Safety Case Arguments

Chris Hobbs
cwlh@farmhall.ca

17th November 2021

International System Safety Society
Canada Chapter

Where are we?

~~Safety~~ Usefulness Standards

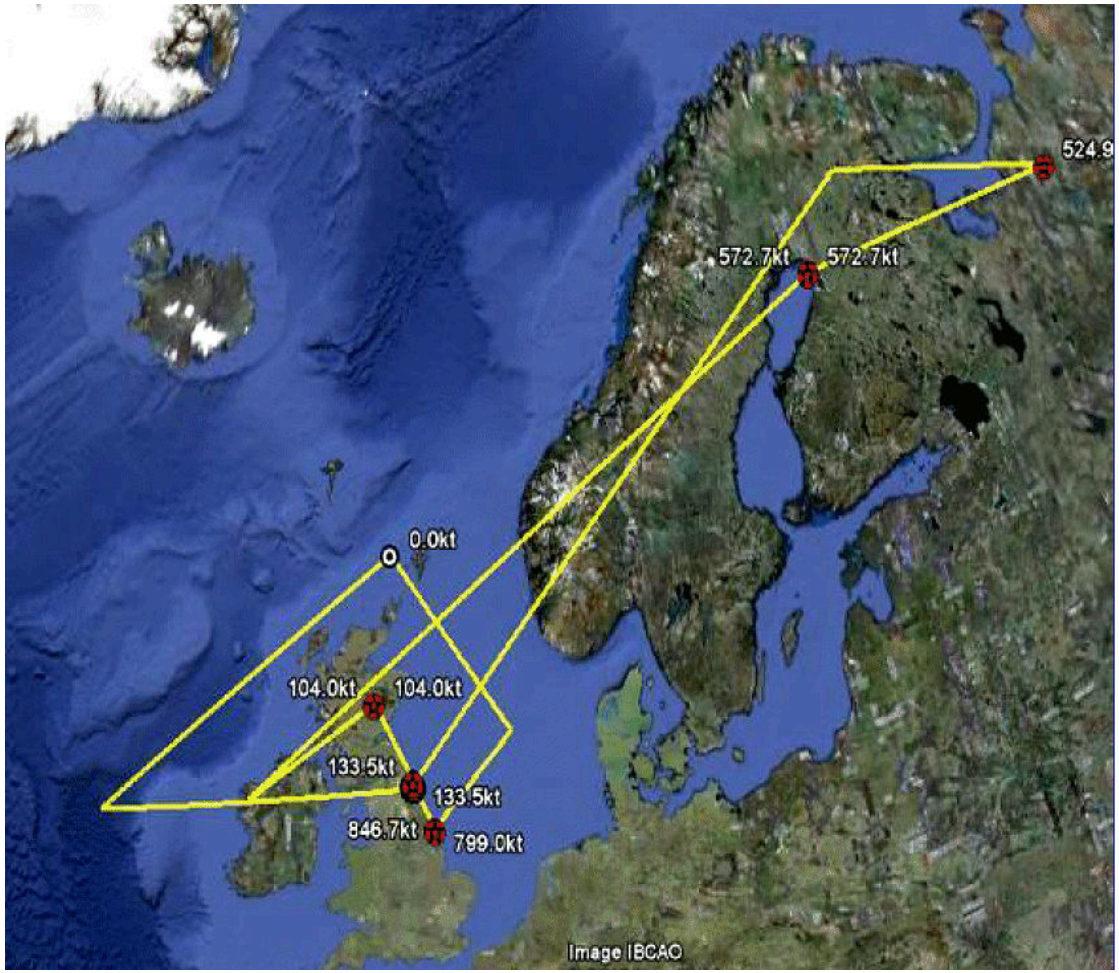
Producing a safe product is easy.

Producing a safe product that is useful is hard.

Terminology: "Accidental System"

2010 GPS Jamming Trials in North Sea

THV *Galatea*



With GPS jammed, position errors were expected.

Terminology: Accidental System

With GPS jammed, position errors were expected.

But the radar also failed.

No one knew that the radar depended on GPS: it was a system accidentally dependent on GPS. No one would have thought of testing it with a GPS failure.

Terminology: Accidental System

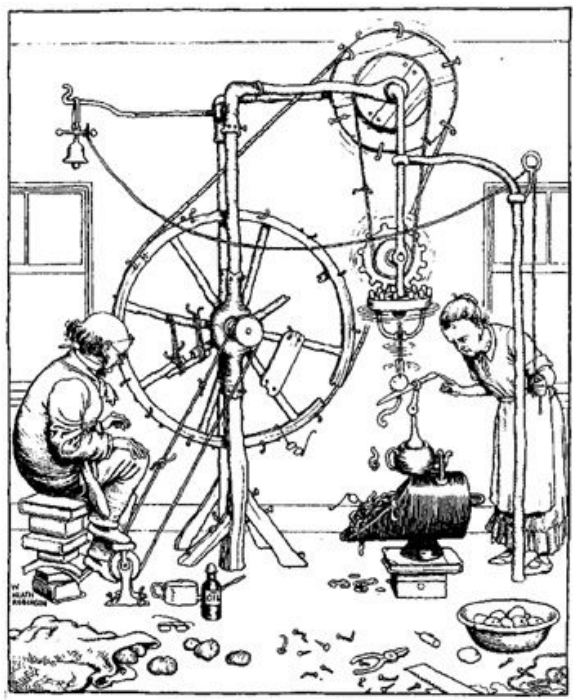
With GPS jammed, position errors were expected.

But the radar also failed.

No one knew that the radar depended on GPS: it was a system accidentally dependent on GPS. No one would have thought of testing it with a GPS failure.

Today, we don't fully understand the systems we build: they are too complex.

Most of our systems are accidental systems. How do we verify them?



The Professor's Invention for Peeling Potatoes

Terminology: SOTIF

SOTIF: Safety Of The Intended Functionality

Nothing broke, nothing malfunctioned, nothing failed. A dangerous situation still occurred.

