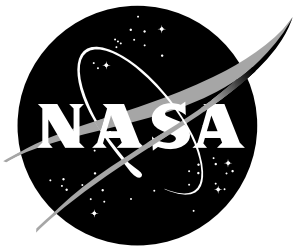
NASA/TM–20230003336



Towards a Coherent View of Evidence in Safety Assurance

Finlay McCardel
University of Glasgow, Glasgow, Scotland, UK

C. Michael Holloway
NASA Langley Research Center, Hampton, Virginia

Kimberly Wasson
Joby Aviation, Crozet, Virginia

Neil McDonnell
University of Glasgow, Glasgow, Scotland, UK

Mallory Graydon
NASA Langley Research Center, Hampton, Virginia

Abel Peña
Intern, NASA Langley Research Center, Hampton, Virginia

Sarah Lehman
NASA Langley Research Center, Hampton Virginia

April 2023

NASA STI Program. . . in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI Program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI Program provides access to the NASA Aeronautics and Space Database and its public interface, the NASA Technical Report Server, thus providing one of the largest collection of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

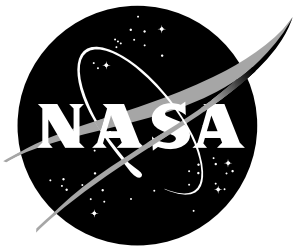
- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counterpart of peer-reviewed formal professional papers, but having less stringent limitations on manuscript length and extent of graphic presentations.
- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.
- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.
- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.
- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.
- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include organizing and publishing research results, distributing specialized research announcements and feeds, providing information desk and personal search support, and enabling data exchange services.

For more information about the NASA STI Program, see the following:

- Access the NASA STI program home page at `http://www.sti.nasa.gov`
- E-mail your question to `help@sti.nasa.gov`
- Phone the NASA STI Information Desk at 757-864-9658
- Write to:
  NASA STI Information Desk
  Mail Stop 148
  NASA Langley Research Center
  Hampton, VA 23681-2199

NASA/TM–20230003336



Towards a Coherent View of Evidence in Safety Assurance

Finlay McCardel
University of Glasgow, Glasgow, Scotland, UK

C. Michael Holloway
NASA Langley Research Center, Hampton, Virginia

Kimberly Wasson
Joby Aviation, Crozet, Virginia

Neil McDonnell
University of Glasgow, Glasgow, Scotland, UK

Mallory Graydon
NASA Langley Research Center, Hampton, Virginia

Abel Peña
Intern, NASA Langley Research Center, Hampton, Virginia

Sarah Lehman
NASA Langley Research Center, Hampton, Virginia

National Aeronautics and
Space Administration

Langley Research Center
Hampton, Virginia 23681-2199

April 2023

Acknowledgments

Finlay McCardel's work on this paper was supported by a scholarship from the Scottish Graduate School for Arts and Humanities Arts and Humanities Research Council Doctoral Training Partnership (SGSAH AHRC DTP).

C. Michael Holloway's work was sponsored in part by the Federal Aviation Administration through Interagency Agreement IA1-30333, Annex 2: Using the Overarching Properties in Novel Examples (Opine). Nothing written here, however, should be considered to represent the official views of the FAA.

Available from:

NASA STI Program / Mail Stop 148
NASA Langley Research Center
Hampton, VA 23681-2199
Fax: 757-864-6500

Abstract

This paper explains why labeling something as evidence does not make it special and proposes a systematic approach for avoiding giving the word more authority than it warrants in rational inquiry about the real world.

1 Introduction

> *'Evidence' is a term that I have used informally in introductory or summary formulations. I have not found it useful in more detailed inquiry.*
> — W. V. O. Quine [1]

Appeals to evidence are ubiquitous. Scientists speak of basing theories on the evidence rather than opinions or personal preferences. Every legal system has detailed rules distinguishing between what may be and may not be exalted as evidence. Evidence-based medicine is hailed as a step forward. Within the domain of safety engineering, an oft-cited standard [2] demands that arguments be supported by a body of evidence. Accident investigators claim to seek evidence before they speculate about causes. In real life and fiction, people perceived to possess an admirable commitment to unbiased inquiry say such things as, "I just follow the evidence wherever it leads." Even conspiracy theorists often appeal to evidence to buttress their theories. But, does the word 'evidence' warrant this lofty status? That is, does the word's modern usage adequately convey the crucial concept it is intended to convey? We assert it does not. Here's why.

Put simply, despite the centrality and importance of the concept, there exist no commonly-accepted parameters for the use of the word 'evidence' in general or even within particular domains.[1] In this paper, we seek to provide a pragmatic solution to the deficiency for the domain of safety engineering. Specifically, we outline some extant features of both the concept (evidence) and the word ('evidence'), and we argue that appeals to the concept can be misleading because of the ambiguity of the word. We offer an interpretation of 'evidence' that (we hope) reduces the likelihood of confusion, and we make use of it in three illustrative examples in the field of safety engineering. Whilst remaining guided by the concept is crucial, equally crucial is not taking any given use of the word as more authoritative than it really is. Simply calling something 'evidence' does not make it special.

