Safety Science 49 (2011) 55–64

Contents lists available at ScienceDirect

Safety Science

journal homepage: www.elsevier.com/locate/ssci

Applying systems thinking to analyze and learn from events

Nancy G. Leveson *

Aeronautics and Astronautics, Engineering Systems, MIT, United States

ARTICLE INFO

Article history:
Received 25 August 2009
Received in revised form 20 December 2009
Accepted 30 December 2009

Keywords:
Operator error
Accident model
Safety versus reliability
Hazard analysis
Causal factors

ABSTRACT

Major accidents keep occurring that seem preventable and that have similar systemic causes. Too often, we fail to learn from the past and make inadequate changes in response to losses. Examining the assumptions and paradigms underlying safety engineering may help identify the problem. The assumptions questioned in this paper involve four different areas: definitions of safety and its relationship to reliability, accident causality models, retrospective vs. prospective analysis, and operator error. Alternatives based on systems thinking are proposed.

© 2010 Elsevier Ltd. All rights reserved.

> *It isn't what we don't know that gives us trouble, it's what we know that ain't so.*[1]

1. Introduction

The prospectus for the workshop that the papers in this special issue arise from ponders why event analysis and learning from experience has not been as successful as predicted. We do not seem to be making much progress lately in reducing accidents in most industries. Major accidents keep occurring that seem preventable and that have similar systemic causes. Too often, we fail to learn from the past and make inadequate changes in response to accidents. The prospectus suggests three potential explanations: (1) our analysis methods do not discover the underlying causes of events, or (2) learning from experience does not work as it is supposed to do, or (3) learning is happening in the wrong places. More generally, why do the approaches we use to learn from events, most of which go back for decades and have been incrementally improved over time, not work well in today's world?

Maybe the answer lies in re-examining the assumptions and paradigms underlying safety engineering, most of which go back decades, to identify any potential disconnects with the world as it exists today. While abstractions and simplifications are useful in dealing with complex systems and problems, those that are counter to reality can hinder us from making forward progress. There are too many beliefs in accident analysis—starting from the assumption that analyzing events and learning from them is adequate—that are accepted without question. Most of the new research in this field never questions these assumptions and paradigms. It is time that we began to question them. Rather than propose some incremental improvement in what we do today—a new tool or a new learning protocol—this paper will instead examine the underlying foundations in this field.

The assumptions questioned in this paper involve four different areas: definitions of safety and its relationship to reliability, accident causality models, retrospective vs. prospective analysis, and operator error. Alternatives based on systems thinking and systems theory are proposed.

2. Safety vs. reliability

> *Assumption*: Safety is increased by increasing the reliability of the individual system components. If components do not fail, then accidents will not occur.

Safety and reliability are *different* system properties. One does not imply nor require the other—a system can be reliable and unsafe or safe and unreliable. In some cases, these two system properties are conflicting, i.e., making the system safer may decrease reliability and enhancing reliability may decrease safety. The confusion on this point is exemplified by some researchers who suggest that high *reliability* organizations will be safe (La Porte and Todd, 1996; Roberts, 1990; Rochlin et al., 1987; Weick et al., 1999) and in the focus on failure events in most accident and incident analysis. This belief is simply not true. In complex systems, accidents often result from interaction among perfectly

* Tel.: +1 617 258 0505; fax: +1 617 253 7397.
E-mail address: leveson@mit.edu.

[1] Attributed to Will Rogers (e.g., New York Times, 10/7/84, p. B4), Mark Twain, and Josh Billings (Oxford Dictionary of Quotations, 1979, p. 491) among others.

0925-7535/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ssci.2009.12.021

functioning components. The loss of the Mars Polar Lander was attributed to noise (spurious signals) generated when the landing legs were deployed during descent (JPL, 2000). This noise was normal and expected and did not represent a failure in the landing leg system. The onboard software interpreted these signals as an indication that landing occurred (which the software engineers were told they would indicate) and shut the engines down prematurely, causing the spacecraft to crash into the Mars surface. The landing legs and the software performed correctly (as specified in their requirements, i.e., neither failed), but the accident occurred because the system designers did not account for all interactions between the leg deployment and the descent-engine control software.

This type of component interaction accident is becoming more common as the complexity of our system designs increases. In the past, our designs were more intellectually manageable and the potential interactions among components could be thoroughly planned, understood, anticipated, and guarded against. In addition, thorough testing was possible and could be used to eliminate system design errors before system use. Modern, high-tech systems no longer satisfy these properties and system design errors are increasingly the cause of major accidents, even when all the components have operated reliably, i.e., have not failed. As Perrow has noted (Perrow and Charles, 1999), such systems will also be harder for operators to manage in a crisis situation.

The same applies to organizational decision-making as illustrated by Rasmussen's analysis of the Zeebrugge ferry mishap (Rasmussen, 1997) shown in Fig. 1. Some information about the accident (Sheen and Barry, 1987) is necessary to understand the figure. On the day the ferry capsized, the *Herald of Free Enterprise* was working the route between Dover and the Belgium port of Bruges-Zeebrugge. This was not her normal route and the linkspan at Zeebrugge had not been designed specifically for the Spirit class of vessels. The linkspan used spanned a single deck and so could not be used to load decks E and G simultaneously. The ramp could also not be raised high enough to meet the level of deck E due to the high spring tides being encountered at that time. This limitation was commonly known and was overcome by filling the forward ballast tanks to lower the ferry's bow in the water. The *Herald* was due to be modified during its refit in 1987 to overcome this problem.

Before dropping moorings, it was normal practice for a member of the crew, the Assistant Bosun, to close the doors. The First Officer also remained on deck to ensure they were closed before returning to the wheel house. To keep on schedule, the First Officer returned to the wheel house before the ship dropped its moorings (which was common practice), leaving the closing of the doors to the Assistant Bosun, who had taken a short break after cleaning the car deck upon arrival at Zeebrugge. He had returned to his cabin and was still asleep when the ship left the dock. The captain could only assume that the doors had been closed because he could not see them from the wheel house due to their construction, and there were no indicator lights in the wheelhouse to show door position. There was confusion as to why no one else closed the doors. A few years earlier, one of the Herald's sister ships sailed from Dover to Zeebrugge with the bow doors open, but she made it to the destination without incident. It was therefore believed that leaving the bow doors open should not alone have caused the ship to capsize.