One study found that over 90% of dangerous situations occurred although nothing broke, malfunctioned or failed — everything behaved exactly as designed.

ISO 26262-1: "**Hazard:** potential source of harm caused by *malfunctioning* behaviour of the item."

Most standards specifically exclude SOTIF.

Traditional methods of failure analysis do not acknowledge SOTIF.

A SOTIF example

- An autonomous car is travelling along a road with a manually-driven car close behind.

A SOTIF example

- An autonomous car is travelling along a road with a manually-driven car close behind.
- A child on a skateboard comes down a hill towards the road.

A SOTIF example

- An autonomous car is travelling along a road with a manually-driven car close behind.
- A child on a skateboard comes down a hill towards the road.
- The camera system **correctly** recognises that it is a child (with 73% probability: Blowing paper bag 15%, Dog 12%).

A SOTIF example

- An autonomous car is travelling along a road with a manually-driven car close behind.
- A child on a skateboard comes down a hill towards the road.
- The camera system **correctly** recognises that it is a child (with 73% probability: Blowing paper bag 15%, Dog 12%).
- The analysis system **correctly** measures its speed as 15 ± 2 km/hr.

A SOTIF example

- An autonomous car is travelling along a road with a manually-driven car close behind.
- A child on a skateboard comes down a hill towards the road.
- The camera system **correctly** recognises that it is a child (with 73% probability: Blowing paper bag 15%, Dog 12%).
- The analysis system **correctly** measures its speed as 15 ± 2 km/hr.
- The decision system **correctly** rejects the identification as a child because children do not travel at 15 km/hr except on bicycles — it must be something else (probably a paper bag).

A SOTIF example

- An autonomous car is travelling along a road with a manually-driven car close behind.
- A child on a skateboard comes down a hill towards the road.
- The camera system **correctly** recognises that it is a child (with 73% probability: Blowing paper bag 15%, Dog 12%).
- The analysis system **correctly** measures its speed as 15 ± 2 km/hr.
- The decision system **correctly** rejects the identification as a child because children do not travel at 15 km/hr except on bicycles — it must be something else (probably a paper bag).
- The decision system is faced with harming a "paper bag" or possibly harming a human by applying the brakes and **correctly** decides not to brake hard.

A SOTIF example

- An autonomous car is travelling along a road with a manually-driven car close behind.
- A child on a skateboard comes down a hill towards the road.
- The camera system **correctly** recognises that it is a child (with 73% probability: Blowing paper bag 15%, Dog 12%).
- The analysis system **correctly** measures its speed as 15 ± 2 km/hr.
- The decision system **correctly** rejects the identification as a child because children do not travel at 15 km/hr except on bicycles — it must be something else (probably a paper bag).
- The decision system is faced with harming a "paper bag" or possibly harming a human by applying the brakes and **correctly** decides not to brake hard.

Everything did what it was designed to do. Nothing failed. Nothing malfunctioned.

Terminology: Safety Case

The traditional definition:

> *A Safety Case is a structured argument, supported by a body of evidence, that provides a compelling, comprehensible and valid case that a system is safe for a given application in a given operating environment.*
>
> *DS 00-56 and many other sources*

This has been recognised as a dangerous definition.

Where are we?

The Safety Case

Also known as the "Safety Assurance Case".

The Boundary of the System

The system includes . . . and excludes . . .

The Safety Case

Also known as the "Safety Assurance Case".

The Boundary of the System

The system includes . . . and excludes . . .

The Claim

I claim that the system, when used as described in the Safety Manual, . . .

The Safety Case

Also known as the "Safety Assurance Case".

The Boundary of the System

The system includes . . . and excludes . . .

The Claim

I claim that the system, when used as described in the Safety Manual, . . .

The Argument

I argue that I meet my claim as follows . . .

Using Goal Structuring Notation (GSN), a Bayesian Belief Network (BBN) or SACM.

The Safety Case

Also known as the "Safety Assurance Case".

The Boundary of the System

The system includes . . . and excludes . . .

The Claim

I claim that the system, when used as described in the Safety Manual, . . .

The Argument

I argue that I meet my claim as follows . . .

The Evidence

The evidence that supports my argument is as follows . . .

Where are we?

The Loss of RAF Nimrod XV230

A Failure of Leadership, Culture and Priorities

From the Nimrod Report

The Nimrod Safety Case represented the best opportunity to capture the serious design flaws in the Nimrod which had lain dormant for years. If the Nimrod Safety Case had been drawn up with proper skill, care and attention, the catastrophic fire risks to the Nimrod MR2 fleet . . . would have been identified and dealt with, and the loss of XV230 in September 2006 would have been avoided.

Unfortunately, the Nimrod Safety Case was a lamentable job from start to finish. It was riddled with errors. It missed the key dangers. Its production is a story of incompetence, complacency, and cynicism. The best opportunity to prevent the accident to XV230 was, tragically, lost.

The Nimrod Safety Case

The term "Safety Case" appears 762 times in the report.
Chapters 9, 10 and 11 are dedicated to the Nimrod Safety Case.

"... the seeds of these problems were partly sown by Business Procedure 1201 which espoused an 'implicit Safety Case' ... based on a 'basic assumption that the aircraft is already operating to acceptable levels of safety.' The notion of an 'implicit' Safety Case is, however, something of an oxymoron. A Safety Case is intended to be an exercise in critical thinking and actual assessment of risk. An 'implicit' Safety Case, based on the assumption there are no actual risks, is the antithesis of this.

