# White Paper on the Use of Safety Cases in Certification and Regulation[1]

Prof. Nancy Leveson
Aeronautics and Astronautics/Engineering Systems
MIT

## Preface

Although I had heard of safety cases some time ago, I never gave the topic much attention until I got involved with the Presidential Oil Spill Commission (on Deepwater Horizon) and started serving on an advisory committee to the new agency set up to regulate offshore oil drilling in the United States. Industry has exerted some pressure to adopt safety cases for this industry so the advisory committee has been carefully studying the ramifications of adopting such an approach. For the last couple of years I have been investigating this topic in the engineering and law literature and have become concerned by the push to use safety cases in the certification and regulation of systems in all industries, particularly for software-intensive systems. This paper describes what I have learned and my conclusions about the usefulness (and dangers) of this approach.

## Introduction

Certification of safety-critical systems is usually based on evaluation of whether a system or product reduces risk of specific losses to an acceptable level. There are major differences, however, in how that decision is made and on what evidence is required. The term Safety Case has become popular recently as a solution to the problem of regulating safety-critical systems. The term arises from the HSE (Health and Safety Executive) in the U.K., but different definitions seem to be rife. To avoid confusion, this paper uses the term “safety case” to denote the use of a structured argument for why the system is safe and examines the use of safety cases and some dangers associated with their use.

First, it should be noted that safety is a system property, not a property of the individual components. It makes no sense to talk about software or even a hardware component by itself being safe or unsafe except in a particular system. Many serious losses have occurred when software that was safe in one system was reused in another one [Leveson, 1995 and 2012].

Certification must consider all the aspects of the system, including the hardware, software, operators, and environment in which it will be used. Potential component failure is only one aspect that must be considered; certification must also consider system design flaws, unsafe interactions among components, human factors arising from the operator or user interacting with the system, and so on.

A final general principle is that certification is not a one-time effort. Systems must continue to operate safely throughout their life even when the system itself changes, the users and operators change their behavior over time, and the environment (and the assumptions about that environment used during design and the original hazard analysis) changes over time.

Any evaluation of certification approaches must consider these principles as well as other aspects, such as those in the legal, political, and moral realm. The first important distinction to consider is between types of regulation.

## Types of Regulation

Certification methods differ greatly among industries and countries. Approaches commonly used can be broken into two general types, which determine the type of evidence used in the certification process:

1. **Prescriptive**: Standards or guidelines for product features or development processes are provided that are used to determine whether a system should be certified.

[1] Parts of this paper appeared in the Nov/Dec 2011 issue of the *Journal of System Safety*.

a. **Product**: Specific design features are required, which may be (a) specific designs or (b) more general features such as fail-safe design or the use of protection systems.

   i. **Specific designs and features** provide a way to encode and pass on knowledge about past experience and lessons learned from past accidents. In some industries, practitioners are licensed based on their knowledge of the standards or codes of practice. An example is the existence of electrical codes based on past experience with various designs. For software, some completeness criteria for requirements have been identified [Leveson, 1995] as well as specific design features [Leveson, 2012] based on common flaws leading to many accidents in the past. Certification then becomes the responsibility of the licensed practitioner, who can lose their license if they fail to follow the standards. Organizations may also be established that produce standards and provide certification, such as the UL rating. It is difficult to fathom any argument that such encoded knowledge should not be included in any certification effort. Requiring reinvention of this past experience for every project would be prohibitively costly and potentially incomplete and error prone without any clear advantage.

   ii. Different industries face different safety problems and therefore the **general approach to safe design** may differ among them. For example, commercial aviation has created various types of fail-safe techniques used to protect against component failures [Follensbee]. Nuclear power, because of differences in the problem, has traditionally used defense in depth and protection systems. For software, such features might include the use of exception handling, checking for out-of-range variables, and designing to reduce the potential for human error when interacting with the software. Certification is usually provided by inspection that the design features provided are effective and implemented properly.

b. **Process**: Here the standards specify the process to be used in producing the product or system or in operating it (e.g., maintenance or change procedures) rather than specific design features of the product or system itself. Assurance is based on whether the process was followed and, sometimes, on the quality of the process or its artifacts. The process requirements may specify

   i. General product or system development processes and their artifacts, such as requirements specifications, test plans, reviews, analyses to be performed and documentation produced (e.g., DO-178B) or

   ii. The process to be used in the safety engineering of the system and not the general development process used for the product (e.g., MIL-STD-882). Only the safety engineering process is specified, not the general development process, which is up to the individual system developers.