Another factor that contributed to the capsizing was the depth of the water: if the ship's speed had been below 18 knots (33 km/h) and the ship had not been in shallow water, people on the car deck would probably have had time to notice the bow doors were open and close them. But open bow doors were not alone enough to cause the final capsizing. Almost all ships are divided into watertight compartments below the water line so that in the event of flooding, the water will be confined to one compartment, keeping the ship afloat. The Herald's design had an open car deck with no dividers, allowing vehicles to drive in and out easily, but this design allowed water to flood the car deck. As the ferry turned, the water on the car deck moved to one side and the vessel capsized. One hundred and ninety three passengers and crew were killed.

In this accident, those making decisions about vessel design, harbor design, cargo management, passenger management, traffic scheduling, and vessel operation were unaware of the impact of their decisions on the others and the overall impact on the process leading to the ferry accident. The type of bottom-up decentralized

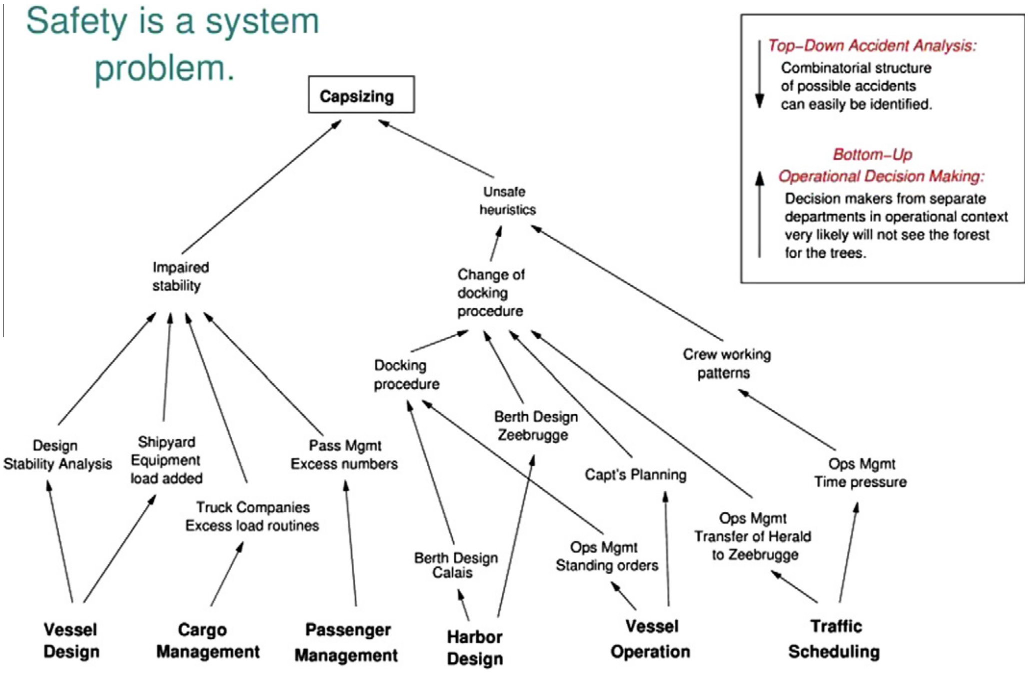
Safety is a system problem.
Capsizing
Impaired stability
Unsafe heuristics
Change of docking procedure
Docking procedure
Crew working patterns
Berth Design Zeebrugge
Capt's Planning
Ops Mgmt Time pressure
Design Stability Analysis
Shipyard Equipment load added
Pass Mgmt Excess numbers
Truck Companies Excess load routines
Ops Mgmt Transfer of Herald to Zeebrugge
Berth Design Calais
Ops Mgmt Standing orders
Vessel Design
Cargo Management
Passenger Management
Harbor Design
Vessel Operation
Traffic Scheduling
Top–Down Accident Analysis:
Combinatorial structure of possible accidents can easily be identified.
Bottom–Up Operational Decision Making:
Decision makers from separate departments in operational context very likely will not see the forest for the trees.

Fig. 1. The complex interactions in the Zeebrugge ferry accident (adapted from Rasmussen, (1997)).

decision-making often advocated for HROs (Weick and Karl, 1987) can lead (and has led) to major accidents in complex socio-technical systems. Each local decision may be "correct" (and "reliable," whatever that might mean in the context of decisions) within the limited context within which it was made but lead to an accident when the independent decisions and organizational behaviors interact in dysfunctional ways. As the interactive complexity grows in the systems we build, accidents caused by dysfunctional interactions among components become more likely. Safety is a system property, not a component property, and must be controlled at the system level not the component level.

The example above is of a system where the components are reliable, but the system is unsafe. The opposite can also be true. As an example of behavior that is unreliable but safe, consider human operators. If a human operator does not follow the specified procedures, then they are not operating reliably. In some cases that can lead to an accident. In other cases, it may prevent an accident when the specified procedures turn out to be unsafe under the particular circumstances. Examples abound of operators ignoring prescribed procedures in order to prevent an accident (Leveson, 1995; Perrow and Charles, 1999). At the same time, accidents have resulted precisely because the operators *did* follow the predetermined instructions provided to them in their training, such as at Three Mile Island. When the results of deviating from procedures are positive, operators are lauded but when the results are negative, they are punished for being unreliable. In the successful case (deviating from specified procedures averts an accident), their behavior is unreliable but safe. It satisfies the emergent safety constraints for the system, but not individual reliability requirements with respect to following specified procedures.