More from the Nimrod Report

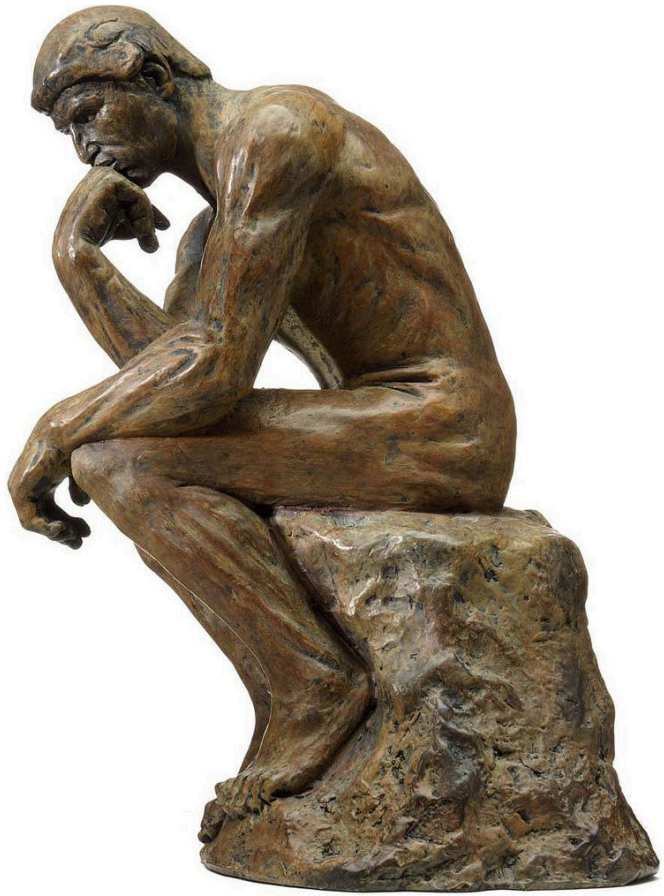
MR HADDON-CAVE QC: *How is it possible that you . . . approved the baseline safety case without ever having looked at it?*

MR MAHY: *Because the meetings that we went to with the IPT, the goalposts continually moved. . . . we were at the point where we've done everything that we've been asked to do . . . But, you know, we couldn't insist on them doing anything. We could only advise them.*

MR HADDON-CAVE QC: *What you could have done was say, "I'm sorry, we haven't seen the baseline safety case report, . . . we haven't read it and we certainly haven't had an opportunity to assess or audit it, . . . therefore we cannot possibly sign off the baseline safety case report. . . ."*

MR MAHY: *In hindsight, it would have been a better answer, yes.*

Where are we?



An Exercise

On the next slide I have written the rule for generating the next number in this sequence. You may guess numbers to discover the rule.

2, 4, 6, 8, 10, . . .

An Exercise

On the next slide I have written the rule for generating the next number in this sequence. You may guess numbers to discover the rule.

2, 4, 6, 8, 10, . . .

YES! The rule is:

Each number must be larger than the previous one

Confirmation Bias

We all suffer from confirmation bias. We only look for evidence that confirms what we already believe.

"The general root of superstition is that men observe when things hit, and not when they miss, and commit to memory the one, and pass over the other.

It is the peculiar and perpetual error of the human intellect to be more moved and excited by affirmatives than by negatives; whereas it ought properly to hold itself indifferently disposed towards both alike."

Francis Bacon
(1561-1626)

The Trap

Given the definition:

"A Safety Case is a structured argument, supported by a body of evidence, that provides a compelling, comprehensible and valid case that a system is safe for a given application in a given operating environment."

What happens when an engineer is asked to create a Safety Case?

The Trap

Given the definition:

"A Safety Case is a structured argument, supported by a body of evidence, that provides a compelling, comprehensible and valid case that a system is safe for a given application in a given operating environment."

What happens when an engineer is asked to create a Safety Case?

She looks for evidence that the system is safe!

The Trap

Given the definition:

"A Safety Case is a structured argument, supported by a body of evidence, that provides a compelling, comprehensible and valid case that a system is safe for a given application in a given operating environment."

What happens when an engineer is asked to create a Safety Case?

She looks for evidence that the system is safe!

And she may look for evidence before structuring the argument.

Pitfall: *Collection of data before an argument has been created is prone to be inappropriately used as evidence for that argument.* (UL 4600)

Where are we?

The challenge

We need to create an argument that our system is safe in a context where:

- our system is probably accidental. We do not know all of its interactions.
- our system will be deployed in environments we have not anticipated.
- we must consider its safety, even when everything works as designed.
- we are inherently biased. We look only for evidence that it **is** safe, and often we gather evidence before structuring the argument.

Guidance from UL 4600

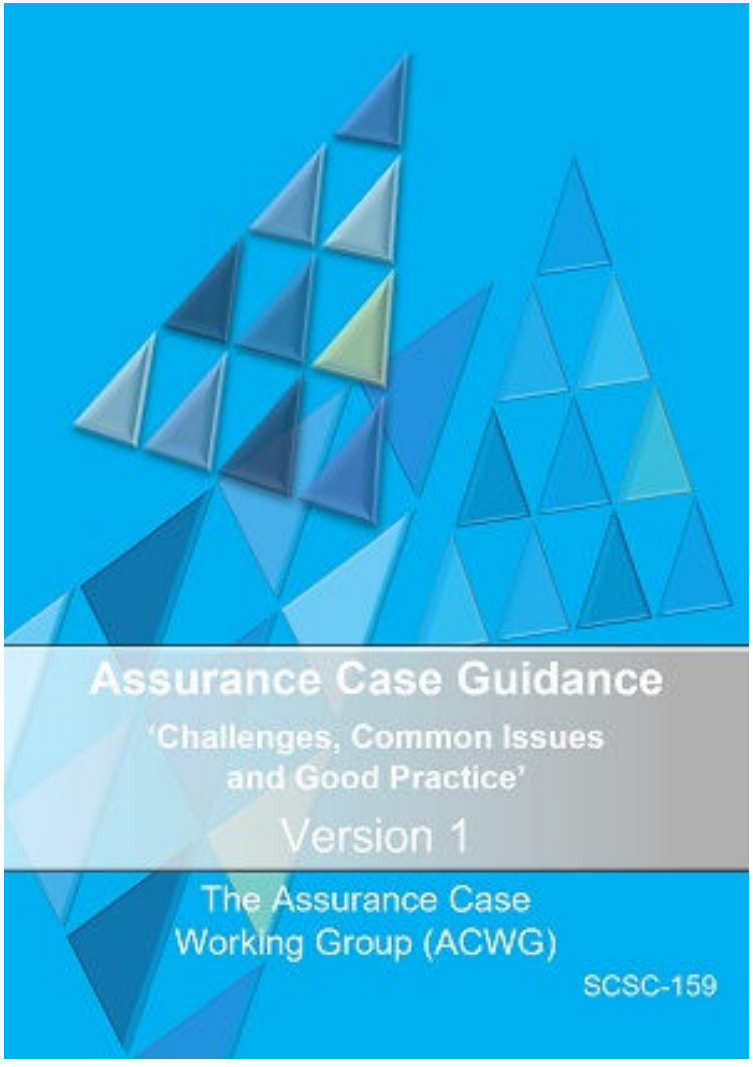
(Standard for Safety for the Evaluation of Autonomous Products)
UL 4600 is a goal-based standard that requires only a Safety Case.