2 The Anatomy of Evidence

> *What is the* point *of evidence? When put that way, it becomes clear that we do not value evidence for its own sake. Evidence is not like happiness, pleasure, or dignity, which are plausibly considered ends and*

[1] A plausible case can be made that the domain of law constitutes an exception [3], at least if we restrict attention to the laws within a particular legal system. For example, the United States Federal Courts adhere to a single standard for what may be admitted in court as 'evidence'. See `https://www.law.cornell.edu/rules/fre`.

> *not means. Rather, evidence is a means to some end, and the end is some factual conclusion of interest to us. And embedded in the factual conclusions that interest us is the assumption that those conclusions are valuable because and when they are* true. *So we can say, conventionally, that evidence is valuable insofar as it leads to truth—or, more precisely, to a belief in things that are true.* — Frederick Schauer [4]

2.1 The Concept

The concept of evidence is highly diverse. Evidence can come in the form of a document or an image, a sound or video recording, oral or written testimony, the presence of an object, the absence of an object, or the occurrence (or non-occurrence) of some process or event. A rock sitting on the floor beyond a broken pane of glass can be evidence but so can a photograph of the rock, or a video of its arrival, the document detailing its position and state, or the testimony of a witness. If we are to be as precise as possible, we might prefer to say that the presence of the rock *in that position* is evidence and that a photograph (or document, or video, etc.) *of the rock in that position* is evidence *of that evidence*. Evidence can come in many forms, and it can be nested.

We may then ask, 'what is the rock evidence *of*?' This question is crucial. If we come to the scenario with a question of how it is that the window came to be broken, then we might take the rock to be evidence in support of the proposition that a rock broke the window. If, on the other hand, we come to the scenario with the question of what scratched the floor, then we will take the rock to be evidence of a different proposition; namely that a rock scratched the floor. This example demonstrates that there is no such thing as free-standing evidence. That is, evidence is not a self-contained, intrinsic feature of the rock or its surroundings. The rock is only evidence *of* some other thing. More generally, ***X* is never simply evidence but always evidence of some *Y***. When we speak otherwise, we are being elliptical. That is, we are expressing an incomplete claim when we say 'X is evidence' and do not specify for which Y we take it to be so. In some discussions, what we take Y to be may be sufficiently obvious that leaving it unsaid is harmless, but in other discussions leaving it unsaid may be harmful, causing confusion and misunderstandings among participants in the discussion.

What sort of thing is Y? There are many different sorts of thing that Y could be — much like with the heterogeneous nature of the X candidates we considered above. Y could be a theory, or a course of action, or a conclusion about the performance of some system. When we choose to consider X to be evidence of Y, however, we consider X to establish that Y is at least more likely to be correct (or true) than it would be without X.[2] So, **whatever *Y* is, it must be the sort of thing that is *truth-evaluable***. Philosophers recognise this as the hallmark of a *proposition*: the thing expressed by a declarative sentence. Propositions are either

[2]Hence the Bayesian conception of evidence, according to which X is evidence for Y only in the case that the conditional probability of Y on X is greater than the unconditional probability of Y. To non-Bayesians out there, from a non-Bayesian in here, this footnote is not intended to imply an acceptance of Bayesianism.

true or false. So, when we consider X to be evidence for a theory we take that to mean that it is evidence that the theory is true; when we consider X to be evidence of some event we take it to mean that it is evidence that the event happened; and when we consider X to be evidence for a course of action, we take that to mean that it is evidence for the truth of the proposition that the course of action is sensible. **Treating Y as a proposition leads us to consider it as a conclusion to an argument for which X is a premise.**

If we consider X to be a premise in the argument for Y, then that would imply that X is also a proposition, since premises are always propositions. Does this mean that evidence is always a proposition? Reflecting on usage suggests not: a rock is *not* a proposition! We recognize that many of the things we pick out with the term 'evidence' are not propositions themselves, but we maintain that it is nevertheless the case that all of the claims we want to make when identifying evidence can be captured by propositions. 'There is a rock on the floor' is a proposition that captures the fact that the rock is on the floor.[3] The importance of treating evidence as a proposition is only apparent when we move from speaking elliptically of X being evidence to giving full expression to the claim that X is evidence of some Y. So, **whether it matches common language usage or not, we should treat X as a proposition**.[4]

Once we understand the concept of evidence as being a part of an argument where the evidence is a premise for a (sometimes unstated) conclusion, we can open up the question of whether the argument is a good one or not. This is not just a matter of whether the premise is true (i.e., whether the rock is really there). One could grant that X is true but not consider that truth to justify the conclusion Y. Using our toy example again, it might be agreed by all parties to some dispute that the rock is on the floor but not agreed that this is evidence that it broke the window, since one party believes that the rock was in that very position before the breaking and the other party does not. In such cases, **whether something counts as evidence depends on these background beliefs**.

In philosophical parlance, we may say that something counting as evidence is relative to an *epistemic situation*.[5] We can capture disparities among parties by making some of these implicit background beliefs *ex*plicit and incorporating them into arguments for Y: one party has the premise 'the rock was in that exact position before the breaking' and the other has the premise 'the rock was *not* in that exact position before the breaking'. In this case, the claim that the rock is evidence is doubly incomplete: it does not state the conclusion for which the presence of the rock is a premise, nor does it state the further suppressed premise(s) about the prior position of the rock. It may not be the case that every claim of evidence *does* have both of these suppressed elements, but it is important to note that any given claim *can*. In this paper we will advocate a way of understanding appeals to evidence that facilitates the exposure of these otherwise suppressed elements whenever they are

[3]Cf. [5]: "[T]he historical evidence is not the physical document itself but various propositions about it, for example that it is signed 'John.' The biochemical evidence is not the experiment as an event but, for example, the proposition that it was carried out with such-and-such results."

