The Danger of a "Safety Case"

Nancy Leveson

The concept seems to be spreading quickly that the safety case is the answer to our problems in safety engineering. After 32 years of experience in system safety engineering, I am afraid that we are in danger of becoming the proverbial lemmings when jumping on this bandwagon and we will all end up plunging into the sea.

There are lots of problems with the safety case but a basic one is that it is impossible to prove (or argue convincingly) for something that is untrue. No system is "safe." Anything, under the right conditions, can do harm. Therefore, every argument that a system is safe is going to be incorrect and leave out some potential cases under which an accident can occur. Such arguments are *always* subject to confirmation bias.

Is there anyone that has ever created a fault tree that had no potential branches? We try to put probabilities on the branches to somehow prove that the branch (cut set) is very unlikely, but it is impossible to provide such probabilities for every factor that could contribute to the cut set. So either numbers are made up or the factors for which such probabilities cannot be determined are omitted. In my 32 years, most of the accidents I have investigated had some type of analysis that showed the accident was "impossible" (10^-9 or even less per some unit of time). Sometimes the probabilistic arguments were used to avoid spending the resources necessary to prevent that accident scenario. I have a list of some of these examples for aircraft accidents on my website that was compiled by Follensbee, a former FAA employee.

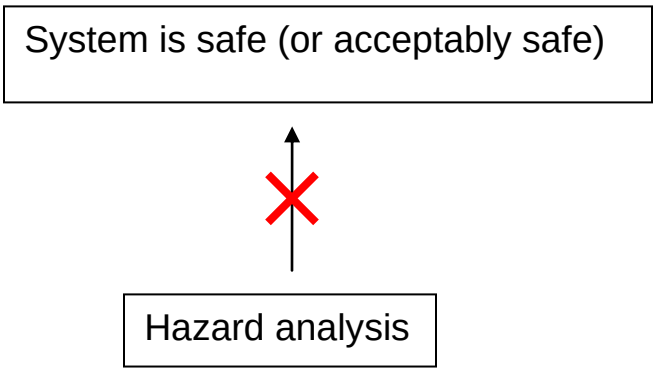
What about instead saying that we are arguing that the system is "acceptably safe"? There are two reasons why this won't work either. First, there is no definition of "acceptably." Acceptable to whom? To the person who is paying the cost of making it safe or to the potential victim? Both of these groups will have a different definition of "acceptable." We fall here into the traps of risk/benefit analysis, of which the Pinto is the most well-known example.

While engineers tell me they know what "safe" means and that they know they are not proving the system is safe, is that true of the management several levels up who will grab onto the fact that the system has been proven "safe"? Haddon-Cave in the Nimrod accident report noted the complacency that can result from doing a "Safety Case" and argued not only that the term not be used but that an argument in a "risk case" (as he suggested we call it) be very different than that used in a "safety case."

Why can't we argue for safety? Because there is no way to show safety, only unsafety. But, some tell me, isn't lack of finding ways the system is unsafe an argument that it is safe? Unfortunately, the answer is no, because our methods are inherently incomplete and easily done poorly.

What would such an argument for safety mean? Well, most of the safety cases I have seen use the following type of argument: My hazard analyses did not find a path to a

particular accident so that accident cannot happen. This logic is totally bogus. Even if done perfectly (and they never are), a hazard analysis can only find paths to particular hazards, they cannot show that such a path does not exist or that there are not unidentified hazards. If such analyses were possible and all identified hazards could be eliminated, we would have eliminated accidents and have perfectly safe systems now. Those of us who perform safety analyses do the best we can and know that we are not perfect. Even STPA, which in practice is finding more paths to accidents than older methods, is not perfect. So any argument in a safety case that:



is inherently incorrect before you begin.

What is safe or acceptably safe is a trans-scientific concept and cannot be answered by engineers or engineering methods.[1] Alvin Weinberg, former head of Oak Ridge National Laboratory, defined trans-scientific questions as those that lie between science and politics and can be stated in scientific terms but that in principle are beyond the proficiency of science to answer.[2]

> "In the current attempts to weigh the benefits of technology against its risks, the protagonists often ask for the impossible: scientific answers to questions that are trans-scientific."[2]

Engineers tread onto VERY dangerous ground when they claim they can answer philosophical, moral, and political questions with mathematics or analytical tools.

What is the alternative? Well, it is to continue to do what system safety has done for at least the past 60+ years (and often before that time but without the label "system safety"). We do hazard analyses that try to show how accidents could potentially happen and then design to eliminate or mitigate the scenarios we identify. We know we are imperfect in this task, but we don't have any better alternative. Even when we eliminate some hazards, we know that we have probably only added others (hopefully of less import) or we may be incorrect in our assumptions or analyses. We do the best we can. Then the documented analyses are reviewed by other experts to get some confidence in them. Even with multiple eyes, we know that there is still a potential for omissions or errors but resources are always limited and there will always be ways that accidents can happen. If our analyses and methods were perfect, we would not have all the accidents we have.

---

[1] See Section 1.4 titled "How Safe is Safe Enough" in my 1995 Safeware book.
[2] Alvin M. Weinberg, Science and Trans-Science, *Minerva*, 10_209-222, 1972,

Fukushima, Deep Water Horizon, Columbia, Bhopal, Air France 447—all of these losses occurred in systems where hazard analyses and sophisticated engineering techniques were used to try to prevent them.

After the hazard analysis is performed and engineers have done the very best they can to create a design that eliminates or controls the hazards judged (by various methods) to be the worst, they assemble what they have done and provide the results to decision-makers. Usually the evidence provided involves a list of the hazards they identified and why they thought these particular hazards were judged to be important as well as the specific analyses and design decisions that were made. It also should include the assumptions under which the analyses were done and the *limitations of what was done* (e.g., hazards that could not be eliminated, uncertainties in the analyses, uncertainties about the usage environment of the system, etc.).

The decision makers must then make a decision about whether the system is usable given the inherent uncertainties involved in the analyses, the design techniques used, and the assumptions under which the entire engineering effort was based. It's a difficult decision to make and I can understand why decision makers, especially those who are not engineers, would LOVE to have someone make an argument that the system is "safe." That makes their decision making easy. But that is not what the system safety engineers have shown. Alternatively, decision making becomes easy if decision makers are presented with a number like "10^-9' which is simply a way of providing them with an argument that the system is "acceptably safe." But such numbers are also inherently incorrect and misleading.

Decision making about safety is very difficult and error prone. A tremendous need exists for researchers to create ways for decision makers to better understand the risks involved in their decision making, not to mislead them by providing arguments that their systems are safe when no such logically correct argument is possible.