2. **Performance-based or goal-setting approaches** focus on desired, measurable outcomes, rather than required product features or prescriptive processes, techniques, or procedures. The certification authority specifies a threshold of acceptable performance and often but not always a means for assuring that the threshold has been met. Basically, the standards set a goal, which may be a risk target, and usually it is up to the assurer to decide how to accomplish that goal. Performance-based regulation specifies defined results without specific direction regarding how those results are to be obtained. An example is a requirement that an aircraft navigation system must be able to estimate its position to within a circle with a radius of 10 nautical miles with some specified probability or that for new aircraft in-trail procedure (ITP) equipment “The likelihood that the ITP equipment provides undetected erroneous information about accuracy and integrity levels of own data shall be less than 1E-3 per flight hour” [RTCA, 2008].

### Common Certification Approaches in the U.S.

While in the past most certification was based on prescriptive methods (either product or process), there has been interest in performance-based regulation and assurance by government agencies, starting in the U.S. during the Reagan administration, often spearheaded by pressure from those being certified. A similar movement, but much more successful, was started in Great Britain around the same time, some of it stemming from the Cullen report on the Piper Alpha accident [Cullen, 1990].

Certification in the U.S. primarily uses prescriptive methods, but mixes the two types (product and process). Commercial aircraft, for example, are certified based on airworthiness standards requiring specific features (such as oxygen systems and life preservers), and more general features such as fail-safe design. Certification also requires the use of various types of safety analysis techniques, such as Fault Hazard Analysis, and general engineering development standards.

While the Nuclear Regulatory Commission requires prescriptive assurance for nuclear power plants, the American Nuclear Society in 2004 called for the use of risk-informed and performance-based regulations for the nuclear industry, arguing that

> "Risk-informed regulations use results and insights from probabilistic risk assessments to focus safety resources on the most risk-significant issues, thereby achieving an increase in safety while simultaneously reducing unnecessary regulatory burden produced by deterministic regulations" [American Nuclear Society, 2004]

Similar arguments have been made about FAA regulations and procedural handbooks being inflexible and inefficient and rule-making taking too long. Recommendations have been made to redesign the rulemaking process by moving to performance-based regulations where appropriate, but this type of certification is controversial, particularly with respect to how the performance goals are set and assured.

Sometimes certification is a one-time activity that follows the development process and occurs before the product or system is allowed to be marketed or used. More commonly and especially for complex systems, such as aircraft, nuclear power plants, and offshore oil exploration, certification may involve both initial approval and oversight of the operational use of the system. Changes to the original system design and certification basis may require recertification activities.

All certification is based on evidence that the certification approach has been followed. Inspection and test may be used if the certification is based on following a product standard. If the certification is based on the process used, engineering artifacts or analyses may be required and reviewed. Performance-based regulation may require a particular type of analysis (such as the use of specific types of probabilistic risk assessment) or may allow any type of reasoning that supports having achieved a particular performance goal.

As an example, the U.S. Department of Defense in Mil-Std-882 uses a prescriptive process that details the steps that must be taken in the development of safety-critical systems to ensure they are safe. The purpose of the SAR (safety assessment report), which is used as the basis for certification, is to describe the results of the prescribed steps in the standard. The SAR contains the artifacts of the prescribed process, such as a Safety Plan (which must be approved by the DoD at the beginning of the development of the system), a Preliminary Hazard Analysis, a System Hazard Analysis, a Subsystem Hazard Analysis, an Operating System Hazard Analysis, etc. The DoD evaluates the quality of the process artifacts provided in the SAR as the basis for approving use of the system.

While NASA has recently been influenced by the nuclear power community emphasis on probabilistic risk analysis, traditionally it has taken (and continues to emphasize) an approach similar to the U.S. DoD. The U.S. FAA (Federal Aviation Authority) approach for civil aviation has also been overwhelmingly prescriptive and the initial certification based on the quality of the prescribed process used to develop the aircraft and the implementation of various airworthiness standards in the aircraft's design. Operational oversight is based on inspection as well as feedback about the safety of the operations process. Recently, the FAA has moved to create a requirement for a safety management system by those developing or operating aviation systems in order to shift more of the responsibility for safety to the airframe manufacturers and airlines.

The type of evidence required is straightforward with prescriptive regulation, but performance-based regulation requires a more complex argument and evaluation strategy. While the term "safety case" may be used in prescriptive regulation, it is more commonly used in a performance or goal-based regulatory regime and that usage is followed here.