Some of the confusion between reliability and safety stems from the use of vague and ambiguous definitions. So let's start there. A *failure* occurs when a component does not satisfy its specified requirements (Leveson, 1995), i.e., a failure is the nonperformance or inability of the system or component to perform its intended function for a specified time under specified environmental conditions. If there are no requirements either specified or assumed, then there can be no failure as any behavior is acceptable. If a component behavior is undesirable from the encompassing system standpoint, then there is a design error in the system, but the component has not failed (it operated as intended by the designers and builders of the component). In a different system, the same component behavior might not have had adverse effects but in either case the component behaved identically and thus did not fail in one case and not the other. The component operated as designed and intended. The problem occurred in the system design as a whole, i.e., in the interaction among the components. A system design error can lead to an accident (unacceptable loss) without any component failures.

Reliability in engineering is defined as the probability that a component satisfies its specified behavioral requirements over time and under given conditions, i.e., it does not fail (Leveson, 1995). Every physical component (and most humans) can be made to "break" given some set of conditions or a long enough time. The limitations in time and operating conditions are required to differentiate between unreliability under assumed operating conditions and situations where no component or component design could have continued to operate. If a driver puts on the brakes too late to avoid hitting the car in front, we would not say that the brakes "failed" because they did not stop the car under circumstances for which they were not designed. The brakes, in this case, were not unreliable. They operated reliably, but the requirements for system safety (in this case) went beyond the capabilities of the brake design. Failure and reliability are always related to component requirements and assumed operating (environmental) conditions.

Safety can be defined as the absence of accidents, where an *accident* is defined as an event involving an unplanned and unacceptable loss[2] (Leveson, 1995). *Safety*, unlike reliability, is a system property, not a component property (Leveson, 2009). Determining whether a nuclear power plant is acceptably safe, for example, is not possible by examining a single valve in the plant, and evaluating the safety of a hospital clinical care unit is not possible by examining a single step in a surgical procedure. In fact, statements about the "safety of the valve" without information about the context in which that valve is used are meaningless. Conclusions can be reached about the reliability of the valve. But safety can only be determined by the relationship between the valve and the other plant components, that is, in the context of the whole. Similarly, statements about the "safety of a total hip replacement operation" are meaningless without information about the context in which the procedure is performed—such as the quality and compliance with pre-operative preparation, the urgency with which it is performed, constraints imposed by the physical facilities (e.g., the availability of laminar flow rooms in the operating theater), and competition for shared resources during the technical performance of the procedure (e.g., high demand for intra-operative X-rays due to concurrent trauma cases), access to post-operative rehabilitation, etc. A component (or procedure) that is perfectly safe in one system may not be when used in another. A hair dryer may be highly reliable and safe when used under most circumstances. When seated in a bathtub, however, it is still very reliable (the reliability has not changed) but it has become unsafe.

In systems theory, complex systems are modeled as a hierarchy of organizational levels, each level more complex than the one below (Ramo and Simon, 1973). The levels are characterized by *emergent* properties that are irreducible and represent constraints on the degree of freedom of components at the level below. Safety is an emergent property and unsafe system behavior is defined in terms of *safety constraints*[3] on the behavior of the system components. Safety is then viewed, using systems thinking and systems theory, as a *control* problem (enforcing the safety constraints) rather than a *failure* or reliability problem. Between each level of the hierarchy, a feedback control loop acts to ensure that the safety constraints are enforced.

Fig. 2 shows an example of a hierarchical safety control structure for a typical US. regulated industry, such as aircraft. Each industry and company will, of course, have its own unique control structure. There are two basic hierarchical control structures in Fig. 2—one for system development (on the left) and one for system operation (on the right)—with interactions between them. An aircraft manufacturer, for example, might only have system development under its immediate control, but safety involves both development and operational use of the aircraft and neither can be accomplished successfully in isolation: Safety must be designed into the aircraft and safety during operations depends partly on the original design and partly on effective control over operations. Manufacturers must communicate to their customers the assumptions about the operational environment in which the original safety analysis was based, e.g., maintenance quality and procedures, as well as information about safe aircraft operating procedures. The operational environment, in turn, provides feedback to the manufacturer about the performance of the system during operations. Each component in the hierarchical safety control structure has responsibilities for enforcing safety constraints appropriate for and assigned to that component; together these responsibilities should result in enforcement of the overall system safety constraints.

[2] This definition is the one used in the US defense and aerospace communities, for example see MIL-STD-882 and the NASA safety standards.

[3] In systems theory, constraints are limitations on the behavioral degree of freedom of the system components.