[4]For a fuller argument for treating evidence as propositional, see [6].

[5]Hence, this is what Achinstein calls *ES-evidence* — see [7].

present.

To summarise, then, the concept of evidence is such that **any given appeal to evidence can be interpreted as a (compressed) argument in which the thing appealed to is one premise, but not necessarily the only premise, in support of some target conclusion.**

2.2 The Word

As noted above, the use of the word 'evidence' in English may refer to a wide range of different phenomena. Our argument above claims that it is a useful, and regimenting, step to treat evidence as a proposition even though that is not a restriction we find in ordinary usage. There are some other features of ordinary usage that we think it is important to note and which we believe may confound the appeal to evidence in high-stakes contexts.

According to our breakdown of the concept, when we assert 'X is evidence', we are expressing, in a compressed form, an argument in which a proposition about X is a premise for some conclusion Y. Note, though, that the statement 'X is evidence' is not necessarily neutral in at least the following three respects:

1. It generally is not neutral on the truth of X: the labeling as 'evidence' carries the implication that X is true.

2. It may not be neutral with regard to the quality of the argument from X to Y: again the labeling as 'evidence' endorses the inference from X to (the greater likelihood of) Y.

3. It generally is not neutral on the truth of Y: Saying that X is 'evidence' of Y subtly implies (defeasibly) that Y is either true, or more likely to be true than not. If the speaker took it to be otherwise, they would have modified the claim, and cancelled the implicature, by describing it as 'misleading evidence' or with some equivalent caveat.

These pragmatic considerations concerning usage are ones that risk confounding the neutral, and empirically respectable, status that evidence enjoys in our reasoning. The concept is vital, but the term 'evidence' is problematically loaded. Indeed, it is precisely because the concept is vital that we need a coherent way of talking about it. To be clear, the issue is not with the things referred to as 'evidence' but with the fact that we use this word to refer to them. Through its various uses in different contexts, the word picks up subtle but significant connotations that we would do better to avoid in high-stakes situations. One might think that the best solution would be just to shake these connotations off, but the word is bound to pick them back up again. Instead, in the following sections, we suggest a neutral way of interpreting appeals to evidence that avoids actually using the term 'evidence'.

2.3 Reconstructing Appeals to Evidence

We have said that evidence is to be understood as a premise and that together with other premises, it lends support to a conclusion. In some arguments, the support

lent to the conclusion is so strong that the conclusion is *guaranteed* to be true if the premises are true. When this is the case, we say that the argument as a whole is *deductively valid*. Think of a deductive argument as any argument that aims to be deductively valid.

We often find deductive arguments in mathematical proofs. Deductive validity is a standard we can meet in the idealised world of mathematics. However, things are generally not so neat and tidy in the world of empirical observations. When we appeal to X as evidence of Y, X does not normally *guarantee* the truth of Y. Another way to put the same point is to say that evidence generally provides *non*-deductive support for its conclusion.

To make this clearer, we will try reconstructing a particular appeal to evidence as an argument from X to Y. Before doing that, however, we need to think about the process by which one identifies the background beliefs[6] that should feature in the argument. Here is a rough guide to identifying the background beliefs that one ought to make explicit when reconstructing an appeal to evidence.

Suppose you believe Y, and you appeal to X as evidence of Y. To reconstruct this as an argument from X to Y, we first imagine that you have a smart and generally well-informed opponent who believes X but who also believes the *negation* of Y. Then we ask, 'What else might this smart and generally well-informed opponent believe in order to justify their belief that not-Y?' Lastly, we ask, 'Which of *your* background beliefs contradict your *opponent*'s beliefs?' and we add the answers as premises in the argument from X to Y.

To illustrate, suppose you appeal to the presence of a rock on the floor as evidence that the window was broken by a rock. We can begin to represent this appeal to evidence as follows:

X There is a rock on the floor beyond the broken window.

...

Y The window was broken by a rock.

In its current state, the argument from X to Y is weak: in philosophical parlance, it is deductively invalid. To strengthen the argument, we can ask, 'What might your smart and generally well-informed opponent believe in order to justify the belief that the window was *not* broken by a rock?' Well, as noted above, they might believe that the rock was in the exact same position before the breaking of the window. And if *you* believe that the rock was *not* in the exact same position before the breaking of the window, then we should add this as a premise to the argument:

X There is a rock on the floor beyond the broken window.

B1 The rock was not in that exact position before the breaking of the window.

...

Y The window was broken by a rock.

[6] 'Background beliefs' here correspond in part to 'presuppositions' in the maxim *Presuppositions Predetermine Plausibility* as discussed in [8].

We can see that the argument is now better (i.e., stronger): the inference to Y from X *and B1* is more reasonable than the inference to Y from X alone.