- Prescriptive standards (e.g., IEC 61508, ISO 26262)
  This is how to build a safe, useful system: do X, Y and Z. Don't do A, B, C.
- Goal-Based standards (e.g., UL 4600)
  This is how to demonstrate that your final product is sufficiently safe: ...

UL 4600 also has useful information about the rôle of the assessor.

Philip Koopman 

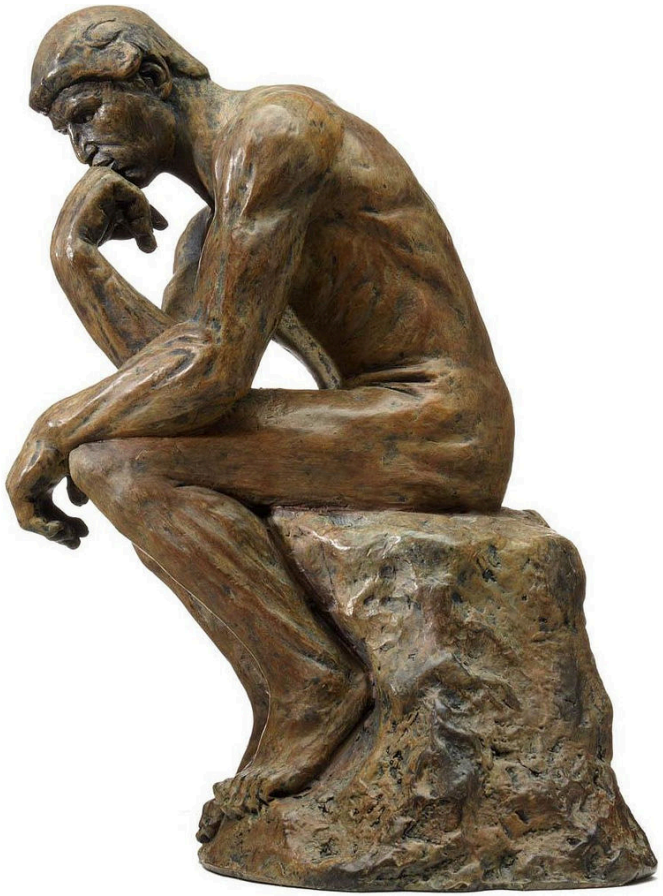
Guidance from the Safety Critical Systems Club



Version 1 issued in July 2021.

Contains guidance on avoiding bias.

Where are we?

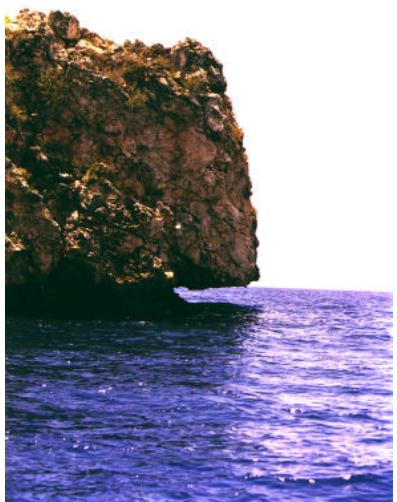


Adding doubt

See *Eliminative Induction: A Basis for Arguing System Confidence* by John B. Goodenough, Charles B. Weinstock and Ari Z. Klein.

Three types of doubt:

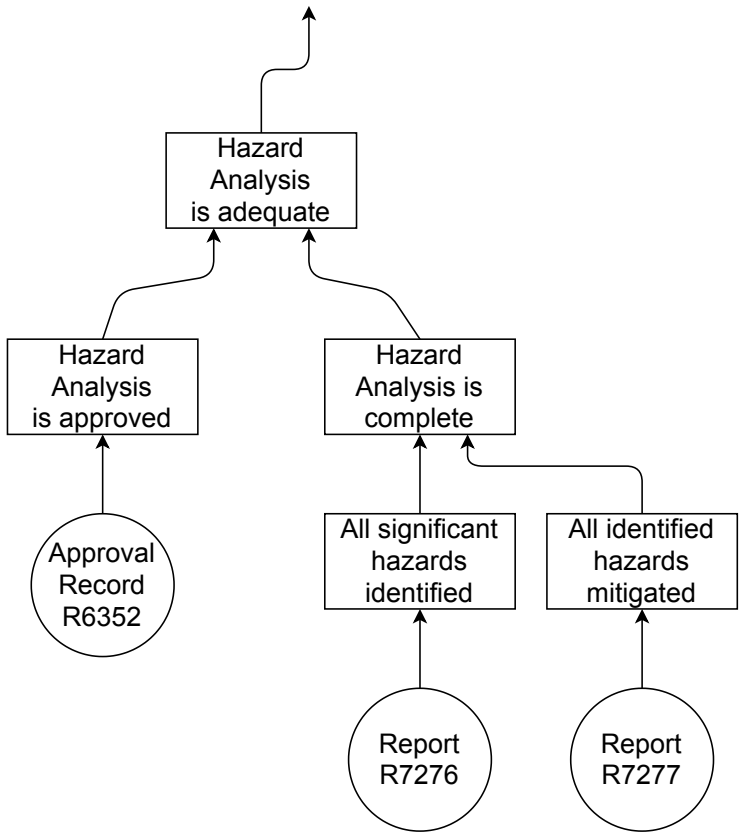
1. Rebutting: The claim is wrong: I have a counterexample.
2. Undermining: The evidence does not convince me.
3. Undercutting: The evidence is convincing, but it does not support the claim.