### Performance-Based Regulation and Safety Cases

Government oversight of safety in England started after the Flixborough explosion in 1974, but the term *safety case* seems to have emerged from a report by Lord Cullen on the Piper Alpha disaster in the offshore oil and gas industry in 1988 where 167 people died. The Cullen report on the Piper Alpha loss, published in 1990, was scathing in its assessment of the state of safety in the industry [Cullen, 1990]. The Cullen report concluded that safety assurance activities in the offshore oil industry were:

- Too superficial;
- Too restrictive or poorly scoped;
- Too generic;
- Overly mechanistic;
- Demonstrated insufficient appreciation of human factors;
- Were carried out by managers who lacked key competences;
- Were applied by managers who lacked understanding;
- Failed to consider interactions between people, components and systems.

The report suggested that regulation should be based around "goal setting" which would require that stated objectives be met, rather than prescribing the detailed measures to be taken [Whyte, 1997], i.e., performance-based rather than prescriptive. In such a regime, responsibility for controlling risks shifted from government to those who create and manage hazardous systems in the form of self-regulation. This approach has been adopted by the British Health and Safety Executive and applied widely to industries in that country.

The British safety case philosophy is based on three principles [Inge, 2007; Sutton]:

- Those who create the risks are responsible for controlling those risks
- Safe operations are achieved by setting and achieving goals rather than by following prescriptive rules. While the government sets goals, the operators develop what they consider to be appropriate methods to achieve those goals. It is up to the managers, technical experts, and the operations/maintenance personnel to determine how accidents should be avoided.
- All risks must be reduced such that they are below a specified threshold of acceptability.

When performance-based or goal-based certification is used, there are differences in how the performance or goals are specified and how the evaluation will be performed. In 1974, the creation of the Health and Safety Executive (HSE) was based on the principle that safety management is a matter of balancing the benefits from undertaking an activity and protecting those that might be affected by it, essentially cost-benefit analysis (CBA). The HSE also instituted the related concept of ALARP or "as low as reasonably practical" and widely used probabilistic risk analysis as the basis for the goals. Each of these is controversial [HSE, 2005].

The nuclear power industry was probably the first to use probabilistic risk analysis as a basis for certification. In the United Kingdom, the Nuclear Installations Act of 1965 required covered facilities to create and maintain a safety case in order to obtain a license to operate. The nuclear industry has placed particular emphasis on the use of Probabilistic Risk Assessment (PRA) with the use of techniques such as Fault Tree and Event Tree Analysis. Because of the use of standard designs in the nuclear power community and very slow introduction of new technology and innovation in designs, historical failure rates are often determinable.

Other potentially high-risk industries, such as the U.S. nuclear submarine community, take the opposite approach. For example, SUBSAFE does not allow the use of PRA [Leveson, 2012]. Instead, they require OQE (Objective Quality Evidence), which may be qualitative or quantitative, but must be based on observations, measurements, or tests that can be verified. Probabilistic risk assessments, for most systems, particularly complex systems, cannot be verified.

A second unique aspect of the British approach to safety assurance and required by the HSE is argumentation and approval based on whether risks have been reduced as low as is reasonably practicable (ALARP). Evaluating ALARP involves an assessment of the risk to be avoided, an assessment of the sacrifice (in money, time and trouble) involved in taking measures to avoid that risk, and a comparison of the two. The assumed level of risk in any activity or system determines how rigorous, exhaustive and transparent the risk analysis effort has been. "The greater the initial level of risk under consideration, the greater the degree of rigor required to demonstrate that risks have been reduced so far as is reasonably practicable" [Heiler, 2005].

The application of ALARP to new systems, where "reasonably practical" has not yet been defined, is questionable. Not increasing the accident rate in civil aviation above what it is today does seem like a reasonable goal given the current low rate, for example, but it is not clear how such an evaluation could be performed for the new technologies (such as satellite navigation and intensive use of computers) and the new and very different procedures that are planned.

There are also ethical and moral questions about the acceptance of the cost-benefit analysis underlying the ALARP principle. Steinzor [2010] claims that the risk levels tolerated by the British system conflict with both the spirit and the letter of American law. For example, British regulations allow safety cases to be no more protective than preventing one in 1,000 worker deaths and require operators to spend no more than $1.5 million per life saved. These standards are far more lax than comparable American legal requirements [Steinzor, 2010]. In addition, safety cases are strictly confidential in the U.K.; only company officials, regulators, and, in limited circumstances, worker representatives are allowed to see the entire plan. This type of confidentiality would be unlikely to be acceptable in the U.S.

While none of these more controversial aspects of assurance and certification need to be present when using a "safety case" approach, they are part and parcel of the history and foundation of safety cases and performance-based regulation.

