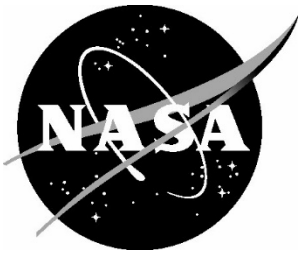
NASA/TM–20240015747



Detangling Safety Knowledge with Contextualism

Mary Z. Chen
Intern, NASA Langley Research Center
University of Oxford, Oxford UK

January 2025

NASA STI Program Report Series

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 NTRS Registered and its public interface, the NASA Technical Reports Server, thus providing one of the largest collections 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 has 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.

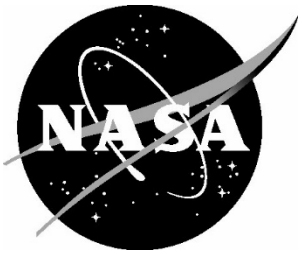
Specialized services also include organizing and publishing research results, distributing specialized research announcements and feeds, providing information desk and personal search support, and enabling data exchange services.

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

- Access the NASA STI program home page at http://www.sti.nasa.gov

.

NASA/TM–20240015747



Detangling Safety Knowledge with Contextualism

Mary Z. Chen
Intern, NASA Langley Research Center
University of Oxford, Oxford UK

National Aeronautics and
Space Administration

Langley Reasearch Center
Hampton, Virginia 23961-2199

January 2025

Acknowledgements

Many thanks to C. Michael Holloway and Sarah Lehman for their support and guidance on this paper, and to Mallory Graydon, George Romanski, and Kim Wasson for helpful comments and discussion on safety regulations.

Available from:

NASA STI Program / Mail Stop 050
NASA Langley Research Center
Hampton, VA 23681-2199

Abstract

We have contradictory intuitions about knowing that things are safe: on the one hand, we claim to have knowledge that many systems (even complex ones) are safe in ordinary language. On the other hand, when challenged by an expert, we easily admit that we do not, in fact, possess very much safety knowledge at all. This Technical Memorandum (TM) attempts to reconcile these two competing intuitions using a philosophical view called **epistemic contextualism**.

We can use epistemic contextualism to explain the importance of safety regulations and the goal in implementing them: first, to differentiate types of disagreement and establish a common basis upon which regulators and regulatees can know that a system is safe; second, to raise ordinary consumers' standards for safety knowledge (given adequate knowledge of the comprehensiveness of the regulation). This TM also uses the contextualist framework of relevant alternatives to elucidate types of failure possibilities and explain complications in attempts to prevent common mode failures.

1 The Problem

I seem to know an awful lot. My brother will disagree, but I do – I know that my dog will be eagerly waiting at the door when I wake up to take her on a morning walk. I know the view outside my window this morning is a one of a quintessential New England autumn, complete with a blaze of vibrant leaves. I know the F3 key on the piano in the living room is ever so slightly out of tune. I know I have a brother named Michael who does not think I know very much at all. Like I said, I seem to know an awful lot.

But it seems it does not take very much for me to reconsider whether I really possess all of this knowledge. Maybe my dog will have slept in this morning for the first time ever and *won't* be waiting outside my door. Maybe overnight, an especially strong storm of which I was unaware blew all the leaves from the trees and the scene outside my window will be more reminiscent of the aftermath of a hurricane. Maybe the humidity in my living room has increased *just* enough for the wood of the F3 key to expand and for the note to recover its proper pitch. And maybe I have just woken up from an especially vivid dream and I do not really have a brother named Michael. In fact, maybe I am Keanu Reeves in *The Matrix* and my entire life has been one long hallucination wherein everything I think I have experienced is not real. Maybe I do not know very much at all.

'Now let's take a step back, Mary,' you might say. 'This isn't a philosophy classroom where I'm being deceived by an evil demon, Descartes-style. This is real life, and no one in real life thinks they live in *The Matrix*.' But even if we toss out the previous *Matrix* reference, it is clear that there is some discord in what exactly we really know, or at least how we talk about what we really know. This is an especially salient question when our safety and lives are at stake. How can we *really* know that something is safe? This matters when we decide what form of transport to take, which life jacket to buy, or whether to participate in some activity at all. Knowledge of

safety (or lack thereof) also bears on attributing responsibility to organizations and individuals when things go wrong.