However, the argument remains deductively invalid: the truth of the premises does not *guarantee* the truth of the conclusion. To strengthen the argument further, we can ask, 'What else might your smart and generally well-informed opponent believe in order to justify their belief that not-Y?' Well, they might believe that there is some other object in the vicinity of the window that would cause it to break if launched at it. And again, if you believe otherwise, then we should add that belief of yours to the argument:

X There is a rock on the floor beyond the broken window.

B1 The rock was not in that exact position before the breaking of the window.

B2 There are no other objects in the vicinity of the window that would cause it to break if launched at it.

...

Y The window was broken by a rock.

As before, we can see that the argument is now even better: the inference to Y from X, *B1*, *and B2* is more reasonable than the inference to Y from X and *B1* alone. This trend ought to continue for all your background beliefs that a smart and generally well-informed opponent might contradict in order to justify their denial of Y. By thinking of background beliefs this way you will address only those points that are truly contentious and avoid listing background beliefs on which all parties to the dispute already agree — such as the laws of thermodynamics, or the ability of a rock to break a window.

However, as much as we might strengthen the argument by adding in premises this way, we generally should not aim for deductive validity. This is in stark contrast with the usual practice in analytic philosophy, but that practice is ill-suited to the contingent and empirical domain of scientific or engineering inquiry. Rocks can be placed on floors without breaking windows along the way; objects that *do* break windows can be removed from the scene after doing so. Premises like *B1* and *B2* do not lend themselves to deductively valid arguments. However, on the assumption that you are smart and generally well-informed, these background beliefs are likely to be all we can add to strengthen the argument from X to Y. Creating *any* deductively valid argument from X to Y is likely to involve at least one premise that claims too much.

For example, if we were to try to make the argument deductively valid, we might add in a conditional premise (C) and represent the inference with the word *therefore* as follows:

Deductive Argument

X There is a rock on the floor beyond the broken window.

B1 The rock was not in that exact position before the breaking of the window.

B2 There are no other objects in the vicinity of the window that would cause it to break if launched at it.

C *If* there is a rock on the floor beyond a broken window, and the rock was not in that exact position before the breaking of the window, and there are no other objects in the vicinity of the window that would cause it to break if launched at it, *then* the window was broken by a rock.

Therefore,

Y The window was broken by a rock.

However, intuitively C claims too much. As such, it ought not to be among your background beliefs. Unfortunately, the same will be true of most (likely all) other premises that make the argument deductively valid.[7]

We contend, then, that it is generally a mistake to try to expand appeals to evidence by representing them as deductive arguments. At this point, however, one might be wondering what the alternative is and whether it is any better. We contend that the following alternative reconstruction *is* better:

Non-deductive Argument

X There is a rock on the floor beyond the broken window.

B1 The rock was not in that exact position before the breaking of the window.

B2 There are no other objects in the vicinity of the window that would cause it to break if launched at it.

Therefore, *likely*:

Y The window was broken by a rock.

The '*likely*' qualifier here is what makes it clear that this is a non-deductive argument: it does not aim at deductive validity. The argument is in good shape as it is, since the conclusion is indeed likely given the premises. For any good appeal to evidence, this qualifier will be appropriate. An especially strong non-deductive argument might deserve a stronger qualifier (e.g., '*very likely*'), but one ought not to expect to construct an argument where the inference needs no qualifier at all.[8]

Why think this way of expanding the appeal to evidence is better? The reason is that it guards against overstating the support that X gives to Y. The goal of appealing to X as evidence may be to persuade others of the truth of Y, but the

[7]To include a conditional premise like this is to presuppose the absence of a counterexample – i.e., to presuppose that either the argument's conclusion is true or some of its premises are false. But if such a presupposition is necessary, then clearly the appeal to evidence is not persuasive.

[8]The approach described here to handle uncertainty differs in mechanics from the approach used by Wasson and Holloway in [9] and associated documents, but it is conceptually consistent. In the Wasson and Holloway approach, all arguments are implicitly qualified; the specifics of the qualification are handled by a definition: 'An argument is cogent if it rationally justifies believing its conclusion to the required standard of confidence.' Qualifiers used in this paper (such as *likely* and *very likely*) correspond to standards of confidence in [9].

goal of *reconstructing* that appeal is not to persuade others of the truth of Y; rather, it is to unpack what we mean when we say that X is evidence for Y. Non-deductive arguments do not guarantee their conclusion, but *that is the point.* **We should not represent our appeals to evidence as more persuasive than they really are, and representing them as non-deductive arguments makes their weaknesses easier to see, regardless of one's subject knowledge or expertise.**

2.4 Evaluating Evidence

Now that we have argued that we ought to represent appeals to evidence as lending *non*-deductive support to some target proposition, one might naturally want to know how to *evaluate* non-deductive arguments. As we will see, this is no easy task.

People with mathematical proclivities will want to quantify how well a conclusion is supported by a set of premises. So, calculating the conditional probability of the conclusion *given* the premises will seem appealing. In a non-deductive argument, it will be less than 1 but (hopefully) greater than 0. Obtaining the calculated probability seems to eliminate vagueness, and vagueness seems less than ideal when reasoning about important subjects. However, **the fact that vagueness is not ideal does not mean that we should pretend to be able to eliminate it.** In some (perhaps nearly all) real world contexts, vague may be as good as we can do. Appeals to evidence are (often) one such context.