### Potential Limitations of Safety Cases

A "safety case" may be and has been defined in many ways. In this paper, the term is used to denote an argument that the system will be acceptably safe in a given operating context. The problem is that it is always possible to find or produce evidence that something is safe. Unlike proving a theorem using mathematics (where the system is essentially "complete" and "closed," i.e., it is based on definitions, theorems and axioms and nothing else), a safety analysis is performed on an engineered and often social system where there is no complete mathematical theory to base arguments and guarantee completeness[2]

In fact, it can be argued that no system is completely safe so the goal of the argument is untrue before starting. If instead the goal is to show the system is "acceptably" safe, the problem simply devolves to defining what is "acceptable" and to whom: to the producer of the system who is paying the cost of making it safe or to the potential victim? The concept of ALARP is the British attempt to answer that question, but, as discussed above, that concept is unlikely to be acceptable in the U.S. legal system and the results of Pinto and other cases have demonstrated the fallacies in using cost/benefit analyses where the costs of fixing the gas tank design problem were clear but the number of victims and amount to be paid in liability claims was underestimated.

The main problem with the use of arguments for safety or acceptable safety in the safety case approach to certification lies in psychology and the notion of a mindset or frame of reference.

> "In decision theory and general systems theory, a *mindset* is a set of assumptions, methods or notations held by one or more people or groups of people which is so established that it creates a powerful incentive within these people or groups to continue to adopt or accept prior behaviors, choices, or tools. This phenomenon of *cognitive bias* is also sometimes described as *mental inertia*, *groupthink*, or a *paradigm*, and it is often difficult to counteract its effects upon analysis and decision-making processes" [Wikipedia].

[2] Even with such a mathematical basis, published and widely accepted mathematical proofs are frequently found later to be incorrect. They are not based on physical laws as in engineering.

An important component of mindset is the concept of confirmation bias. *Confirmation bias* is a tendency for people to favor information that confirms their preconceptions or hypotheses regardless of whether the information is true. People will focus on and interpret evidence in a way that confirms the goal they have set for themselves. If the goal is to prove the system is safe, they will focus on the evidence that shows it is safe and create an argument for safety. If the goal is to show the system is unsafe, the evidence used and the interpretation of available evidence will be quite different. People also tend to interpret ambiguous evidence as supporting their existing position [Dekker, 2006].

Experiments have repeatedly found that people tend to test hypotheses in a one-sided way, by searching for evidence consistent with the hypothesis they hold at a given time [Kunda, 1999; Nickerson, 1998]. Rather than searching through all the relevant evidence, they ask questions that are phrased so that an affirmative answer supports their hypothesis. A related aspect is the tendency for people to focus on one possibility and ignore alternatives. In combination with other effects, this one-sided strategy can obviously bias the conclusions that are reached.

Confirmation biases are not limited to the collection of evidence. The specification and interpretation of the information is also critical. Fischoff, Slavin, and Lichtenstein conducted an experiment in which information was left out of fault trees. Both novices and experts failed to use the omitted information in their arguments, even though the experts could be expected to be aware of this information. Fischoff *et al* [1978] attributed the results to an “out of sight, out of mind” phenomenon. In related experiments, an incomplete problem representation impaired performance because the subjects tended to rely on it as a comprehensive and truthful representation—they failed to consider important factors omitted from the specification [Vicente and Rasmussen, 1992]. It is very likely that this same non-recognition of omissions in the argument will occur when certifiers evaluate the evidence provided in safety cases.

Does this type of confirmation bias in safety cases occur or is it just a theoretical phenomenon? Every safety case published in papers promoting safety cases for software has seemed terribly flawed to me but obviously not to the authors or reviewers. For example, one argument was advanced to support the case that “the software was fault free.”[3] It’s not clear what “fault free” means in software or that this goal has ever been achieved in real software—errors have been found in nearly every piece of non-trivial software when used in operational settings. Part of the argument for this goal in the safety case is that the fault tree did not find any contribution of the software to an accident. While this argument has nothing to do with being fault free, it could potentially apply to the goal “the software will be safe.” The problem is that most fault trees are incomplete and not including software in them is common. In addition, design errors and interaction problems are rarely identified in fault trees. A second argument in support of the goal showing the software is fault free involves the evidence that “hazard-directed test results did not find any software faults leading to a hazard. It’s not clear what “hazard-directed testing” might be as it is not possible to test for safety.[4] More important, Dijkstra’s aphorism about testing being able only to show the presence of errors and not their absence is clearly true. Just because the testing did not find a fault in the software leading to a hazard is not proof that such faults to do exist. The same is true for hazard analysis: System safety engineers who perform hazard analyses are well aware that their analysis can, like testing, only identify some paths to some hazards (the ones analyzed) and does not in any way “prove” that the system will be safe or hazard free.