These are all big questions. I will now (somewhat) sidestep them and deal with a smaller, though related and still significant, one: how can we make sense of the fact that we seem to make all of these contradictory knowledge claims? How can we claim on the one hand that we know an awful lot, and on the other that we do not *really* know because we cannot be certain? An extension to this question, as we will see, applies to our attribution of knowledge to other people.

The big picture claim of this article is that *we can appeal to context to explain why our knowledge attributions in various situations differ*. The more specific and controversial claim that I will argue for is that the best context-based solution is one where *the context of the attributor of a particular knowledge claim is the context that ultimately matters* – a philosophical theory called epistemic contextualism.

An important application of this theory is in knowledge about the safety of complex systems. We can use epistemic contextualism to explain the importance of safety regulations and the goals in implementing them: (1) to establish a common basis upon which regulators and regulatees can know that a system is safe and (2) to bring consumers to a minimum level of ordinary 'knowledge' that a system is safe.

1.1 Case study: An ordinary problem about safety, knowledge, expertise.

I will work with and add to the following short story throughout this article to help clarify the claims that I am making. The purpose of the story, as with this introduction, is to make seemingly outlandish claims (for instance, 'we seem to think we know a lot and very little at the same time') more palatable and convincing.

New Car. Before he finally takes the leap and purchases a new vehicle, John makes sure to do his due diligence. He checks reviews from the most respected car experts, goes to the dealership to test drive potential options, and has even seen the crash test statistics and videos online. He eventually chooses a car that is widely thought of as clunky-looking but reliable. His friends, who have never gone beyond the first page of an internet search, are bemused by John's meticulousness. They usually find the dealership nearest to their house and are easily swayed by a limited-edition color and a savvy salesman. But John argues, 'Even if all my research has not resulted in my buying the flashiest or trendiest car, <u>at least I know it is safe</u>,' and his friends concede this to be true.

So far, so good? John and his friends each make reasonable comments, none of which seem to raise any immediate red flags. He does know that the car is safe, as both he and his friends have agreed upon – he has done all of his research and prioritized safety as the most important feature of the vehicle he intends on purchasing. Keeping all of this 'reasonableness' in mind, let's carry on.

New Car, continued. Jeanine disagrees with John and his friends. She is a safety engineer who was, in fact, responsible for the design of an integral component of the vehicle's safety system. 'John does *not* know that his car is safe,' she argues. 'He doesn't really understand how the safety system works. Moreover, even *I* don't truly know that the car is safe, because I only designed one feature of the car, and so do not possess a complete understanding of how my safety system fits in with the rest of the vehicle.'

John and Jeanine make competing and prima facie (on the face of it) contradictory claims. Broadly we can sort people who agree and disagree with sets of their knowledge. These views might be summarized in the following table, categorized by how each response to the statements:

Ordinary. John (the layperson) knows the car is safe.

Expert. Jeanine (the expert) knows the car is safe.

Table 1: Attitudes Toward Safety Knowledge

		Ordinary	
		Affirm	*Deny*
Expert	*Affirm*	Ordinary (John)	Deferrer to Experts
	Deny	Conspiracy Theorist	Skeptic (Jeanine)

These four views are detailed below:

> *Ordinary* (i.e. John). The ordinary individual like John might think that we can know lots of stuff! With enough research and due diligence, we can even have knowledge of complex systems even though we don't understand the ins and outs of these systems. For instance, I know that the foundations of my house are safe even though I am not a civil engineer. I know that my phone will charge when I connect it to my usual USB-C charger and cable that is plugged into the appropriate outlet. More generally, a person can possess knowledge that a system is safe and will function as expected without full corresponding knowledge of exactly how the system works.
>
> *Deferrer to Experts*. The deferrer to experts will affirm that experts have knowledge of complex systems but deny that ordinary people have that same knowledge. This is intuitively a very reasonable position – for instance, we might say that a doctor knows that a certain course of treatment is best for a niche medical condition, and that an ordinary person would not be in the position to know because they do not understand the intricacies of the case.
>
> *Skeptic*. The skeptic, or also just a humble expert like Jeanine, will deny both ordinary and expert knowledge. This stanceis well-captured by the question, 'But do you *really* know?'. This captures a key intuition about system knowledge, that is, that one needs to have a full understanding of how a system works in order to truly know that it will be safe. According to the skeptic, the standard of knowledge about systems is very high, and neither the expert nor the lay person can meet that standard.