When representing an appeal to evidence as an argument for some proposition Y, the conditional probability of Y depends on what the premises are; however, we can only provide a *rough* guide to incorporating background beliefs as premises: there is no *precise* way to identify which background beliefs are relevant. Moreover, even once we have decided on a set of premises, we usually are not able to calculate the precise probability of a proposition conditional on them. Evidence tends not to come as part of a pre-defined package like a deck of playing cards, where the set of possibilities is known and their precise probabilities are calculable. Consider Hansson:

> For good or bad, life is usually more like an expedition into an unknown jungle than a visit to the casino. Most of the time we have to deal with dangers without knowing their probabilities, and often we do not even know what dangers we have ahead of us. [10]

Hansson cautions us against what he calls the *tuxedo fallacy*, i.e.,

> to proceed as if reasonably reliable probability estimates were available for all possible outcomes ... treating all decisions as if they took place under epistemic conditions analogous to gambling at the roulette table. The tuxedo fallacy is dangerous since it may lead to an illusion of control and to neglect of uncertainties that should have a significant impact on decisions. [10]

Eliminating vagueness is often the right thing to do, but representing rational support with numbers can also do harm by giving a sense of clarity where there is little or none.[9]

A different (but compatible) way to evaluate non-deductive arguments is in terms of how well the conclusion *explains* the premises (or some subset thereof). For the *Non-deductive Argument* example above, the answer is something like 'fairly well': if someone were to ask why there is a rock on the floor beyond the broken window, we would give a fairly good explanation by saying that the window was broken by a rock. This question is one that can only be answered in vague terms, but, as above, perhaps vague is as good as we can do.

Either way, these approaches concern the *structure* of the argument on offer rather than the plausibility or truth of the premises within that structure. To fully understand the overall rational support that the argument gives for the conclusion, however, we must take both the structure *and* the content of the argument into account, and when it comes to non-deductive arguments, there is usually no way to justify putting a precise figure on that. The takeaway message of this section is thus a simple, cautionary one: **if you see a precise figure on strength of evidence, you should *doubt* it.**

3 Illustrations in Safety Engineering

> *How are we going to be sure we have achieved safety if we think of safety as an absolute? ... In practice, our goal with safety has to be tempered by reality. We can make systems safe by never using them, but that is not really what we want.* — John C. Knight [12]

The toy case of the rock and the window is intentionally mundane so as to avoid unnecessary controversy. We turn now to three illustrations of appeals to evidence that are more relevant to the practice of safety engineering. In the following section, we conclude by presenting a practical procedure for applying our observations and reemphasizing some of our key points.

3.1 Illustration 1: Pilot Fit for Flight

Suppose we are trying to defend the claim that a particular pilot (call her Jess) is eligible and fit for flight. In our attempt to defend this claim, we may appeal to evidence *of Jess's eligibility and fitness*. For example, we may point to the fact that Jess has obtained a valid pilot license in accordance with the requirements for her license class in her country (e.g., 14 CFR Part 61, Subpart E in the United States for a private pilot license), holds a current medical certificate authorized by a qualified Aviation Medical Examiner, and is appropriately rated for the aircraft category, class, and type in question (e.g., single-engine piston).

[9] For a relevant example of a use of numbers that is potentially illusory in this way, see de la Vara, García, Valero, & Ayora [11], wherein it is suggested that the strength of an evidence artifact should be represented as a value between 0 and 100.

Bearing in mind the propositional nature of evidence, we can begin to construct an argument from this evidence to the claim that Jess is eligible and fit for flight:

> *X* Jess meets eligibility criteria to pilot the aircraft in question, in which she must hold an applicable license, medical certificate, and type rating.
>
> ...
>
> *Y* Jess is eligible and fit for flight.

Next, we need to decide which background beliefs to make explicit. If we marshal in a conditional such as '*If* Jess meets eligibility criteria to pilot the aircraft in question, *then* she is eligible and fit for flight', then we will have a deductively valid argument; but of course, this premise claims (far) too much. Instead, we should represent the appeal to evidence as a non-deductive argument, and strengthen it by including as premises the beliefs that a smart and generally well-informed opponent might contradict in order to justify their denial of *Y*. This process might yield something like the following:

> *X* Jess meets eligibility criteria to pilot the aircraft in question, in which she must hold an applicable license, medical certificate, and type rating.
>
> *B1* The licensing qualifications cover a codified base of knowledge and practical skills specific to the safe and efficient operation of aircraft.
>
> *B2* The medical qualifications cover a codified base of medical conditions and history specific to potential aircraft operational hazards.
>
> *B3* The type rating qualifications cover extensions to the licensing qualifications to account for additional operational and safety features of particular aircraft types.
>
> Therefore, *very likely*:
>
> *Y* Jess is eligible and fit for flight.

The non-deductive structure of the argument is plain to see: there is some recognized degree of risk of not being eligible and fit to fly despite meeting specific eligibility criteria.[10] Presenting this evidence as a premise in a non-deductive argument makes this risk recognizable without specialist knowledge and thus (we hope) harder to ignore.[11] Reconstructing this appeal to evidence as a non-deductive argument also encourages consideration of the background beliefs upon which the

[10] For example, Jess is taking antihistamines for a cold, or has not had adequate rest, or is in another condition explicitly defined as 'unfit' according to regulations.