Confirmation bias problems are not easy to eliminate. But they can be reduced by changing the goal. A company in which the author is a co-owner was recently hired to conduct a non-advocate safety assessment of the new U.S. Missile Defense system for the hazard “inadvertent launch,” which was the major concern at the time [Pereira, Lee, Howard, 2006]. The system safety engineers conducting the independent safety assessment did not try to demonstrate that the system was safe, everyone was already convinced of that and they were going to deploy the system on that belief. The developers thought they had done everything they could to make it

[3] A reference is purposely not provided here for obvious reasons.

[4] Accidents often occur when the requirements are incomplete or wrong or assumptions about the usage environment are incorrect. Testing necessarily is based on the requirements and assumptions about the design and usage of the system.

safe. They had basically already constructed a “safety case” argument during development that would justify their belief in its safety. By law, however, the government was required to perform an independent risk analysis before deployment and field testing would be allowed. The goal of our independent assessment was to show that there were scenarios where inadvertent launch could occur, not to show the system was safe. The analysis found numerous such scenarios that had to be fixed before the system could be deployed, resulting in a six month delay for the Missile Defense Agency and expenditure of a large amount of money to fix the design flaws. The difference in results was partly due to the use of a new, more powerful analysis method but it also involved the different mindset and the different goal, which was to identify unrecognized hazards (ways inadvertent launch could occur) rather than to argue that the system was safe (that inadvertent launch could *not* occur).

Engineers always try to build safe systems and to verify to themselves that the system will be safe. The value that is added by system safety engineering is that it takes the opposite goal: to show that the system is unsafe. Otherwise, safety assurance becomes simply a paper exercise that repeats what the engineers are most likely to have already considered. It is for exactly this reason that Haddon-Cave recommended in the Nimrod accident report that safety cases should be relabeled “risk cases” and the goal should be “to demonstrate that the major hazards of the installation and the risks to personnel therein have been identified and appropriate controls provided” [Haddon-Cave, 2009], not to argue the system is safe.

A final potential problem with safety cases, which has been criticized in the off-shore oil industry approach to safety cases and with respect to the Deepwater Horizon accident (and was also involved in the Fukushima Daichi nuclear power plant events), is not using worst-case analysis [Houck, 2010]. The analysis is often limited to what is likely or expected, not what could be catastrophic. Simply arguing that the most likely case will be safe is not adequate: Most accidents involve unlikely events, often because of wrong assumptions about what is likely to happen and about how the system will operate or be operated in practice. Effective safety analysis requires considering worst cases.

But while theoretical arguments against safety cases are interesting, the proof is really “in the pudding.” How well have they worked in practice?

### Experience with Safety Cases

Unfortunately, careful evaluation and comparison between certification approaches has not been done. Most papers about safety cases express personal opinions or deal with how to prepare a safety case, but do not provide data on whether it is effective. As a result, there is no real evidence that one type of regulation or certification is better than another.

One way to compare approaches is to use past experience. The use of performance-based regulation has not necessarily proven to be better than the other approaches in actual use. One of the most effective safety programs ever established, SUBSAFE [Leveson, 2012], which has had no losses in the past 48 years despite operating under very dangerous conditions, is the almost total opposite of the goal-based orientation of the British form of the safety case. The spectacular SUBSAFE record is in contrast to the U.S. experience prior to the initiation of SUBSAFE, when a submarine loss occurred on average every two to three years. SUBSAFE uses a very prescriptive approach as does the civil aviation community, which has also been able to reduce accident rates down to extremely low levels and keep them there despite the tendency to become complacent after years of having very few accidents.

Despite claims of successful use of safety cases in offshore oil exploration in the U.K., accidents have occurred. In addition, a British study of conditions in the North Sea suggest alarming neglect of the physical infrastructure that ensures safety, further undermining, according to Steinzor [2010], claims that use of safety cases is as effective as its advocates claim.

The use or at least poor use of safety cases has been implicated in some accident reports. The best known of these is the Nimrod aircraft crash in Afghanistan in 2006. A safety case had been prepared for the Nimrod, but the accident report concluded that the quality of that safety case was gravely inadequate [Haddon-Cave, 2009]:

> “. . . the Nimrod safety case was a lamentable job from start to finish. It was riddled with errors. . . Its production is a story of incompetence, complacency, and cynicism … The

> Nimrod Safety Case process was fatally undermined by a general malaise: a widespread assumption by those involved that the Nimrod was 'safe anyway' (because it had successfully flown for 30 years) and the task of drawing up the Safety Case became essentially a paperwork and 'tickbox' exercise."