The fourth possible position is the *Conspiracy Theorist*, who denies expert knowledge but affirms ordinary knowledge. This is not quite the classic conspiracy theorist position, e.g. someone who believes that pigeons are spies for the US military and all the bird experts are in on it, because it requires the expert to genuinely believe that pigeons are real (the standard conspiracy theorist would have it that the expert intentionally misleads the public about their actual beliefs). In any case, I do not consider this a tenable position whose intuitions need to be validated by contextualism.

Excluding *Conspiracy Theorist*, these models reflect individually intuitive and yet competing and inconsistent sets of beliefs: they cannot be true at the same time. This leads to an unappealing but inevitable conclusion that we should reject at least two of them. If I want to be consistent in my communication, it seems I am forced to concede that I do not know the view outside my window, the flatness of my F3 key, and even that I have a brother named Michael.

Is this a truly inevitable conclusion or is there a way out? A less immediately obvious option is that the three models are not actually inconsistent; they only *seem* inconsistent. We would then need a theory that can explain both of these claims (i.e., why the models are consistent and why the models seem inconsistent). If we succeed with this attempt, we will have a theory that reconciles the three intuitive models and denies the fourth by fulfilling the following conditions:

1. *(Ordinary)* Theory grants laypeople some kind of knowledge about systems they do not completely understand (in at least some cases),
2. *(Deferrer to Experts)* Theory respects the knowledge of experts and the value of rigorous testing to ensure safety,
3. *(Skeptic)* Theory is intellectually humble and recognizes the limitations, even of experts, to predict all possible scenarios, and
4. *(contra Conspiracy Theorist)* Theory does not give undue weight to non-evidence-based opinions.

We've got a big task ahead of us, but one philosophical theory of knowledge might be up to the task: **epistemic contextualism**.

2 What is Epistemic Contextualism?

Epistemic contextualism is, roughly, the philosophical position that:

EC. The meaning of a knowledge attribution (a claim of the form '*Someone* knows that *some claim that can be evaluated true/false'*, e.g. '*Jeanine* knows that *the car is safe'*) depends partly on the context of the attributor (the person making the knowledge claim) (Rysiew, 2023).

There are probably a few questions that jump out immediately, and one of the most important might be, how exactly is this supposed to help us? The simple answer to that question is that **EC** opens the door for both John's claim ('I know that the car is safe') and Jeanine's claim ('John

does not know that the car is safe') to be true on the basis that the *attributor's* context is different. Therefore, **EC** helps us reconcile John and Jeanine's seemingly contradictory yet individually intuitive knowledge attributions. The rest of this section sets out and defends contextualism in a question-and-answer format.

2.1 Q: Don't we mean what we say?

The basic idea here is that we are in control of the language that we use, and that if we meant something different then we would have said something different. Why should we think that the meaning of what we say changes sneakily without our noticing?

A: No, not always. Other ordinary examples include 'all', 'flat', and 'tall'.

I will argue that a general contextualism – that the meaning of the things that we say depends on the context in which we say them – is highly intuitive. For instance, after doing a final sweep of my room for hoarded empty mugs before starting the dishwasher, I might say, 'All the mugs are in the dishwasher.' It is *implicit* in my statement that I am referring to all of the mugs *in my room* or all of the mugs *in my house*. Clearly, I am not referring to all of the mugs in my neighborhood, state, or country; the strictness of the term 'all', or what counts in the total set, is limited by my context. Other contextual terms that we ordinarily use include 'flat' or 'tall': what it means when we call an object 'flat' or someone 'tall' will depend on what the standard of comparison and the practical purpose of evaluation.

When we look at these other ordinary examples, I think we should at least concede that it would not be so strange for something like 'know' to also be context-dependent. Ordinary language is also inherently vague, and so it would not be surprising for 'know' to capture an umbrella of different meanings and different levels of specificity.

2.2 Q: But doesn't 'knowing' requires certainty that other context-dependent terms don't?

Wait a second, you might say. 'Know' is not context-sensitive in the way that some of these other terms are. Knowing something seems to imply certainty in a way that these other terms do not. If I were not certain about something, I would use another word like 'pretty sure' or 'think' or 'believe' instead of 'know', and so intuitions from the previous cases I have mentioned are not applicable to knowing.

A: 'Knowing' does seem to require certainty. But so does 'flat'.

I agree that 'know' seems to imply certainty – in other words, that it seems that knowledge must be infallible and if I cannot eliminate a possibility in which the thing that I claim to know is false, then it seems that I cannot really know said belief. But conceding this point seems to lead to, as philosopher David Lewis puts it, a 'whirlpool of skepticism' (1996, p.550), where we do not have *any knowledge at all* because there are almost no cases in which we can eliminate every single one of those possibilities of falsity.

