot.

The final argument for the safety of the repository asserts, “No issues have been identified that have the potential to compromise safety.” This claim is supported by “an analysis of a wide range of assessment cases that were derived in a careful and methodical way” that did not identify any outstanding issues with the potential to compromise safety. We initially flagged this reasoning as an instance of arguing from ignorance because the claim of safety is based upon a lack of apparent evidence to the contrary. An important difference in the Opalinus Clay argument, however, is that the claim of safety is not based upon *ignorance* of any counter-evidence but rather a thorough search for such evidence that has turned up none. The argument avoids the fallacy because there is reason to believe that such evidence does not exist as opposed to the developers’ simply being unaware of it.

It is likely that the developers of the Opalinus Clay safety case included the “absence of outstanding issues” argument to anticipate criticism that they had not considered evidence unfavorable to their safety claims. Accompanied with other lines of reasoning that point to positive evidence for the system’s safety, this argument strengthens the overall quality of the safety case. If the absence of outstanding issues had been the sole argument in the safety case, however, it *would* have been an instance of arguing from ignorance because it would have attempted to shift the burden of proof onto skeptics who doubted the safety of the system. As its creator and primary advocate, the developer of a system must carry the burden of proof in establishing its safety.



**Figure 3: Top-Level Structure of Rule-Based Whole Airspace Argument [8]**

### 4.3 EUR Whole Airspace Preliminary Safety Case

Although the two arguments in this safety case were intended to give separate lines of reasoning for the safety of the airspace, the reviewers found them to be very much the same once broken down into subgoals and evidence. The argument based on geographic areas split the top-level goal into subgoals claiming the safety of each geographic region in the whole airspace. Similarly, the rule-based argument argued that the ATM rules were implemented safely by claiming that they were implemented safely in each region. Neither argument considered possible interactions between geographic areas, such as when an aircraft is handed off by one air traffic controller to another in an adjacent region. Even if the safety rules are respected within each region, without special considerations for interactions between regions and with external airspace, the rules might be violated in the context of the broader whole airspace system. Both reviewers flagged these arguments as instances of fallacious composition.

Figure 3 illustrates the top-level structure of the rule-based argument. Because the argument is preliminary, it contains placeholders for elements that would need to be filled in if it were developed further. The contextual placeholder C01 suggests that a single definition of “safe” is being used throughout the argument when this cannot actually be the case. The top-level goal, “the airspace is safe,” is split into two subgoals that effectively define safe in the context of the airspace as the safe implementation of safe rules. The argument uses the term “safe” ambiguously because it neglects to specify what “safe” means in the context of the whole airspace rules or their implementation, and to use the previous definition would lead to recursion. Moreover, it does not provide a definition of “airspace” even though understanding what is meant by this term is crucial to evaluating the argument. Additional context should be provided to give a precise description of the airspace to which the argument refers.

## 5. Conclusions and Future Work

Safety cases are a powerful tool for merging the various elements of a system safety argument and presenting them in a structured form. A well-structured safety case built upon a cogent argument is a testament to a system’s safety in its operational context. Unfortunately, if the safety argument is incomplete or

faulty, the value of the safety case is lost because basic conclusions about safety might be invalidated. We have shown that fallacies are present and in surprising numbers in the safety cases used to document the safety of important applications, and based upon taxonomies of fallacies in classical arguments, we have developed a taxonomy of fallacies specifically tailored to system safety cases.

We believe our taxonomy will assist developers in avoiding these fallacies in their safety arguments. Our case study presented in Section 2 was undertaken before and was the motivation for the development of our taxonomy; however the disparity of results from our case study suggests that further study is warranted into ensuring the consistency of results across those who review safety arguments. Our next step is to determine the effectiveness of our taxonomy in aligning the critiques of independent reviewers.

## Acknowledgements

This work was funded in part by NASA Langley Research Center grants NAG-1-2290 and NAG-1-02103.

## References

- [1] Bishop, P. & R. Bloomfield. "A Methodology for Safety Case Development." *Industrial Perspectives of Safety Critical Systems: Proc. Sixth Safety-Critical Systems Symposium*, Birmingham. Springer-Verlag: 1998.
- [2] Damer, T.E. *Attacking Faulty Reasoning: A Practical Guide to Fallacy-Free Arguments*, 5<sup>th</sup> ed. Australia: Wadsworth. 2005.
- [3] Dependability Research Group. "Safety Case Repository." Department of Computer Science, University of Virginia. 2004. <<http://dependability.cs.virginia.edu/research/safetycases/safetycasesexamples.php>>
- [4] Govier, T. *A Practical Study of Argument*, 6<sup>th</sup> ed. Australia: Wadsworth. 2005.
- [5] EUROCONTROL. "The EUR RVSM Pre-Implementation Safety Case," ver. 2.0. Document RVSM 691. 14 August 2001. <<http://www.ecacnav.com/rvsm/library.htm>>
- [6] Greenwell, W.S., E.A. Strunk & J.C. Knight "Failure Analysis and the Safety Case Lifecycle." *Proc. 7<sup>th</sup> Working Conference on Human Error, Safety, and Systems Development*. Toulouse, France. August 2004. C.W. Johnson & P. Palanque (eds.)
- [7] Kelly, T. & R. Weaver. "The Goal Structuring Notation - A Safety Argument Notation." *Proc. DSN Workshop on Assurance Cases: Best Practices, Possible Outcomes, and Future Opportunities*. Florence, Italy. July 2004.
- [8] Kinnersly, S. "Whole Airspace ATM Safety Case - Preliminary Study." November 2001. <<http://www.eurocontrol.int/care/innovative/studies2001/aeat/aeat.htm>>
- [9] Lions, J.L. "Ariane 501 Failure: Report by the Inquiry Board." European Space Agency. 19 July 1996.
- [10] Nagra. "Project Opalinus Clay: Safety Report." Technical report NTB 02-05. December 2002. <[http://www.nagra.ch/english/aktuell/e\\_nachweis/ntb02\\_05.htm](http://www.nagra.ch/english/aktuell/e_nachweis/ntb02_05.htm)>