Adding doubt

Three types of doubt:

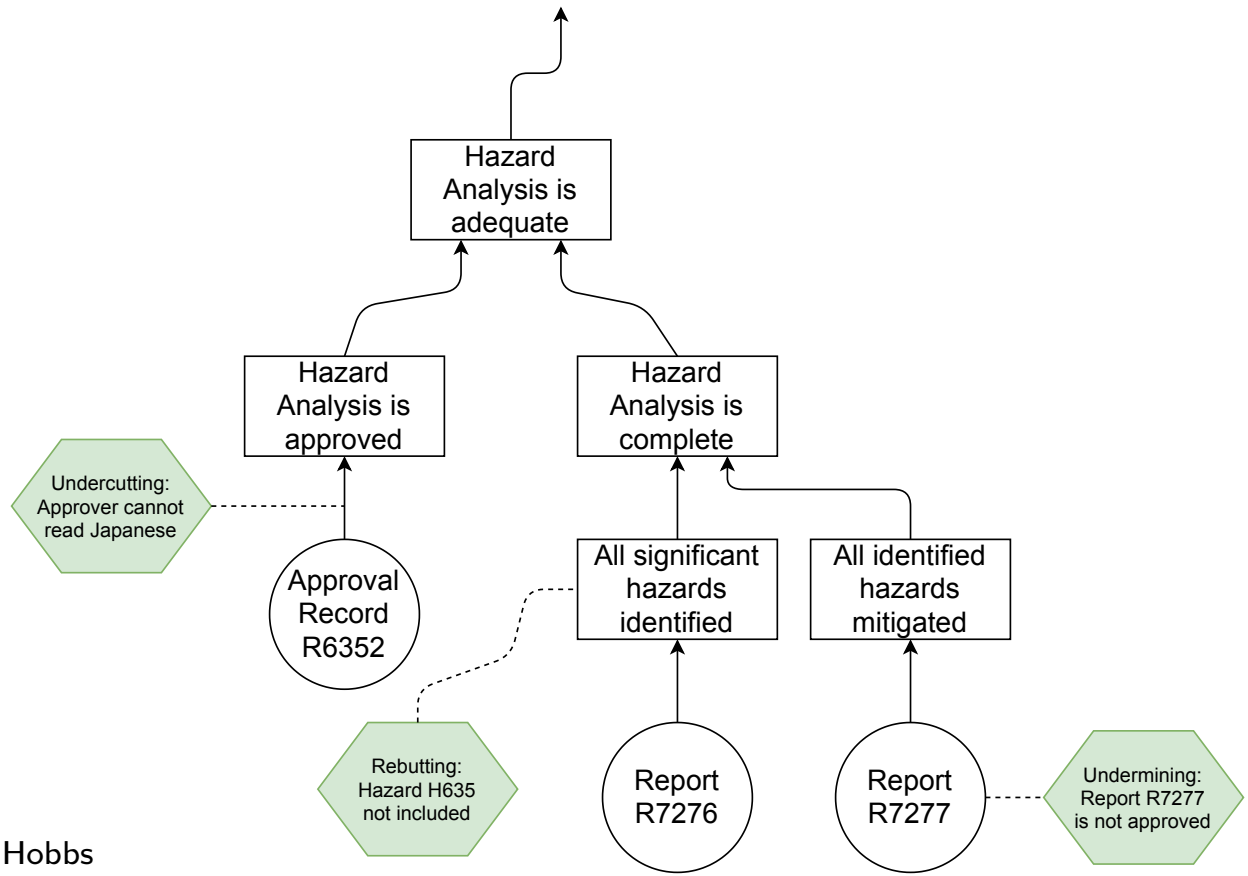
1. Rebutting: The claim is wrong: I have a counterexample.
2. Undermining: The evidence does not convince me.
3. Undercutting: The evidence is convincing, but it does not support the claim.



Adding doubt

Three types of doubt:

1. Rebutting: The claim is wrong: I have a counterexample.
2. Undermining: The evidence does not convince me.
3. Undercutting: The evidence is convincing, but it does not support the claim.



Adding doubt

Three types of doubt:

1. Rebutting: The claim is wrong: I can find a counterexample.
2. Undermining: The evidence does not convince me.
3. Undercutting: The evidence is convincing, but it does not support the claim.

~~Tell the engineer to produce a Safety Case to demonstrate that the system is safe.~~

Tell the engineer to collect everyone's doubts about the system's safety. And then try to eliminate those doubts.

This approach uses Confirmation Bias positively.

QNX's Experience

QNX first certified its Neutrino Operating System in 2010.

It recertified the OS several times with different certification bodies, each time producing a Safety Case acceptable to the assessor.

QNX's Experience

QNX first certified its Neutrino Operating System in 2010.

It recertified the OS several times with different certification bodies, each time producing a Safety Case acceptable to the assessor.

In 2018 it introduced "Eliminative Induction" and found 25 problems that had not been identified before!