The criticisms of safety cases contained in the Nimrod report include:

- Use of safety cases has led to a culture of 'paper safety' at the expense of real safety. It currently does not represent value for money.
- The current shortcomings of safety cases in the military environment include: bureaucratic length; their obscure language; a failure to see the wood for the trees; archaeological documentary exercises; routine outsourcing to industry; lack of vital operator input; disproportionality; ignoring of age issues; compliance-only exercises; audits of process only; and prior assumptions of safety and 'shelf-ware'.
- Safety cases were intended to be an aid to thinking about risk but they have become an end in themselves.
- Safety cases for 'legacy' aircraft are drawn up on an 'as designed' basis, ignoring the real safety, deterioration, maintenance and other issues inherent in their age.
- Safety cases are compliance-driven, i.e., written in a manner driven by the need to comply with the requirements of the regulations, rather than being working documents to improve safety controls. Compliance becomes the overriding objective and the argumentation tends to follow the same, repetitive, mechanical format which amounts to no more than a secretarial exercise (and, in some cases, have actually been prepared by secretaries in outside consultant firms). Such safety cases tend also to give the answer that the customer or designer wants, i.e. that the platform is safe.
- Large amount of money are spent on things that do not improve the safety of the system

Haddon-Cave, the author of the Nimrod accident report, concluded that safety cases should be renamed "risk cases" and made the following recommendations (among others):

- Care should be taken when utilizing techniques such as Goal Structured Notation or 'Claims-Arguments-Evidence' to avoid falling into the trap of assuming the conclusion ('the platform is safe'), or looking for supporting evidence for the conclusion instead of carrying out a proper analysis of risk. (Note the similarity to the concerns expressed in earlier about mindset and confirmation bias.)
- Care should be taken when using quantitative probabilities, i.e. numerical probabilities such as $1 \times 10^{-6}$ equating to "Remote". Such figures and their associated nomenclature give the illusion and comfort of accuracy and a well-honed scientific approach. Outside the world of structures, numbers are far from exact.
- Care should be taken when using historical or past statistical data. The fact that something has not happened in the past is no guarantee that it will not happen in the future. Piper Alpha was ostensibly "safe" on the day before the explosion on this basis. The better approach is to analyze the particular details of a hazard and make a decision on whether it represents a risk that needs to be addressed.
- Care needs to be taken to define the process whereby new hazards can be added to the Risk Case, incorporated in the Hazard Log, and dealt with in due course, and how original assumptions about hazards or zones are to be re-examined in light of new events.
- Once written, the safety case should be used as an on-going operational and training tool. There are all too many situations where a comprehensive safety case is written, and then it sits on a shelf, gathering dust, with no one paying attention to it. In such situations there is a danger that operations personnel may take the attitude, "We know we are safe because we have a safety case".

**Practical Considerations in the Use of Safety Cases**

The validity of the argument in each safety case needs to be evaluated individually by a certification authority [Wassyng, 2011]. How can this be done? If every submission is different, it will be difficult to evaluate them in a systematic way, so certification of different systems may occur using different criteria. If certification only occurs within a company in an unregulated industry, will the argument really be evaluated in an independent and unbiased way?

Where will people qualified to do this evaluation be found? Such evaluation would be extremely time and resource consuming and potentially costly [Wassyng, 2011]. Companies complain already about the long waits involved in certification. In addition, regulatory agencies are notoriously understaffed and underfunded but the safety case requires a well-resourced and competent regulator.

As an example, consider again offshore oil exploration. The UK and Norway, which use the safety case approach for the North Sea, employ a large number of highly educated personnel and technical specialists to perform audits, inspections and review required documents. In Norway, the regulatory authority has 160 employees, of which approximately 100 perform compliance and audit related tasks regulating 105 offshore installations. Each of these 100 employees has a postgraduate (Masters Degree) or equivalent level of training in one or more areas of expertise, including drilling, petroleum engineering, structural engineering, and reliability engineering. In contrast, the U.S. Bureau of Safety and Environmental Enforcement (BSEE) and the U.S Coast Guard share approximately 60 billeted offshore inspectors for over 3,500 offshore installations [Steinzor, 2010].