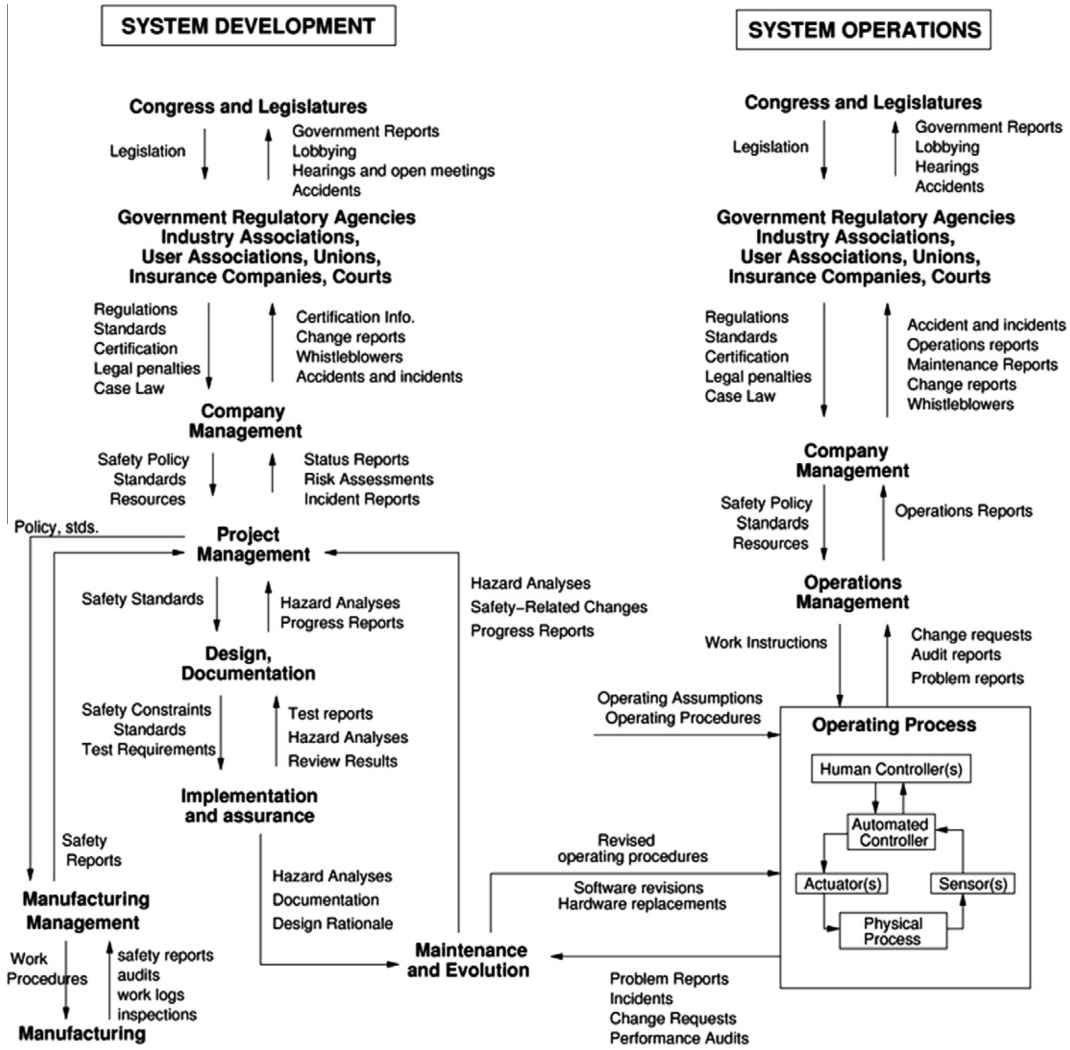

Fig. 2. An example of a hierarchical safety control structure.

Hierarchies, in system theory, are characterized by control and communication processes operating at the interfaces between levels (Checkland and Peter, 1981). The downward communication channel between levels in the hierarchy provides information necessary to impose behavioral constraints on the level below and an upward feedback channel provides information about how effectively the constraints were enforced. For example, in Fig. 2, company management in the development safety control structure (on the left in the figure) may provide safety policy, standards and resources to project management and, in return, receive status reports, risk assessment, and incident reports as feedback about the status of the project with respect to the safety constraints.

Not only are reliability and safety different qualities, but they often conflict: Increasing reliability may decrease safety and increasing safety may decrease reliability. One of the challenges of engineering is to find ways to increase safety without decreasing reliability. Understanding the conflicts between reliability and safety requires distinguishing between requirements and constraints. *Requirements* represent the mission or reason for existence of the system. *Constraints* represent acceptable ways the system can achieve those goals. While in some systems, safety is part of the mission or reason for existence, e.g., air traffic control and healthcare, in others safety is not the mission but a constraint on how the mission can be achieved. For example, the mission of a chemical manufacturing plant is to produce chemicals. Not exposing bystanders to toxins or not polluting the environment are constraints on the way the mission (producing chemicals) can be achieved. The best way to achieve the safety and environmental constraints is not to build or operate the system at all. A particular plant may very reliably produce chemicals while occasionally releasing toxic materials into the surrounding environment. The plant is reliable but unsafe. Sometimes the safety requirements conflict among themselves. For example, one safety constraint on an automated train door system is that the doors must not open unless the train is stopped and properly aligned with a station platform. Another safety constraint is that the doors must open anywhere for emergency evacuation. There are always multiple goals and constraints for any system—the trick in engineering and risk management decision-making is to analyze and implement tradeoffs among these multiple requirements and constraints.

Analyzing and learning from accidents requires going beyond focusing on component failure and reliability. High component reliability is neither necessary nor sufficient for system safety. Safety is an emergent property that exists only at the system level, not the component level. Top-down analysis and control is necessary to handle safety.

3. Retrospective vs. prospective analysis

Assumption: Retrospective analysis of adverse events is required and perhaps the best way to improve safety.

Retrospective event analysis is necessarily limited because systems are rarely if ever static. As Rasmussen has argued, systems and organizations continually experience change as adaptations are made in response to local pressures and short-term productivity and cost goals (Rasmussen, 1997). People adapt to their environment or they change their environment to better suit their purposes. A corollary of this propensity for systems and people to adapt over time is that safety defenses are likely to degenerate systematically through time, particularly when pressure toward cost-effectiveness and increased productivity is the dominant element in decision-making. The critical factor here, as noted by Rasmussen, is that such adaptation is not a random process—it is an optimization process depending on search strategies—and thus should be predictable and potentially controllable.

Reliance on accident and incident analysis has been most effective in industries like nuclear power and commercial aircraft design where the basic designs change very slowly (if at all) and common designs are used by lots of people (e.g., all US nuclear reactors are based on the original reactor designed for nuclear submarines). In these cases, reactive learning from event analysis has been very successful—accidents and incidents are analyzed in great depth and the causes are eliminated from the basic designs. In other industries, design changes are always occurring and simply looking at past causal factors is not enough. And even in the commercial aircraft and nuclear power industries, technical innovation and the introduction of digital components (e.g., computers and software) is making reliance on past experience less effective. Simply looking at past events in such systems will be as ineffective as attempting to catch a moving train by going to where it was previously.