There is a big difference between asking:

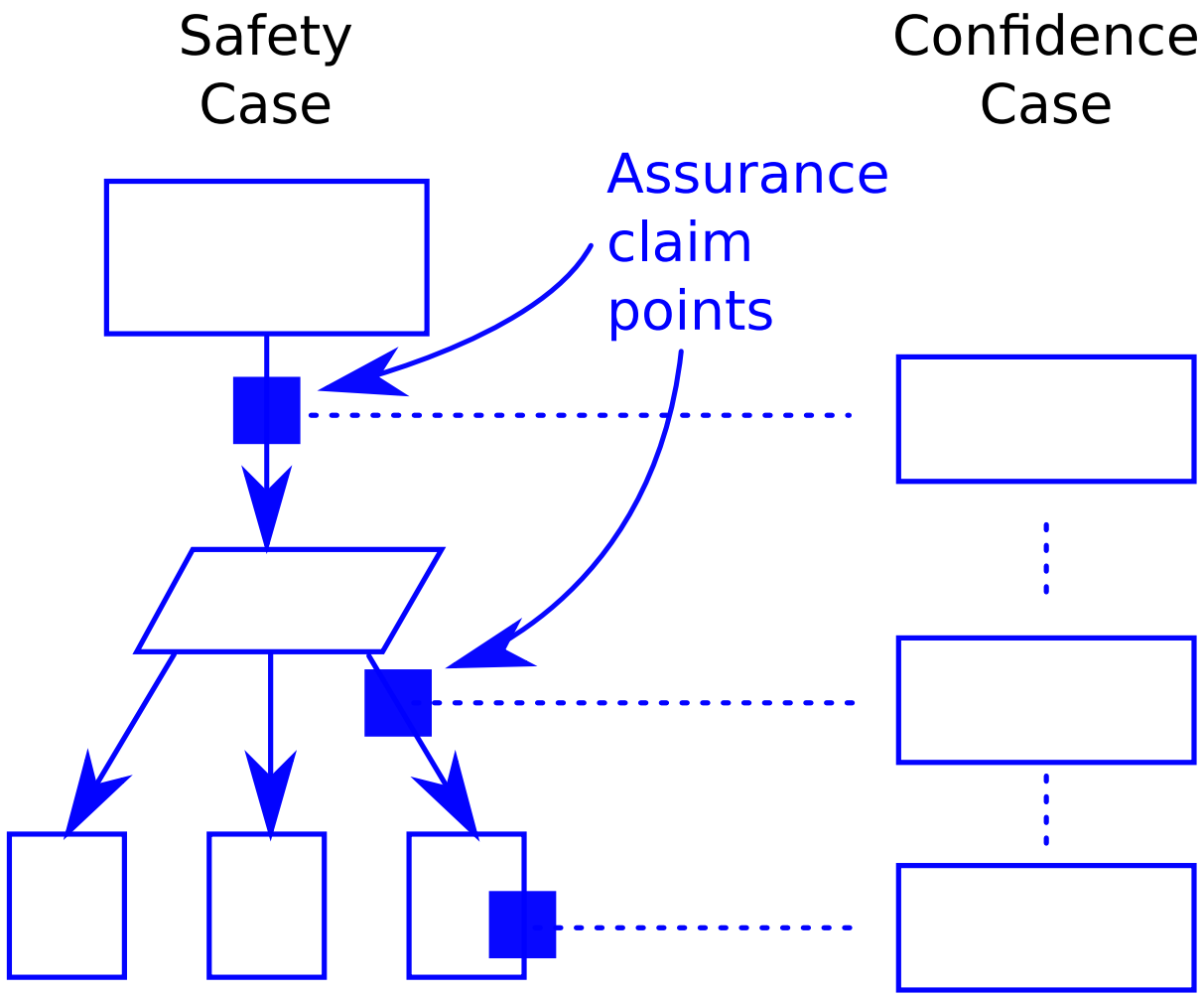
"Is process X being followed?"

and

"Can you think of any time when process X was not followed?"