Another practical consideration is the role of stakeholders in the certification process. Historically, users of products and systems were expected to evaluate the safety of these products themselves and assume responsibility for using them. But systems have become so complex that this process is no longer possible. I cannot evaluate the safety of an aircraft before I decide to fly on it so I must trust in the regulatory authorities to make this determination. However, historically stakeholders and engineers and scientists outside the company involved have participated in creating the product-based and process-based standards and sometimes even the certification process itself so blind trust in government agencies to ensure my safety is not required. External evaluation of the certification processes by stakeholders such as airline passengers and pilots or those living in the vicinity of nuclear power plants is possible. For example, in the commercial aviation realm, RTCA committees that define certification procedures are open to everyone, including pilots and pilot unions and airline passenger associations. For TCAS, pilots participated directly in the certification process as did anyone in the country with an interest in participating. Stakeholders cannot see proprietary company data, but they can help determine the process for evaluating that data.

With the safety case approach, each case could be presented using different evidence and the company itself determines what argument is provided. Stakeholders are effectively shut out of the certification process.

## An Alternative to Safety Cases in the Certification of Complex Systems

Any viable certification approach will almost certainly contain both types of prescriptive certification. Product-based certification ensures that lessons of the past are considered in new systems and that more general design approaches determined over time to be useful in that industry and for the types of systems being designed and operated are implemented in new designs. Process-based certification ensures consistency of certification decisions, stakeholder inputs and participation, and more efficient and cost-effective implementation of the certification process.

There may be many potentially useful certification approaches that have these characteristics. In addition, there may be instances where goal or performance based certification is appropriate, particularly if there is an accepted and scientifically viable means for evaluating the achievement of the goal.

Some general features of acceptable (at least to the author) certification approaches include the following:

1. Certification should include a plan submitted early to the certifying authority outlining the way the system will be certified to be safe. Such a plan is required in MIL-STD-882. It allows flexibility in the way that the standard is applied. Tailorable standards are also important in providing flexibility. MIL-STD-882, for example, is written as a set of independent tasks, not all of which are appropriate for every project. The certification

authority can approve the company's plan or request modifications. Note that this approach is similar to that used in approval of pharmaceuticals by the FDA. The plan will also include the accidents (losses) and specific hazards to be considered. In some industries, these losses and hazards are actually defined by the government regulator and are not at the discretion of the designer or operator of the system. This plan should include how safety will be designed into the system from the beginning and not just argued in at the end. It should also include a plan for maintaining the certification throughout the life time of the product or system.

2. The certification process should require a hazard analysis and information about how the hazards considered were eliminated or controlled. Ideally, the hazard analysis techniques used should be at the discretion of the applicant in order to allow practices to include new approaches without waiting years for them to be approved through a leaden international committee process. The agency or industry standards can specify the types of results that should be obtained and the factors to be considered (including software, human factors, and operations) but should allow discretion in the way they are obtained. The specific hazard analysis techniques to be used can be included in the certification plan and thus approved or not by the certifier at the beginning of the development process.
3. Deliverables to the agency should include the limitations of what was done and the uncertainties, and assumptions underlying the analyses and design procedures used.

**Conclusions**

To avoid confirmation bias and compliance-only exercises, certification should focus not on showing that the system is safe but in attempting to show that it is unsafe. It is the emphasis and focus on identifying hazards and flaws in the system that provides the "value-added" of system safety engineering. The system engineers have already created arguments for why their design is safe. The effectiveness in finding safety flaws by system safety engineers has usually resulted from the application of an opposite mindset from that of the developers.

Whatever is included in the certification process, the following characteristics seem important:

- The process should be started early. The analysis done for certification is only useful if it can influence design decisions. That means it should not be done after a design is completed or prepared in isolation from the system engineering effort. If safety cases are created only to argue that what already exists is safe, then the effort will not improve safety and becomes, as apparently has happened in the past, simply a paper exercise to get a system certified. One unfortunate result might be unjustified complacency by those operating and using the systems.
- The assumptions underlying the certification decision should be continually monitored during operations and procedures established to accomplish this goal. The system may be working, but not the way it was designed or the assumptions may turn out to be wrong, perhaps because of poor prediction or because the environment has changed. Changes to the system and its environment may have been made for all the right reasons, but the drift between the system as designed and the system as enacted is rarely if ever analyzed or understood as a whole, rather than each particular deviation appearing sensible or even helpful to the individuals involved.
- To make maintaining the certification feasible, the analysis needs to be integrated into system engineering and system documentation so it can be maintained and updated. Certification should not be just a one-time activity but must continue through the lifetime of the system, including checking during operations that the assumptions made in the certification decision remain true for the system components and the system environment. In the author's experience, the major problems in updating and maintaining documentation and certification arise in relating the analysis to the detailed design decisions. When a system design or operating environment is changed, it must be possible to determine what assumptions in the hazard analysis are involved and must be revisited. Starting a new hazard analysis from scratch is just not feasible.