In fact, reliance on retrospective analysis of events has contributed to losses, for example, the loss of a satellite where quality assurance only checked for those things that had led to a satellite loss in the past (Pavlovich, 1999). In this case, the proximate cause was a typo in the software data used for the launch. Such a typo had never occurred before so the correctness of the software load tape was never checked. With software, so many potential errors can be made that simply looking at past errors is hopeless in preventing software-related accidents. In addition, software that is used many times safely in one system can cause an accident when used in another system or when environmental factors change over time. Reuse of software is common along with the assumption that it will be safe because no accidents occurred in the previous system.

The introduction of new technology, particularly digital technology, has led to an increase in the complexity of the systems we are building and the introduction of new potential causal factors. This does not mean that retrospective analysis cannot be useful, especially if systemic factors are identified in the analysis and not just symptoms or the proximate events and their immediate causes: Systemic factors tend to change much more slowly than physical design. But it does imply that proactive analysis and control of the system hazards is becoming more important and will become increasingly important in the future. In addition, of course, the consequences of an accident in some of our new systems are so great that we cannot afford to wait until accidents occur to determine how to prevent them. Hazard analysis, which has been used for very dangerous systems for 50 years, can identify the causes of accidents that have never occurred previously so they can be prevented from occurring the first time. It does not start by looking only at known failure modes or interactions among system components but instead starts by identifying all potentially hazardous states and conditions and then determining whether they are possible. If the consequences are serious enough, hazards are eliminated or controlled even if we cannot determine how or if they might occur.

4. Accident causation models

Assumption: Accidents are caused by chains of directly related failure events.

This assumption implies that working backward from the loss event and identifying directly related predecessor events (usually technical failures or human errors) will identify the "root cause" for the loss. Using this information, either the "root cause" event is eliminated or an attempt is made to stop the propagation of events by adding barriers between events, by preventing individual failure events in the chain, or by redesigning the system so that multiple failures are required before propagation can occur (putting "*and*" gates into the event chain). Fig. 3 shows a typical chain-of-events model for a tank rupture. The events are annotated in the figure with standard engineering "fixes" to eliminate the event.

The problem with the chain-of-events[4] model of accident causation is that it oversimplifies causality and the accident process and excludes many of the systemic factors in accidents and indirect or non-linear interactions among events. It also does not hold for accidents where the cause(s) lies in the interaction among system components, none of which may have failed.

Let's take the sinking of the Herald of Freedom as an example again. Working back through the chain-of-events, it would appear that the root cause was the Assistant Bosun not closing the doors and the First Officer not remaining on deck to check the doors before returning to the wheel house. There was a redundant design here (with the First Officer checking the work of the Assistant Bosun), but it did not prevent the accident, as redundancy often does not (Leveson, 1995; Perrow and Charles, 1999). Many of the important factors in the Herald of Freedom accident description above do not appear in an event chain, either because the relationships are indirect or they involve things that did not happen (which are not events).

In general, accident causation can be viewed as involving three levels (see Fig. 4) (Leveson, 1995). The lowest level is the basic proximate event chain, which includes the failure to close the doors by the Assistant Bosun and the premature return of the First Officer to the wheelhouse. The second level represents the conditions that allowed the events to occur, i.e., the high spring tides, the inadequate design of the ferry loading ramp for this harbor, and the desire of the First Officer to stay on schedule (thus leaving the car deck before the doors were closed). The top level contains the systemic factors that contribute to the conditions and events, such as the owner of the ferry (Townsend Thoresen) needing ships that were designed to permit fast loading and unloading and quick acceleration in order to remain competitive in the ferry business. Another possible systemic factor might have been pressure by the company management on the Captain and First Officer to strictly adhere to schedules.

Most accident analysis techniques identify the proximate chain-of-events and often the conditions underlying those events. They are based on the classic assumption that cause and effect must be directly related. Almost none include systemic factors, often because those factors only have an indirect relationship to the events and conditions. A few attempt to include systemic factors but are severely limited in their success in achieving this goal. To see the limitations of these techniques, consider *5 Whys*. This technique is perhaps the most simplistic and it leads to the least amount of learning from events, but it provides an illustrative example of the problems in most current causal analysis techniques.

Using 5 Whys, the investigation team questions "why" the incident happened or "why" the unfavorable conditions existed.

[4] Note that fault trees are simply a notational convenience for specifying multiple chains (with some common parts) and thus the same arguments apply.

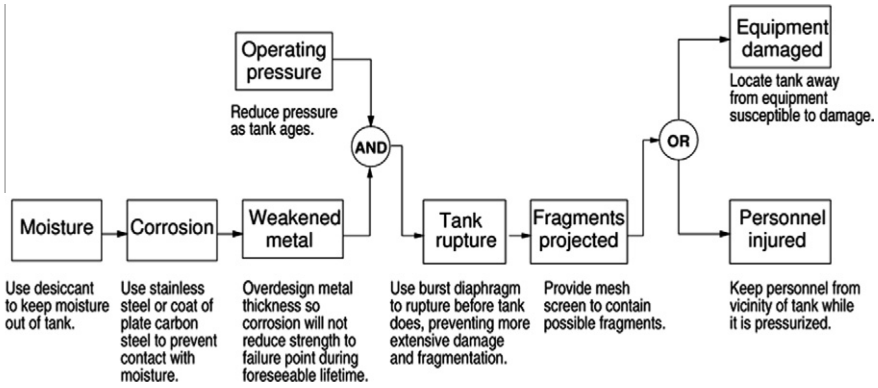
Operating pressure
Reduce pressure as tank ages.
AND
Equipment damaged
Locate tank away from equipment susceptible to damage.
OR
Moisture
Corrosion
Weakened metal
Tank rupture
Fragments projected
Personnel injured
Use desiccant to keep moisture out of tank.
Use stainless steel or coat of plate carbon steel to prevent contact with moisture.
Overdesign metal thickness so corrosion will not reduce strength to failure point during foreseeable lifetime.
Use burst diaphragm to rupture before tank does, preventing more extensive damage and fragmentation.
Provide mesh screen to contain possible fragments.
Keep personnel from vicinity of tank while it is pressurized.

Fig. 3. An example chain-of-events model for a tank rupture from (Leveson, 2009).