But the same can be said about flatness! Philosopher Peter Unger argues that flat is *not* a contextual term but an absolute one, on the following basis: for anything that you might call 'flat', if there exists something flatter, then the original thing is not actually flat (Unger, 1975). That means that hardly anything is actually flat: a mid-western highway is less flat than my dining room table, is less flat than a window, is less flat than a high-precision optical flat for laboratory use, is less flat than a superconducting surface at extremely low temperatures. If this is true, then it seems like hardly anything at all is flat!

But I think (and I believe most people would think) that there is intuitive reason to reject this because we meaningfully and truthfully make statements about such and such thing being flat in everyday life. David Lewis offers one way of doing this in 'Scorekeeping in a Language Game', in which he argues that skeptics like Unger are 'changing the score' on us (Lewis, 1979, p. 353).

What does this mean? Take the example of the mid-western highway going straight through Indiana's corn and soybean fields – when considering whether this highway is flat, the pebbles and occasional potholes are not relevant. This is because our standard of precision for road flatness is low. But then Unger brings up the table, which requires a raised standard of evaluation, and under this standard, the highway is no longer flat. But, as Lewis argues, 'That does not alter the fact that it *was* true enough *in its original context'* and the fact that the highway is flat in the original context and not flat in the new context do not contradict one another (Lewis 1997, p. 353). Lewis concludes, '[Unger] can indeed create an unusual context in which hardly anything can acceptably be called "flat", but he has not thereby cast any discredit on the more usual contexts in which lower standards of precision are in force' (Lewis 1997, p. 353); in other words, just because a mid-western highway is not considered flat by the standards of superconducting surfaces at near absolute-zero temperatures does not contradict its flatness at the lower standards of precision which apply to its normal use.

I've defended two points here: (1) that 'flat' implies certainty in the same way that 'know' does, and (2) that we can get out of Unger's unhappy skeptical conclusion that very few things are flat by appealing to contextualism about flatness.

The third claim I want to make is supported by the initial two, which is that (3) we can apply the same contextualism about knowing: when John talks about his knowing that the car is safe or Jeanine knowing that the car is safe, he is using an ordinary standard of knowledge analogous to an ordinary flatness standard for a highway. But then, Jeanine points out that John lacks some information that would help him *really* be sure that the car is safe. In doing so, she uses a higher standard of evaluation due to her status as an auto safety expert.[1]

---

[1] Here, I am using an idealization in which Jeanine and John are reasonable members of their communities (of experts and ordinary users, respectively). The higher standard and lower standard therefore ones that are accepted by each community, respectively.

2.3 Q: What is a standard, and how exactly does it matter for contextualism?

Thus far, I have been vague about what I mean about 'standards for knowledge attribution'. What exactly are these standards, and what does it mean to raise and lower them?

One immediate response is this: that standards for knowledge just refer to how 'sure' we must be of some claim in order to know it. The higher the stakes are for when we get it wrong, the surer we have to be. But this is not quite right: surely knowing something requires being sure of it! When I say that I am 'pretty sure' that the car is in the garage, we intuitively think that this means that I do not actually know the car is in the garage though I am reasonably confident that it is. In other words, we think that knowledge must be *infallible*, so a set of standards based on how sure we must be of a claim in order to know it will not work.

A: A standard is a sphere of relevant alternatives. The wider the sphere of alternatives, the higher the standard for correct knowledge attribution.

A more promising solution is this: that *relevant alternatives* are the key to setting standards of knowledge attribution (Lewis 1996, see also Goldman 1976, Dretske 1981, Heller 1995). Going back to the road example, what matters when calling the road flat is *comparison to other roads*. In most circumstances, dining room tables, windows, and optical flats are not relevant alternatives when evaluating the flatness of a highway, and so their flatness should not be considered when thinking about the flatness of the highway. To know that a highway is flat, we must eliminate all possibilities of **relevant** items of comparison that are flatter than the highway is. Different alternatives are relevant depending on the thing that we are evaluating and the purpose of comparison.

Let us now apply the same solution to knowledge attribution: correct knowledge attribution of a claim is a result of eliminating all **relevant** alternatives in which the claim is false. The sphere of relevant alternatives in which the car is *not* safe is much narrower for an ordinary car user than it is for a safety engineer whose job precisely *is* to think about less likely possibilities of when the car malfunctions. Jeanine's standard for safety knowledge attribution is therefore higher exactly because there are many more alternatives that are relevant to her in her role as a safety engineer.