- The analysis should consider worst cases, not just the likely or expected case (the latter is called a *design basis accident* in nuclear power plant regulation).
- The analysis needs to include all factors, that is, it must be comprehensive. It should include not just hardware failures and operator errors but also management structure and decision-making. It must also consider operations and the updating process for the analysis must not be limited to development and certification but must continue through the operational part of the system life cycle.
- To be most useful, qualitative and verifiable quantitative information must be used, not just probabilistic models of the system.
- The integrated system must be considered and not just each hazard or component in isolation.
- The certification process must be practical to implement with the personnel available in the government certification agency (or the company if external certification is not involved) or by those licensed by them to accept responsibility for approval.

**References**

American Nuclear Society (2004), "Risk-Informed and Performance-Based Regulations for Nuclear Power Plants," Position Statement 46, June.

The Hon. Lord Cullen (1990), The Public Inquiry into the Piper Alpha Disaster, Vols. 1 and 2 (Report to Parliament by the Secretary of State for Energy by Command of Her Majesty, November).

Sidney Dekker (2006), *The Field Guide to Understanding Human Error*, Ashgate Publishers.

B. Fischoff, P. Slovic, and S. Lichtenstein (19780, "Fault Trees: Sensitivity of Estimated Failure Probabilities to problem Representation," *Journal of. Experimental Psychology: Human Perception and Performance*, vol. 4.

Charles Haddon-Cave (2009), *The Nimrod Review*, HC 1025, London: The Stationery Office Limited, Oct. 28.

Health and Safety Executive (2005), "Safety Case Regulations for Offshore Oil Drilling."

Kathryn Heiler (2005), "Is the Australian Mining Industry Ready for a Safety Case Regime," *31st International Conference of Safety in Mines Research Institute*, Brisbane, Australia, Oct. 2005.

Oliver A. Houck (2010), "Worst Case and the Deepwater Horizon Blowout: There Ought to be a Law," *Environmental Law Reporter*, 40 ELR 11036, Nov.

J.R. Inge (2007), "The Safety Case: Its Development and Use in the United Kingdom," *Equipment Safety Assurance Symposium 2007*, Bristol U.K.

Kunda, Ziva (1999), *Social Cognition: Making Sense of People*, MIT Press, ISBN 9780262611435, OCLC 40618974.

N.G. Leveson (1995), *Safeware: System Safety and Computers*, Addison Wesley Publishers.

N.G. Leveson (2012), *Engineering a Safer World*, MIT Press.

Nickerson, Raymond S. (1998), "Confirmation Bias; A Ubiquitous Phenomenon in Many Guises", *Review of General Psychology* (Educational Publishing Foundation) 2 (2): 175–220,

NOPSA (2005), http://nopsa.gov.au/safety.asp.

Steven J. Pereira, Grady Lee, and Jeffrey Howard (2006). "A System-Theoretic Hazard Analysis Methodology for a Non-advocate Safety Assessment of the Ballistic Missile Defense System," *AIAA Missile Sciences Conference*, Monterey, CA, Nov.

Rasche, T (2001), "Development of a safety case methodology for the Minerals Industry – a discussion paper," MISHC, University of Queensland.

RTCA (2008), "Safety, Performance and Interoperability Requirements Document for the In-Trail Procedure in the Oceanic Airspace (ATSA-ITP) Application," DO-312.

Rena Steinzor (2010), “Lessons from the North Sea: Should ‘Safety Cases’ Come to America?” *Boston College Symposium on Environmental Affairs Law Review*.

Ian Sutton, "Preparing and Managing a Safety Case in the Process Industries," http://knol.google.com/k/ian-sutton/safety-cases/2vu500dgllb4m/33#.

Vectra Group, "Literature Review on the Perceived Benefits and Disadvantages of the UK Safety Case Regime”, at http://www.hse-databases.co.uk/research/misc/sc402083.pdf.

K.J. Vicente and J. Rasmussen (1992), "Ecological Interface Design: Theoretical Foundations," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 22, no. 4, July/Aug.

Alan Wassyng, Tom Maibaum, Mark Lawford, and Hans Bherer (2011), “Software certification: Is there a case against safety cases?,” in R. Calinescu and E. Jackson (eds.), *Monterey Workshops 2010*, LNCS 6662, Springer-Verlag, pp. 206-227.

D. Whyte (1997), “Moving the goalposts: The deregulation of safety in the post piper alpha offshore oil industry,” http://www.psa.ac.uk/cps/1997/whyt.pdf.

P. Wilkinson (2002), “Safety case: success or failure?” *Seminar paper 2 National Research Centre for OHS Regulation*, ANU Canberra

Wikipedia, Mindset, http://en.wikipedia.org/wiki/Mindset.