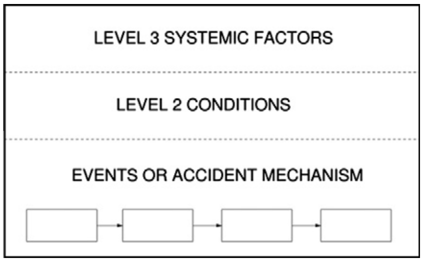

Fig. 4. A hierarchical view of accident causation (Leveson, 1995).

Specifically, the team selects an event associated with the incident, asks why this event occurred, and identifies multiple sub-events or conditions that gave rise to the event. For each of these sub-events or conditions, the team again asks why it occurred. The team records the sub-events or conditions as an event tree. The team then repeats this process five times to identify the root cause. The number five is arbitrary, of course, but the same limitations occur if the question is asked more times. Some problems with this approach include the fact that the results are not reproducible and consistent among investigation teams, systemic causes may not be identified, and a single cause is assumed along with one linear path to an event. Most important, the impacts of the potential ways to eliminate this cause are not evaluated.

Consider the following widely used example (iSixSigma, 2008).

The Washington Monument is disintegrating.

Question (Why?)	Answer
Why is it disintegrating?	Because we use harsh chemicals
Why do we use harsh chemicals?	To clean pigeon droppings off the monument
Why are there so many pigeons?	They eat spiders and there are a lot of spiders at monument
Why are there so many spiders?	They eat gnats and lots of gnats at monument
Why so many gnats?	They are attracted to the lights at dusk

Solution: turn on the lights at a later time.

The identified root cause is the lights attracting gnats to the monument at dusk. The elimination of this root cause then points the investigators to recommendations involving gnats and turning on lights. There may be (and probably are), however, several reasons for the disintegration and not simply the harsh chemicals used (e.g., pollution and perhaps the original building materials). The 5 Whys approach often will not identify them because it relies on only one analytical chain of reasoning. Other techniques, such as fault trees and their equivalent (e.g., fishbone diagrams and why-because diagrams, which are simply a different notation for the classic fault tree), find more paths but suffer from the same potential omissions.

Note that the solution "turn on the lights at a later time" is not the only solution—less harsh chemicals might be substituted, for example, or it might be possible to treat the surface of the monument in a way that provides protection. More important, the solution chosen may cause other more serious problems, e.g., tourists may be more likely to be mugged, but this unintended consequence is not considered because the problem has been analyzed so superficially.

In general, all these techniques are limited in that they are unable to go beyond the investigator's current knowledge, i.e., they cannot find causes that they do not already know, they lack support to help the investigator ask the right "why" questions, and the results are not repeatable. Different people using them will come up with different causes for the same problem. A more powerful approach is illustrated below.

Most current accident models and accident analysis techniques suffer from the limitation of considering only the events underlying an accident and not the entire accident *process*. Accidents are often viewed as some unfortunate coincidence of factors that come together at one particular point in time and lead to the loss. This belief arises from too narrow a view of the causal time line. As argued above, systems are not static. Rather than accidents being a chance occurrence of multiple independent events, they tend to involve a migration to a state of increasing risk over time. A point is reached where an accident is inevitable (unless the high risk is detected and reduced) and the particular events involved are somewhat irrelevant: if those events had not occurred, something else would have led to the loss. This concept is reflected in the common observation that a loss was "an accident waiting to happen." The proximate cause of the Challenger Space Shuttle was the foam coming loose from the external tank and damaging the re-entry heat control structure. But many potential problems that could have caused the loss of the Shuttle had preceded this event and an accident was avoided by luck or unusual circumstances. The economic and political pressures had led the Shuttle program to drift to a state where any slight deviation could have led to a loss (Leveson, 2007).

Understanding and preventing or detecting system migration to states of higher risk requires that our accident models consider the *processes* involved in accidents and not simply the events and conditions: Processes control a sequence of events and describe system and human behavior as it changes and adapts over time (perhaps as a result of feedback or a changing environment) rather than considering individual events and human actions. Accident causation is a complex process involving the entire socio-technical system including legislators, government regulatory agencies, industry associations and insurance companies, company management, technical and engineering personnel, operators, etc. To understand why an accident has occurred, the entire process needs to be examined, not just the proximate events in the event chain. Otherwise, only symptoms will be identified and fixed, and accidents will continue to recur.

Better accident analysis techniques that avoid the limitations of those based on chain-of-events models are possible by instead using systems thinking and systems theory. Systems theory dates from the thirties and forties and was a response to the limitations of the classic analysis techniques in coping with the increasingly complex systems being built (Checkland and Peter, 1981). In the traditional scientific method, sometimes referred to as *divide and conquer*, systems are broken into distinct parts so that the parts can be examined separately: physical aspects are decomposed into separate physical components while behavior is decomposed into events over time. Such decomposition (formally called *analytic reduction*) assumes that such separation is feasible: that is, each component or subsystem operates independently and analysis results are not distorted when these components are considered separately. This assumption in turn implies that the components or events are not subject to feedback loops and other non-linear interactions and that the behavior of the components is the same when examined singly as when they are playing their part in the whole. A third fundamental assumption is that the principles governing the assembly of the components into the whole are straightforward, that is, the interactions among the subsystems are simple enough that they can be considered separate from the behavior of the subsystems themselves. These are reasonable assumptions, it turns out, for many of the physical regularities of the Universe. They do not hold, however, for the complex, software-intensive systems we are now engineering (Leveson, 2009) nor do they hold for social systems (Checkland and Peter, 1981).