2.4 Q: But if anyone's context matters, wouldn't it be the subject's context and not the attributor's?

Exactly whose context matters? In cases of self-knowledge, such as when John claims that 'I know the car is safe,' the answer is obvious – his own context is the one that matters. This is because John is both attributor (the one making the claim about knowledge) and the subject (the one to whom the knowledge is attributed). But what about the case where attributor and subject come apart, such as when John claims that 'Jeanine knows the car is safe', or when Jeanine claims that 'John does not know the car is safe'? More generally, in cases with an attributor, *A*, a subject *S*, and a proposition *p*, whose context matters to evaluate the truth of *A*'s claim that '*S* knows (or does not know) that *p*'?

A: Our intuitive answer of subject contextualism will not get us very far…

I think the obvious initial answer is that it is the *subject's* context that matters; after all, it is the subject about whom the knowledge claim is being made. This position is often called Subject-Sensitive Invariantism (SSI) or Interest Relative Invariantism (IRI) and is defended by, among others, John Hawthorne (2004) and Jason Stanley (2005). However, SSI/IRI will not be sufficient to solve our safety knowledge problem, because as a contextualism based on the subject is, in fact, not really a contextualism at all. Consider the following:

> John claims, 'Jeanine knows the car is safe.'
>
> Jeanine claims, 'I do not know the car is safe.'

If what mattered for these 'knowledge' claims to be true was the context of the subject, then these two claims cannot be true at the same time because Jeanine's context is what matters in both cases. So, we are left with two very unappealing options to choose from.

The first is that John's claim is true. This implies that John knows more about Jeanine's context than Jeanine does, which makes no sense! Surely Jeanine knows her own context better than John does, and we should reject John's claim that Jeanine knows the car is safe – after all, she is the expert. But as we will see, the second option is equally, if not more, unappealing.

The second is that Jeanine's claim is true. Why is this worse? This latter option brings us straight to the **Conspiracy Theorist** position, that the layperson has knowledge that the expert does not. If whether Jeanine knows the car is safe is always evaluated by Jeanine's relevant alternatives and whether John knows the car is safe is always evaluated by John's relevant alternatives, we end up in a situation where Jeanine does not know the car is safe because of her high standards but John *does* know that the car is safe.

A(continued.): … and the alternative, attributor contextualism, better reflects and explains the seemingly contradictory ways we speak about knowledge in the safety knowledge problem.

Attributor contextualism, though initially unintuitive, is what will help us reconcile John and Jeanine's claims because it gives us the correct relevant alternatives needed to vindicate our intuitions in this safety knowledge problem.

John's sphere of relevant alternatives is informed by his ordinary-ness. We might say that when he makes claims about people 'knowing that the car is safe', he is using 'know' in an ordinary-standard kind of way. Therefore, John can truthfully say, 'I know that the car is safe' because he has done his research, eliminated the relevant alternatives (e.g., that this car has failed safety tests, that this other car's crash statistics are bad, etc.), and properly ignored the rest of the alternatives. He can also, on the basis of Jeanine's expertise, truthfully say that 'Jeanine knows that the car is safe' using the same or similar standard of 'know.'

Jeanine's sphere of relevant alternatives is informed by her experience as an expert. Because she knows that John does not have her experience and does not have a thorough understanding of how automotive safety works, she can truthfully say that 'John does not know that the car is

safe'. In Jeanine's case, the term 'know' adheres to a much higher standard than what John says when he claims, 'I know that the car is safe.' Jeanine can *also* truthfully say 'I do not know that the car is safe' because she is utilizing a higher standard of knowledge and acknowledges the possibility that her lack of comprehensive knowledge of the entire car's system will mean that she does not know whether some feature(s) of the car is(are) not actually safe.

2.5 Q: Aren't Jeanine and John expressing legitimate disagreement, rather than merely talking over each other?

The contextualist can be construed to say something like the following: 'John's knowledge claim is true if he knows the car is safe according to an ordinary standard. Jeanine's knowledge claim is true if John does not know the car is safe according to a raised standard. It is perfectly possible for both of these statements to be true at the same time.'

This should probably raise some eyebrows, which naturally moves us into the next objection: but John and Jeanine need to be disagreeing about *something*, and the contextualist seems to say that they are simply talking over each other. If contextualism cannot account for disagreement, we should throw it out.