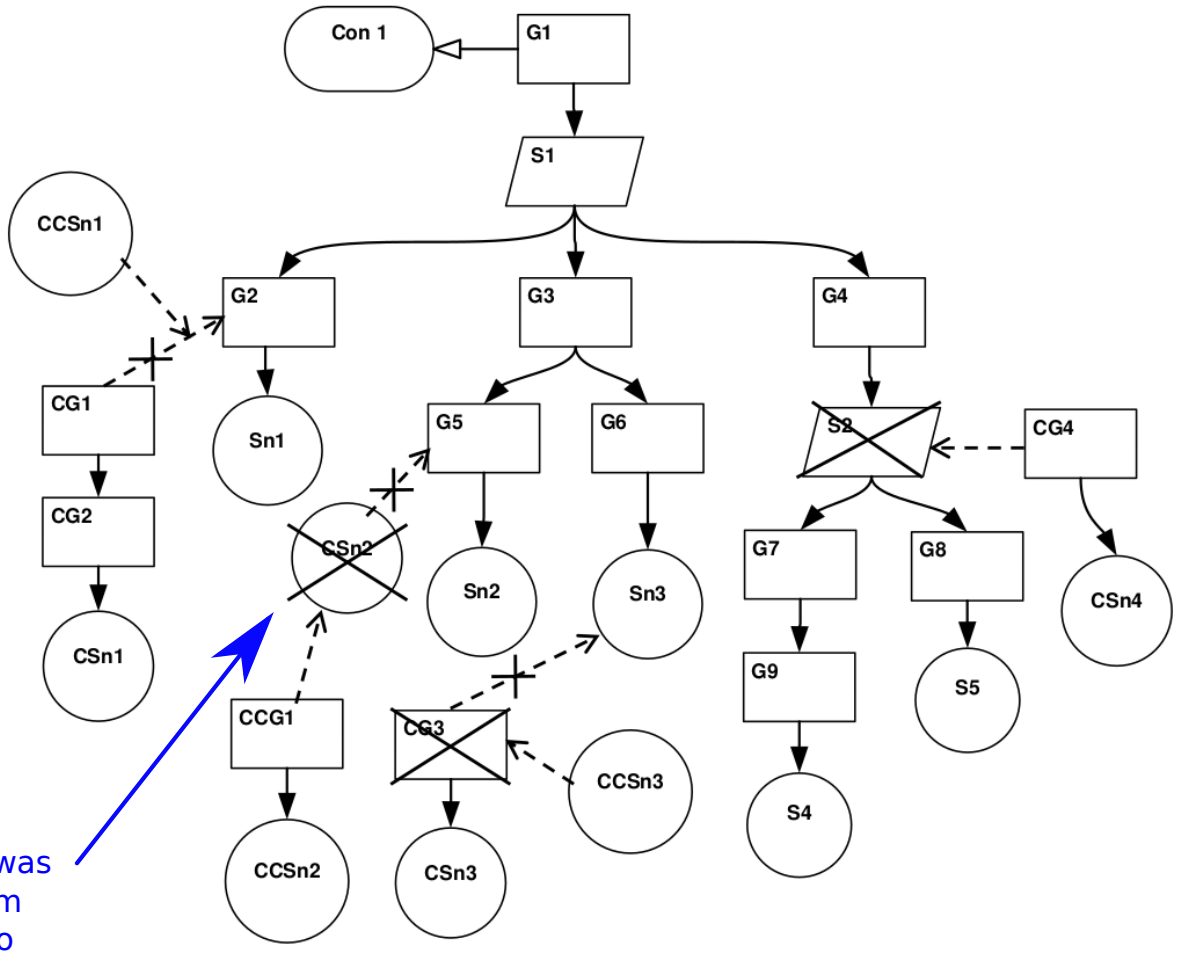
The Confidence Case

It can be useful to keep the
Safety Case and Confidence Case separate.



Big question

How should we record the doubts once they have been resolved?



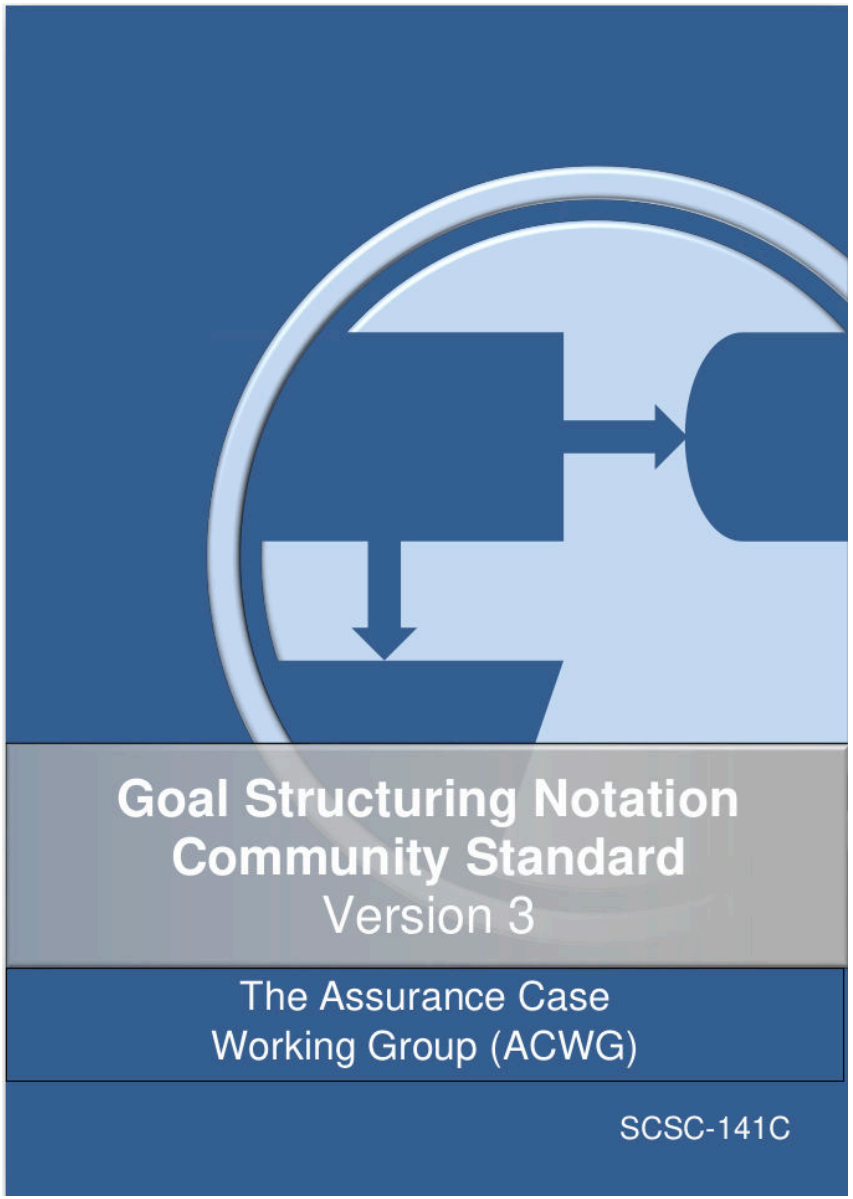
Counter-evidence was found, but the claim has been revised to accommodate it. The counter-evidence is no longer correct.

Big question

How should we record the doubts once they have been resolved?

- Keep them in the drawing?
  Makes the drawing difficult to read.
- Remove them from the drawing and just display the final version?
  Loses the history.
- Rely on the document management system to keep all the old copies?
  Difficult to follow the history.