Instead of decomposing behavior into events over time, the systems approach focuses on systems taken as a whole. It assumes that some system properties can only be treated adequately in their entirety, taking into account all facets relating the social to the technical aspects (Ramo and Simon, 1973). These system properties derive from the relationships among the parts of the system: how the parts interact and fit together (Ackoff, 1971). Thus, the system approach concentrates on the analysis and design of the whole as distinct from the components or the parts and provides a means for studying emergent system properties, such as safety (Leveson, 2009). Using this approach as a foundation, new types of accident analysis (both retroactive and proactive) can be devised that go beyond simply looking at events and can identify the processes and systemic factors behind the losses and also the factors (reasons) for migration toward states of increasing risk. This information can be used to design controls that prevent hazardous states by changing the design to prevent or control the hazards and migration and, in operational systems, detect the increasing risk before a loss occurs. Leveson et al. provides an example of how this analysis can be accomplished for a complex, socio-technical system, in this case, Space Shuttle operations (Leveson et al., 2005).

5. Operator error

> *Assumption*: (1) Most accidents are caused by operator error and (2) rewarding "correct" behavior and punishing "incorrect" behavior will eliminate or reduce accidents significantly.

These assumptions underlie the common behavioral approach to occupational safety. The problem here is the fact that human behavior is always influenced by the environment in which it takes place. Changing that environment will be much more effective in reducing operator error than reward and punishment. Without changing the environment, human error cannot be reduced for long. We design systems in which human error is inevitable and then blame the human and not the system design.

Part of the problem stems from the chain-of-events approach to accident investigation where it is usually difficult to find an "event" preceding and causal to the operator behavior. If the problem is the system design, there is no proximal event to explain the error (only a decision during system design). Even if a technical failure precedes the human action, the tendency is to put the blame on an inadequate response to the failure by an operator. Perrow claims that even in the best of industries, there is rampant attribution of accidents to operator error, to the neglect of errors by designers or managers (Perrow and Charles, 1999). He cites a US Air Force study of aviation accidents that concludes that the designation of human error, or pilot error, is a convenient classification for mishaps whose real cause is uncertain, complex, or embarrassing to the organization.

Traditional event-based accident and risk models are particularly poor at dealing with human error and decision-making. Human error is usually defined as any deviation from the performance of a specified or prescribed sequence of actions. However, instructions and written procedures are almost never followed exactly, as operators strive to become more efficient and productive and to deal with time and other pressures (Rasmussen, 1997). In studies of operators, even in such highly constrained and high-risk environments as nuclear power plants, modification of instructions is repeatedly found and the violation of rules appears to be quite rational, given the actual workload and timing constraints under which the operators must do their job (Fujita, 1991; Vicente and Kim, 1995; Woods and David, 1984). In these situations, a basic conflict exists between error viewed as a deviation from *normative procedure* versus error viewed as a deviation from the rational and normally used *effective procedure* (Rasmussen et al., 1994).

Mental models play a significant role here. The ability to adapt mental models through experience in interacting with the operating system is what makes the human operator so valuable (see Fig. 5). Designers deal with systems as ideals or averages, and they provide procedures to operators with respect to this ideal. Systems may deviate from the ideal through manufacturing and construction variances or through evolution over time. Operators must deal with the existing system and change their mental models (and operational procedures) using operational experience and experimentation. While procedures may be updated over time, there is necessarily a time lag in this updating process and operators must deal with the existing system state. Based on current information, the operators actual behavior may differ from the prescribed procedures. When the deviation is correct (the designers models are less accurate than the operators' models at that particular instant in time), then the operators are considered to be doing their job. When the operators' models are incorrect, they are often blamed for any unfortunate results, even though their incorrect actions may have been reasonable given the information they had at the time.

Flawed decisions may also result from limitations in the boundaries of the model used, but the boundaries relevant to a particular decision maker may depend on activities of several other decision makers found within the total system (Rasmussen, 1997). Accidents may then result from the interaction of the potential side effects of the performance of the decision makers during their normal work. It is difficult if not impossible for any individual to judge the safety of their decisions when it is dependent on the decisions made by other people in other departments and organizations. Fig. 1 illustrates this problem.

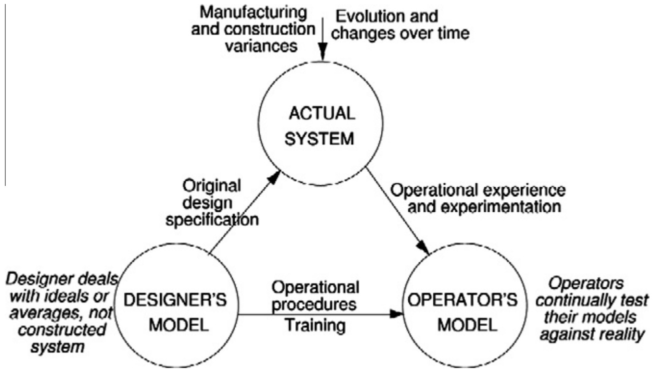
Manufacturing and construction variances
Evolution and changes over time
ACTUAL SYSTEM
Original design specification
Operational experience and experimentation
Designer deals with ideals or averages, not constructed system
DESIGNER'S MODEL
Operational procedures
Training
OPERATOR'S MODEL
Operators continually test their models against reality

Fig. 5. The role of mental models in operations (taken from (Leveson, 2009)).

Traditional decision theory research perceives decisions as discrete processes that can be separated from the context and studied as an isolated phenomenon. Edwards, back in 1962, was one of the first to argue that decisions can only be understood as part of an ongoing process (Edwards, 1962). The state of the system is perceived in terms of possible actions, one of these actions is chosen, and the resulting response from the controlled system acts as a background for the next action. Errors then are difficult to localize in the stream of behavior: *the effects of less successful actions are a natural part of the search on the part of the operator for optimal performance*. Not only are separate decisions difficult to identify in this model of human control, but the study of decision-making then cannot be separated from a simultaneous study of the social context, the value system in which it takes place, and the dynamic work process it is intended to control (Rasmussen, 1990). This view is the foundation of *dynamic decision-making* (Brehmer, 1992) and the new field of *naturalistic decision-making* (Klein et al., 1993; Zsambok et al., 1997).