A: Contextualism accounts for disagreement and does so with greater nuance by allowing for disagreement about standards.

The contextualist might argue something like this in response: that Jeanine's claim is actually twofold, if the contextualist is to be believed: (1) that John does not know the car is safe according to a raised standard and additionally (2) *that the raised standard is the better one*. This is intuitive for knowledge about safety in particular, where well-being and life are potentially at stake: we are more inclined to think that it is better to be safe than sorry, and therefore that the raised safety standard is superior when compared to the ordinary. On the other hand, we might think that the ordinary standard is better in terms of practical actionability, in particular, because, for a higher standard, the cost of greater certainty may outweigh the marginal decrease in risk. An additional realm of disagreement would arise if Jeanine were to claim that John does not know that the car is safe even according to his own standard.

In this way, contextualism allows for a more nuanced explanation of disagreement by separating out a different way that we might disagree other than the normal making of contradictory statements. It elucidates how we *do* sometimes talk over each other, and our seeming disagreement is not about the surface-level 'fact-of-the-matter' but about underlying assumptions and terminology which must be clarified.

2.6 Summary

Hopefully, contextualism about safety knowledge is sounding a bit less strange than it did at the beginning of this TM. To recap, **EC** says that the meaning of knowledge attributions (e.g., 'Jeanine knows the car is safe') depends partly on the context of the attributor of said knowledge

(e.g., John or Jeanine). This sort of contextualism about 'knowing' is not an ad-hoc solution to our safety knowledge problem. Rather, attributions of 'knowing' are similar to attributions of 'flatness'. Both are, in a sense, objective notions unlike attributions of characteristics like 'tastiness' that are purely subjective. But at the same time, 'knowing', like 'flatness', depends on the setting of standards; though 'flatness' is an objective notion, the standard upon which we can truthfully say '[some item] is flat' varies depending on the nature of the item and purpose of the evaluation. This standard is the set of relevant alternatives that we need to consider in our evaluation of flatness. Likewise, the standards, or set of relevant alternatives, upon which we can truthfully say '[someone] knows [some claim about safety]' also vary, but on the basis of the attributor of knowledge (the person making the meta-claim).

Contextualism, far from preventing legitimate disagreement, actually highlights a more nuanced perspective on how we disagree with each other: that is, sometimes we make statements that contradict each other, and sometimes we simply talk over one another using different sets of standards.

3 A Brief Extension

Before drawing some more general conclusions about the value of safety regulations in our ordinary discourse, I want to briefly extend the scope of this article (which has largely focused on the possibility and quality of *ordinary* safety knowledge) and turn to the development of safety regulations themselves. In this addendum, I will show how contextualism can be used to elucidate types of relevant possibilities of failures and highlight key, though often overlooked, risks.

3.1 Three types of failure possibilities and transformational integrity

As Wasson and Voros (2024) point out, at the core of U.S. federal regulation of civil aircraft is the simple idea that 'they must perform their intended function safely in expected and foreseeable conditions'. There are therefore two often (but not always) overlapping spheres of relevant possibilities of failure that are considered in safety regulation: (1) those in which aircraft do not do what they are supposed to, and (2) those in which the aircraft are not safe. The overlap between these spheres can further be delineated into types of relevant possibilities of failure, the first two of which are types of functional failures:

> **(1) without (2):** The aircraft does not perform an intended function, but not in a way that makes it unsafe.
>
> **Both (1) and (2)**: The aircraft does not perform an intended function and this failure has safety consequences.

The third essential, and easily overlooked, possibility of failure can occur even when the aircraft performs the intended function:

(2) without (1) The aircraft performs an intended (and not inherently unsafe) function, but not in a way that is safe.

The goal to eliminate these three types of failure possibilities can be summarized by an "all and only" principle: 'the aircraft must be designed to do all that it is intended to and only that which is acceptable' (Wasson & Voros, 2024).