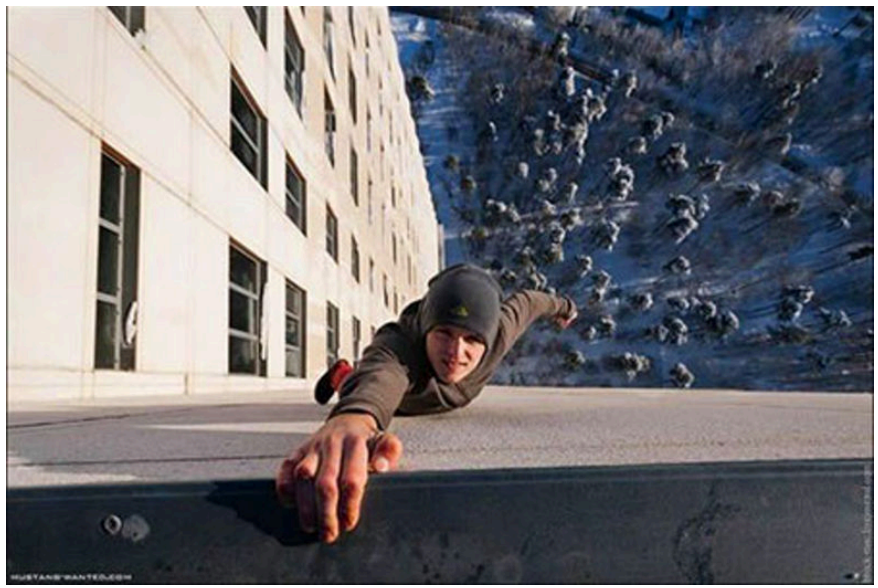
Version 3 of the GSN Standard



Version 3 of the Goal Structuring Notation (GSN) standard incorporates the symbols for adding doubt: known as "Dialectics".

Bayesian Belief Network (BBN) representations allow doubt to be incorporated.

Where are we?

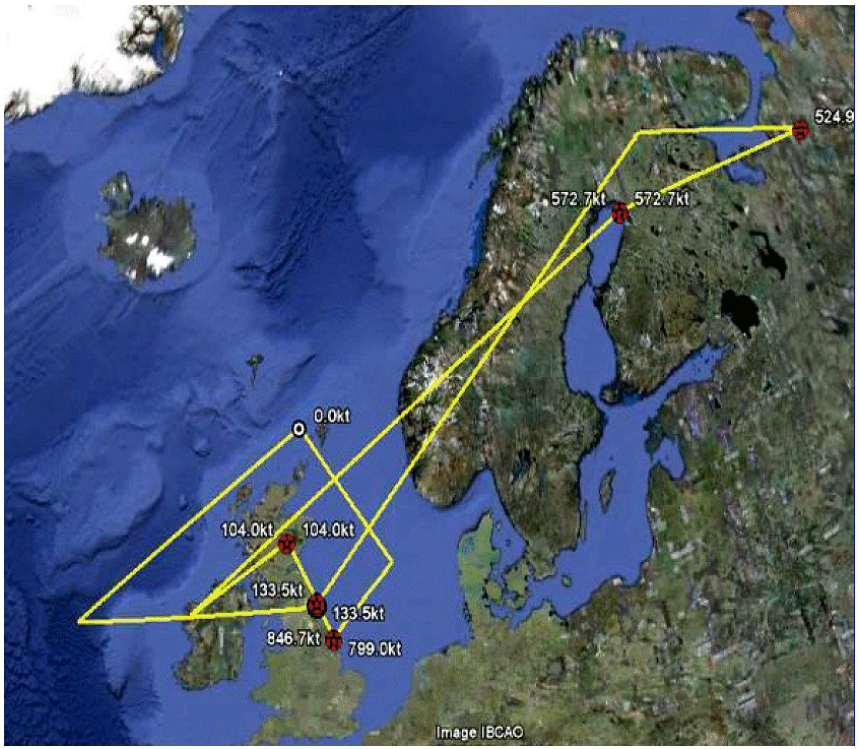


Updating the Safety Case

An autonomous or accidental system will meet conditions that were not anticipated when the Safety Case was created.

Remember the THV Galatea?

The Safety Case would not have considered GPS failure, because no one knew it depended on GPS.



A Dynamic Safety Case?

If we produce a (semi-)formal Safety Case, could the system in the field detect a condition not covered by its Safety Case?

If so, could the device:

- report this and allow human engineers to assess the new conditions rapidly to see whether the system is still safe?
- itself assess whether it is still safe?

The Safety Case for a drone assumes that there will never be more than 10 aircraft within a radius of 50 km. There are suddenly 11.

A Digital Twin as a Dynamic Safety Case?

"A digital twin is a computational model that evolves over time and continuously represents the structure, behavior and context of a unique physical asset such as a spacecraft, a person or even an entire city."

First use: Apollo 13 in 1970?

Now every Tesla car has a digital twin.

Can the digital twin act as a dynamic safety case?

Where are we?

Summary

- A Safety Case presents your argument as to why your system is sufficiently safe.

Summary

- A Safety Case presents your argument as to why your system is sufficiently safe.
- A Safety Case consists of a Claim, an Argument and Evidence.

Summary

- A Safety Case presents your argument as to why your system is sufficiently safe.
- A Safety Case consists of a Claim, an Argument and Evidence.
- Confirmation Bias makes it difficult for humans to create honest Safety Cases.

Summary

- A Safety Case presents your argument as to why your system is sufficiently safe.
- A Safety Case consists of a Claim, an Argument and Evidence.
- Confirmation Bias makes it difficult for humans to create honest Safety Cases.
- Adding doubts can present counterarguments.

Summary

- A Safety Case presents your argument as to why your system is sufficiently safe.
- A Safety Case consists of a Claim, an Argument and Evidence.
- Confirmation Bias makes it difficult for humans to create honest Safety Cases.
- Adding doubts can present counterarguments.
- Version 3 of the GSN standard includes the nomenclature for doubt.

Summary

- A Safety Case presents your argument as to why your system is sufficiently safe.
- A Safety Case consists of a Claim, an Argument and Evidence.
- Confirmation Bias makes it difficult for humans to create honest Safety Cases.
- Adding doubts can present counterarguments.
- Version 3 of the GSN standard includes the nomenclature for doubt.
- SOTIF and accidental systems are driving the need for a dynamic safety case.

Summary

- A Safety Case presents your argument as to why your system is sufficiently safe.
- A Safety Case consists of a Claim, an Argument and Evidence.
- Confirmation Bias makes it difficult for humans to create honest Safety Cases.
- Adding doubts can present counterarguments.
- Version 3 of the GSN standard includes the nomenclature for doubt.
- SOTIF and accidental systems are driving the need for a dynamic safety case.
- SOTIF and accidental systems are driving us away from prescriptive towards goal-based standards.

Questions? Answers?