[11] As Schauer [4] insightfully explains, "[N]onexperts often have the ability to identify and evaluate the rationality of what experts conclude, even if the nonexperts do not understand the underlying methods and conclusions. When so-called experts offer conclusions and the reasons for those conclusions that are internally contradictory or rely on implausible initial premises, nonexpert assessment can reject the expert conclusions even if the assessors are not themselves aware of the expert methods that are allegedly being used. You do not have to be an astronomer to know that the moon is not made out of green cheese, and if someone purporting to be an astronomer says that it is, then non-astronomers have good reason to reject what is advertised as an expert conclusion."

evidence's support (and even its *status as evidence*) is conditional. Often, overlooking the fact that evidential support is conditional on background beliefs will not do any harm; however, when dealing with high-stakes cases such as this, it is good not to subtly presume that the evidence in question provides support all by itself. Even if we often communicate perfectly well by appealing to evidence simply as 'evidence', best practice involves unpacking this elliptical way of speaking so that the weaknesses in our appeals to evidence are fully exposed for all to see, regardless of one's epistemic situation.

3.2 Illustration 2: Adequate Aircraft Hazard Identification

Suppose we are trying to defend the claim that the hazardous states for a particular aircraft[12] have been identified and adequately characterized. In our attempt to defend this claim, we may cite the hazard analysis conducted on the aircraft as evidence *of the identification and adequate characterization of those hazardous states*

As before, we can begin to construct an argument, making use of our propositional understanding of evidence:

X Aircraft Functional Hazard Assessment (AFHA) has been conducted.

...

Y Aircraft-level hazards have been identified and adequately characterized.

Again, we need to determine which background beliefs to make explicit; and again, any attempt to make this a deductively valid argument will most likely involve an implausibly strong premise. Certainly, it would be implausible to suggest that if *any* AFHA has been conducted, then all aircraft-level hazardous states have been identified and adequately characterized. Perhaps the AFHA was conducted by an inexperienced or incomplete staff. Perhaps crucial details of either the aircraft or its intended functions were changed after the AFHA was conducted, invalidating its results. Perhaps assumptions about how crew will react when the aircraft enters a given hazardous state are inaccurate, undermining characterization of the severity of that state. Perhaps the aircraft is of a novel type, or intended for novel operations such that additional inputs to the AFHA as well as new failure modes must be considered. Even if all these possibilities are ruled out, others will remain. Once again, a non-deductive representation of the appeal to evidence is required. For example:

X Aircraft Functional Hazard Assessment (AFHA) has been conducted.

B1 The AFHA was conducted according to standard practice by appropriate personnel.

[12] Readers familiar with SAE ARP4754A [13] and SAE-ARP4761 [14] might be more familiar with identification of an aircraft's 'failure conditions.' Here, we use the general systems-safety concept of a hazardous state instead.

B2 The aircraft and air operation specifications assumed during the AFHA are accurate.

B3 Assumptions about crew mitigations of hazardous aircraft states are accurate.

B4 The AFHA process has been endorsed by relevant aviation regulators.

Therefore, *very likely*:

Y Aircraft-level hazards have been identified and adequately characterized.

There are of course real world examples of aircraft-level hazard assessments failing to identify or adequately characterize hazards. We do not mention any specific examples, to avoid leading any readers down a path away from the central messages of this paper.

3.3 Illustration 3: Satisfactory Software Development

Suppose we are trying to defend the claim that the software for a particular aircraft has been developed to ensure it will meet the requirements (including the safety requirements) allocated to it during aircraft design.[13] In our attempt to defend this claim, we may cite as evidence *meeting the objectives of a relevant standard*.

As before, we can begin to construct an argument, making use of our propositional understanding of evidence:

X The aviation software item was deemed to meet the objectives of RTCA DO-178C at software level B.

...

Y The software satisfies the system requirements allocated to it in this application.

Again, we need to determine which background beliefs to make explicit; and again, any attempt to make this a deductively valid argument will most likely involve an implausibly strong premise. Certainly, it would be implausible to suggest that *any* software found to conform to RTCA DO-178C [15] meets all safety requirements allocated to it. Perhaps the rigor of a given conformance audit is suspect. Perhaps something will be missed during even a rigorous conformance audit; no assessor has the time to revisit every part of every development artifact. Perhaps the standards' objectives are themselves insufficient to show the necessary quality, either because the standard is flawed or because it is ill-suited to techniques such as neural networks used in the software construction. Even if all these possibilities are ruled out, others

[13] Uniting safety-specific requirements with all other system requirements is a foundational principle for handling software in aviation. This unification places the burden for ensuring safety requirements are *captured* where it belongs: on safety engineers. Software engineers are responsible for ensuring the software *satisfies* all requirements. Failure to understand this fundamental unification principle has resulted in some ill-founded criticisms from (primarily) academics about whether aviation software standards / guidelines handle safety.

will remain. Once again, a non-deductive representation of the appeal to evidence is preferable. For example:

> *X* The aviation software item was deemed to meet the objectives of RTCA DO-178C at software level B.
>
> *B1* Assessment of DO-178C conformance was conducted by qualified personnel.
>
> *B2* Conformance assessment by qualified personnel is sufficient to show satisfaction of DO-178C objectives.
>
> *B3* Meeting the objectives of DO-178C ensures adequate quality of this software.
>
> Therefore, *very likely*:
>
> *Y* The software satisfies the system requirements allocated to it in this application.