The way that this core idea is practically implemented in aircraft assurance involves *transformational integrity:* as regulatees move from abstract ideas to a final product, each more concrete actualization must remain consistent with the initial all and only principle. Through the lens of contextualism, preserving transformational integrity means that, at each step, relevant possibilities of failure are contained. These include possibilities of functional failure: the new iteration must be equivalent to the previous in terms of how well the aircraft does what it is supposed to. Importantly, however, preserving transformational integrity also means that the differences between the current and previous iteration do not create safety issues elsewhere in the system.[2]

3.2 Application: 'common mode failure', independent systems, and complications arising

One important way in which this third kind of failure appears is in the elimination of common-mode failures (CMFs), implemented through regulations like **FAR25-1309:** *No single event that causes a catastrophic error.*

What is a CMF? Consider a case where a plane has three computers that back each other up (if computer 1 fails, then one can use computer 2, then computer 3). Let's say that there is an error in the operating system in computer 1. What *should* happen in this case is that the flight management system can simply identify that computer 1 has run into an error and move to computer 2. But if computer 2 uses the same operating system as computer 1, a bug in the operating system will affect computer 2 as well, and then computer 3! This bug would be a case of a common-mode failure, where multiple failures across a system result from a single failure.

Let's say that we consider a relevant possibility of failure to be one in which the failure is serious enough to be considered a 'catastrophic error', and that in the case of the computers, the failure is serious if all three computers fail at the same time. One might be tempted to simply calculate this based on the assumption that the failure probabilities are independent: P(total computer failure) = P(C1 fails) * P(C2 fails) * P(C3 fails). Let's say that if it really *were* the case that the

[2] This process has been made more consistent and efficient by the creation of Advisory Circulars (ACs) by the Federal Aviation Administration (FAA), which are standards and guidelines that detail acceptable ways of preserving transformational integrity to comply with federal regulations.

failure probabilities were independent, P(total computer failure) will be under the threshold probability to count as a relevant possibility of catastrophic failure.

But in our simplistic case of a CMF, we can see that the chance of failure is actually P(total computer failure) = P(C1 fails) = P(C2 fails) = P(C3 fails); this represents a significantly higher probability that P(computer failure) will exceed the threshold probability and remain a relevant possibility of catastrophic failure.

Some regulatory bodies give a seemingly straightforward solution: if what matters is that the errors are independent from each other, then each of the three computers should simply be independently programmed so that the software is different in each. This does seem to provide a straightforward solution to our problem. If there is a mistake in computer 1's code that generates an error in a certain situation, that is unlikely to affect computer 2 because the two computers were programmed independently. In so doing, we have reduced the probability of that catastrophic failure, though by an unknown amount that depends on the degree to which the unidentified errors remaining in the independently developed software are disjoint.

However, reducing the probability of catastrophic failure in that one instance of error is not the only effect of implementing computers with different operating systems and more broadly, it is essential to consider other effects of switching out the computer software. More generally and from a contextualist framework, a particular system, let's call it *X*, is made up of different components *a, b, c,* and so on. When we think about the relevant possibilities of failure for *X*, we do just that – those possibilities are *specific to X*. So when we change a component of *X* to mitigate one possibility of failure, it follows that we have changed *X* and therefore the total sphere of relevant failure possibilities too! When I switch component *a* to a_1, the total change is not only whatever the difference between *a* and a_1, but also depends on how differences in the interaction between the new component a_1 and *b, c,* and all the other parts that comprise *X*.

This shows that there is hidden tradeoff in diversifying components of a system. Though we eliminate known CMFs by simply using different components, we also multiply relevant possibilities of failure because we need to account for the ways these new components interact with other parts of the system. To demonstrate that a system is safe, we need to show more than merely that the new component a_1 mitigates a possibility of failure that was present when *X* had component *a*. As systems become more complicated, this because an increasingly difficult task.

The moral of the story? The straightforward solution is not necessarily so straightforward after all, and diversifying components of a system to eliminate common-mode failures is useful only insofar as we can detect and eliminate the additional error possibilities that stem from the diversification itself.

4 General Conclusions and Applications

We return to our original question in Section 1 of how we can make sense of our ordinary discourse about knowledge, and safety knowledge in particular. In Section 2, I argued that epistemic contextualism provides a framework in which these seemingly contradictory intuitions

can be reconciled by identifying different standards (defined as spheres of relevant possibilities of failure) for safety knowledge that apply in different contexts. In Section 3, I zoomed in on aicraft safety assurance processes and have used the contextualist framework to identify spheres of relevant possibilities of failure, show how aircraft safety assurance processes work to contain these spheres, and elucidate hidden tradeoffs in component diversification.

In this final section, I will explain two more general ways in which safety regulations are valuable within the contextualist framework in converging and raising standards of safety knowledge between (1) regulators and regulatees, and (2) between ordinary consumers.

4.1 Regulation converges safety knowledge standards between regulators and regulatees

First, contextualism helps us see how regulations allow regulators and regulatees to arrive at a common epistemic ground. Recent work in epistemology has focused on the *social* aspects of knowledge and belief – how groups come to know and form justified beliefs. Jonathan Ichikawa points out how people in a conversation 'face considerable metasemantic pressure to converge on an epistemic standard, which will govern both [their] utterances. [They] are in effect deciding together whether to be in a high-standards context or a low-standards one' (Ichikawa, 2024).

The safety assurance and approval process is a formal way in which this convergence is done: regulators and regulatees come to an agreement to share beliefs about how a system will behave and a corresponding level of confidence in these beliefs (Wasson & Voros, 2024). In the contextualist framework, they must be able to determine the relevant particular possibilities of failure and eliminate those possibilities. In the extension to this article, I will elaborate further on this point and argue that contextualism helps to explain some issues arising from the regulations themselves.

4.2 Regulation raises and converges ordinary standards of safety knowledge

Second, contextualism helps explain how safety regulations have great potential to improve the quality of ordinary safety knowledge: they raise the ordinary standard of knowledge to give the layman *better* knowledge. When safety regulations function well, John might not know every safety feature that allows the automatic braking system to function properly, but he *does* know that a car he buys from a licensed brand and dealership has passed federal safety regulations to be eligible for sale. He knows that those regulations were written by experts who possess knowledge of the complex system that he does not possess. Knowing that the car complies with safety regulations and is road-worthy eliminates a host of relevant alternatives that John otherwise would not be able to eliminate.

A caveat and danger here is that safety regulations are not always comprehensive, and therefore may give ordinary consumers a false sense of confidence when the stringency of the regulations do not live up to consumers' expectations. This highlights the point made in the first sentence of this subsection: that safety regulations have great *potential* to improve the quality of ordinary

safety knowledge and will do so in cases where ordinary consumers have a reasonably accurate sense of the regulations' comprehensiveness. Regulators therefore have an important role in giving consumers a 'sense of' what the regulations are like, even when communicating the finer details is not necessary or feasible.

So, in a way, we decide whether we know an awful lot or not because, if contextualism is to be believed, we set conversational standards that govern how we speak about knowledge and truth. This is applicable not just in trivial matters like the color of fall leaves and flat piano keys, but also in significantly important matters like human safety and the preservation of life. Contextualism helps us see why it is good and essential for us to formalize these standards in the form of safety regulations.

References

1. Dretske, F. (1981). The Pragmatic Dimension of Knowledge [Article]. Philosophical Studies, 40(3), 363–378. https://doi.org/10.1007/BF00646423
2. Goldman, A. I. (1976). Discrimination and Perceptual Knowledge [Article]. The Journal of Philosophy, 73(20), 771–791. https://doi.org/10.2307/2025679
3. Hawthorne, J. (John P.). (2004). Knowledge and lotteries [Book]. Clarendon.
4. Heller, M. (1995). The Simple Solution to the Problem of Generality [Article]. Noûs (Bloomington, Indiana), 29(4), 501–515. https://doi.org/10.2307/2216284
5. Ichikawa, J. J. (2024). Epistemic Contextualism and the Sociality of Knowledge. In J. Lackey & A. McGlynn (Eds.), Oxford Handbook of Social Epistemology. Oxford University Press.
6. Lewis, D. (1979). Scorekeeping in a Language Game [Article]. Journal of Philosophical Logic, 8(3), 339–359.
7. Lewis, D. (1996). Elusive knowledge. Australasian Journal of Philosophy, 74(4), 549–567. https://doi.org/10.1080/00048409612347521
8. Rysiew, P. (2023). Epistemic Contextualism. In E. N. Zalta & U. Nodelman (Eds.), The {Stanford} Encyclopedia of Philosophy (Winter2023 ed.). Metaphysics Research Lab, Stanford University.
9. Stanley, J. (2005). Knowledge and Practical Interests (1st ed.) [Book]. Clarendon Press. https://doi.org/10.1093/0199288038.001.0001
10. Unger, P. K. (1975). Ignorance : a case for scepticism (1st ed.) [Book]. Clarendon Press. https://doi.org/10.2307/2218871
11. Wasson, K. S., & Voros, R. (2024). Deobfuscating Machine Learning Assurance and Approval [Proceeding]. IEEE/AIAA Digital Avionics Systems Conference, 1–10. https://doi.org/10.1109/DASC62030.2024.10749638