While aviation software developed to DO-178C has an excellent safety record, defects are sometimes found in aviation software built to that standard.

4 Concluding Remarks

> *I see no reason why I should be consciously wrong today because I was unconsciously wrong yesterday.* — Justice Robert Jackson [16]

So far in this paper we have identified and highlighted the shortcoming of simple appeals to evidence, proposed an alternative approach to articulating evidential claims, and demonstrated the alternative approach through examples. We now make our proposal explicit in the form of a four-step procedure:

X Identify the X that is being presented as evidence. State this in the form of a proposition[14] concerning X.

Y Identify the conclusion Y for which X is evidence. State this in the form of a proposition.

B Identify the *background beliefs* that a smart and generally well-informed opponent might contradict in order to justify believing X but *not-*Y.

A Put all these propositions into a non-deductive *argument*, and pick a qualifier (e.g., *'likely'*) that is proportional to the support that X and the background beliefs provide for Y.[15]

[14] A proposition, recall, is a statement that can be true or false. "A photograph of the rock" cannot be true or false but "there is a photograph of the rock" can.

[15] For people following the [9] approach, this step will be done a bit differently, but the concept remains the same, and the two approaches are mutually consistent.

This list captures the primary steps required but likely does not represent the typical workflow of completing them. Building a compelling argument often requires iterative re-writing of the individual claims to best balance the strength of the argument as a whole. It is common to modify the conclusion — to weaken it, ordinarily — when one reflects explicitly on the premises one can confidently add in support. It is also the case that as the argument is worked on, new background beliefs might be added, and existing background beliefs might be removed. The end result is to be gauged holistically considering the plausibility of the premises together with their structural support for the conclusion.

Nowhere in the non-deductive arguments arising from applying our approach is the word 'evidence' necessary. This is a good thing, since (as we have outlined) the word is often problematically loaded. One often treats the first layer of evidence as a sturdy, sufficient foundation for supporting empirical hypotheses, but good practice requires digging deeper. We have plans for future papers that will dig deeper into what this digging deeper entails.

Although we know some readers will find our doing so redundant, we (almost) conclude this paper by emphasizing the following three key points. If you remember nothing else, please remember the step-by-step procedure just outlined and these points:[16]

1. The *concept* of evidence is crucial to making rational inferences about the empirical world.[17]

2. Because of the variety of meanings and connotations associated with the *word* 'evidence', its use can be a confounding hinderance to rational inference.

3. We believe our proposed approach provides a way to avoid these hindrances.[18]

May the concept of evidence live long, and may the word 'evidence' rest in peace.

[16] If your memory is up to the task, consider also remembering all of the points bolded in the text and not just those reiterated here.

[17] At the risk of over-emphasizing the obvious, engineering of all forms (including most importantly for our purposes, engineering to ensure the safety of systems) is about the empirical world.

[18] Or if not to avoid the hindrances entirely, at least to reduce the likelihood that these hindrances will result in more weight given to what it ought not, simply due to what a thing is called.

References

1. Quine, W. V. O.: Response to Lewis and Holdcroft. *Revue Internationale de Philosophie*, vol. 51, no. 202, 1997, pp. 575–577.

2. Ministry of Defence: *00-56 Safety Management Requirements for Defence Systems—Part 1: Requirements and Guidance, version 7*. (U.K.) Ministry of Defence, Glasgow, UK, Feb. 2017.

3. Kerr, O. S.: A Theory of Law. *Green Bag*, vol. 16, 2012, p. 111.

4. Schauer, F.: *The Proof: Uses of Evidence in Law, Politics, and Everything Else*. Harvard University Press, 2022.

5. Williamson, T.: *The Philosophy of Philosophy*. Wiley-Blackwell, 2007.

6. Williamson, T.: *Knowledge and its Limits*. Oxford University Press, 2000.

7. Achinstein, P.: *The Book of Evidence*. Oxford University Press, 2001.

8. Holloway, C. M.; and Wasson, K. S.: A Primer on Argument. White paper, National Aeronautics and Space Administration, Hampton, VA, USA, June 2021. URL `https://ntrs.nasa.gov/citations/20210019993`.

9. Wasson, K. S.; and Holloway, M.: An Introduction to Constructing and Assessing Overarching Properties Related Arguments (OPRAs). White paper, NASA Langley Research Center, January 2022. URL `https://ntrs.nasa.gov/citations/20210025425`.

10. Hansson, S. O.: From the casino to the jungle: Dealing with uncertainty in technological risk management. *Synthese*, vol. 168, no. 3, 2009, pp. 423–432.

11. de la Vara, J. L.; García, A.; Valero, J.; and Ayora, C.: Model-based assurance evidence management for safety–critical systems. *Software and Systems Modeling*, 2022.

12. Knight, J.: *Fundamentals of Dependable Computing for Software Engineers*. CRC Press, Boca Raton, Florida, 2012.

13. ARP4754A: *Guidelines for Development of Civil Aircraft and Systems*. SAE International, Dec. 2010.

14. ARP4761: *Guidelines and Methods for Conducting the Safety Assessment Process on Civil Airborne Systems and Equipment*. SAE International, Dec. 1996.

15. RTCA: *DO-178C Software Considerations in Airborne Systems and Equipment Certification*. Washington DC, USA (Also published as EUROCAE ED-12C), 2011.

16. *Massachusetts v. United States*, 333 U.S. 611, 639-640 (1948) (Jackson, R., dissenting).

17. Holloway, C. M.: The Friendly Argument Notation (FAN). Technical Memorandum NASA/TM-2020-5002931, National Aeronautics and Space Administration, Hampton, VA, USA, June 2020. URL https://ntrs.nasa.gov/citations/20205002931.

Appendix: Examples recast into FAN

Here are all of the examples used in the paper written in the Friendly Argument Notation (FAN) [17]. The notation[19] is designed to be self-explanatory.

The deductive version for the argument for the rock breaking the window is as follows:

`Believing`

> The window was broken by a rock. {Y}

`is required by these premises`

> There is a rock on the floor beyond the broken window. {X}
>
> The rock was not in that exact position before the breaking of the window. {B1}
>
> There are no other objects in the vicinity of the window that would cause it to break if launched at it. {B2}
>
> If there is a rock on the floor beyond a broken window, and the rock was not in that exact position before the breaking of the window, and there are no other objects in the vicinity of the window that would cause it to break if launched at it, then the window was broken by a rock. {C}

Here is the non-deductive version of the same argument:

`Believing`

> The window was broken by a rock. {Y}

`is justified by these premises`

> There is a rock on the floor beyond the broken window. {X}
>
> The rock was not in that exact position before the breaking of the window. {B1}
>
> There are no other objects in the vicinity of the window that would cause it to break if launched at it. {B2}

[19] The version of FAN used here is a modest enhancement of the original version described in the referenced paper. A new paper explaining the enhancements will be published soon.

The argument in Illustration 1 looks like this:

`Believing`

> Jess is eligible and fit for flight. {Y}

`is justified by these premises`

> Jess meets eligibility criteria to pilot the aircraft in question, in which she must hold an applicable license, medical certificate, and type rating. {X}
>
> The licensing qualifications cover a codified base of knowledge and practical skills specific to the safe and efficient operation of aircraft. {B1}
>
> The medical qualifications cover a codified base of medical conditions and history specific to potential aircraft operational hazards. {B2}
>
> The type rating qualifications cover extensions to the licensing qualifications to account for additional operational and safety features of particular aircraft types. {B3}

Here is the Illustration 2 argument:

`Believing`

> Aircraft-level hazards have been identified and adequately characterized. {Y}

`is justified by these premises`

> Aircraft Functional Hazard Assessment (AFHA) has been conducted. {X}
>
> The AFHA was conducted according to standard practice by appropriate personnel. {B1}
>
> The aircraft and air operation details assumed during the AFHA are accurate. {B2}
>
> Assumptions about crew mitigations of hazardous aircraft states are accurate. {B3}
>
> The AFHA process has been endorsed by relevant aviation regulators. {B4}

Here is the argument used for Illustration 3:

`Believing`

> The software satisfies the system requirements allocated to it in this application. {Y}

`is justified by these premises`

> The aviation software item was deemed to meet the objectives of RTCA DO-178C at software level B. {X}
>
> Assessment of DO-178C conformance was conducted by qualified personnel. {B1}
>
> Conformance assessment by qualified personnel is sufficient to show satisfaction of DO-178C objectives. {B2}
>
> Meeting the objectives of DO-178C ensures adequate quality of this software. {B3}

As an added bonus for those who have taken the time to read this Appendix, here is an argument that was not made explicitly in the main body. It is intended to capture the essence of the text in Footnote 13. Does it? Is it a deductive argument?

Bonus Argument for Illustration 3:

`Believing`

> The software satisfies its safety objectives. {Y}

`is justified by these premises`

> The software satisfies the system requirements allocated to it in this application. {X}
>
> The system requirements allocated to the software include requirements to ensure satisfaction of safety objectives. {B1}

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
01-04-2023	Technical Memorandum	3/2021-3/2023

4. TITLE AND SUBTITLE
Towards a Coherent View of Evidence in Safety Assurance

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
The Rock Hunters: Finlay McCardel, C. Michael Holloway, Kimberly Wasson, Neil McDonnell, Mallory Graydon, Abel Peña, Sarah Lehman

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
NASA Langley Research Center
Hampton, Virginia 23681-2199

8. PERFORMING ORGANIZATION REPORT NUMBER
L–

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
National Aeronautics and Space Administration
Washington, DC 20546-0001

10. SPONSOR/MONITOR'S ACRONYM(S)
NASA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
NASA/TM–2023–20230003336

12. DISTRIBUTION/AVAILABILITY STATEMENT
Unclassified-Unlimited
Subject Category 60
Availability: NASA STI Program (757) 864-9658

13. SUPPLEMENTARY NOTES
An electronic version can be found at http://ntrs.nasa.gov.

14. ABSTRACT
This paper explains why labeling something as evidence does not make it special and proposes a systematic approach for avoiding giving the word more authority than it warrants in rational inquiry about the real world.

15. SUBJECT TERMS
philosophy, evidence, argument, safety, epistemology

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			STI Information Desk (help@sti.nasa.gov)
U	U	U	UU	24	19b. TELEPHONE NUMBER *(Include area code)* (757) 864-9658