As argued by Rasmussen and others, devising more effective accident causality models requires shifting the emphasis in explaining the role of humans in accidents from error (deviations from normative procedures) to focus on the mechanisms and factors that shape human behavior, i.e., the performance-shaping mechanisms and context in which human actions take place and decisions are made. Modeling behavior by decomposing it into decisions and actions (i.e., events) and studying it as a phenomenon isolated from the context in which the behavior takes place is not an effective way to understand behavior (Rasmussen, 1997).

To completely understand the cause of accidents and to prevent future ones, the system's hierarchical safety control structure must be examined to determine why the controls at each level were inadequate to maintain the constraints on safe behavior at the level below and why the events occurred. The goal is not to assign blame—blame is the enemy of safety[5]—but to determine why well-meaning people acted in ways that contributed to the loss.

To get a deep enough understanding of the causal factors in an accident such as the Herald of Freedom loss, the reasons for the events and the conditions leading to those events as well as systemic causes need to be identified. Accomplishing this goal requires documenting the hierarchical safety control structure (see Fig. 2 for an example), if such documentation does not already exist, and using that structure to identify and understand the safety control inadequacies in the engineered system (the physical system), the aspects of the design and the environment that affected the loss, and the systemic factors that contributed to the loss.

The first step in the accident analysis is to understand the physical factors involved in the loss, including the limitation of the physical system design (e.g., the Herald of Freedom had a loading ramp that was too low to reach the upper car deck at high tide), the failures and dysfunctional interactions among the physical system components (the Assistant Bosun not closing the doors), and the environmental factors (e.g., the high spring tides) that interacted with the physical system design. Most accident analyses include this information, although they may omit dysfunctional interactions and look only for component failures.

Understanding the physical factors leading to the loss is only the first step, however, in understanding why the accident occurred. The next step is understanding how the engineering design practices contributed to the accident and how they could be changed to prevent such an accident in the future. Why was the hazard (capsizing as a result of flooding) not adequately controlled in the design? Some controls were installed to prevent this hazard (for example, the doors themselves and the assignment to close them to the Assistant Bosun), but some controls were inadequate or missing (a lack of watertight compartments). What parts of the design and analysis process allowed this flawed design to be accepted? What changes in that process, e.g., better hazard analysis, design, or review processes, could be used to ensure that designs have adequate hazard controls in the future?

Many of the reasons underlying poor design and operational practices stem from management and oversight inadequacies due to conflicting requirements and pressures. Identifying the factors lying behind the physical design starts with identifying the safety-related responsibilities (requirements) assigned to each component in the hierarchical safety control structure along with their safety constraints. As an example, a responsibility of the First Officer on the Herald of Freedom is to ensure that the doors are closed before the ferry leaves the dock, management has the responsibility to ensure their ferries have a safe design and are operated safely, and the responsibility of the International Maritime Organization is to provide regulations and oversight to ensure that unsafe ships are not used for passenger transportation, etc. Using these safety-related responsibilities, the inadequate control actions for each of the components in the control structure can be identified. In most major accidents, there is inadequate control exhibited throughout the structure, assuming an adequate control structure was designed to begin with. But simply finding out how each person or group contributed to the loss is only the start of the process necessary to learn what needs to be changed to prevent future accidents. We must first understand *why* the "controllers" provided inadequate control. The analysis process must identify the systemic factors in the accident causation, not just the symptoms.

To understand why people behave the way they do, we must examine their mental models and the environmental factors affecting their decision-making. All human decision-making is based on the person's mental model of the state and operation of the system being controlled (see Fig. 5). For example, the First Officer assumed that the Assistant Bosun had closed the doors, the Assistant Bosun may have thought that someone else would notice that the doors were open and close them, and the Captain thought the doors had been closed.

Preventing inadequate control actions in the future requires not only identifying the flaws in the controllers' mental models (including those of the management and government components of the hierarchical safety control structure) but also why these flaws existed. For example, the Captain's inadequate mental model (thinking the doors were closed) was not corrected in time to prevent the accident due to lack of feedback about the state of the doors. All of them thought that leaving the doors open would not cause a loss of the ferry because a year earlier one of the Herald's

[5] In the Herald of Freedom loss, for example, many of the individuals at Townsend Thoresen were prosecuted for manslaughter, as was the operating company. Such reactions do not increase safety. See recent work on *Just Culture*, e.g., Dekker, 2007.

sister ships sailed from Dover to Zeebrugge with bow doors open without incident, i.e., they had inadequate knowledge about the potential ferry hazards.

The impact of the operating environment (including environmental conditions, cultural values, etc.) must also be identified. For example, the problematic ferry design features were influenced by the competitive ferry environment in which the ferry was to operate. The accident report blamed a ''disease of sloppiness and negligence at every level of the corporation's hierarchy'' (Sheen and Barry, 1987). But this superficial level of analysis (management sloppiness and negligence) is not useful in preventing future accidents—it simply provides someone to blame and to prosecute. It does not eliminate the underlying pressures that led to the poor decision-making nor the inadequate design of the hierarchical safety control structure. Without changes that respond to those factors, similarly flawed and risky decision-making is likely again in the future, although the actual accident details may be very different. We have used system dynamic models to understand the complex environmental, social, and economic factors contributing to poor decision-making in order to provide policy and other changes to improve risk-related decision-making in the future (Dulac et al., 2007; Leveson et al., 2005).

A complete accident/incident analysis based on systems theory usually finds dozens of causal factors contributing to the accident process and points to many changes that could prevent future losses. Leveson provides several examples of such analyses of major accidents (Leveson, 2009). Both retrospective analysis and prospective analysis (which is simply the investigation of an accident that has not yet occurred) would be much more effective if more sophisticated models of accident causation and human behavior were used.

6. Conclusions

The question tackled by these set of papers is why accidents, with the same or similar causes, keep happening. Why do we not seem able to learn from events? Let's examine the three questions posed in the workshop prospectus in the light of the discussion above.

- Could it be that our analysis methods do not discover the underlying causes of the events?
- Does learning from experience not work as it is supposed to do?
- Is learning happening in the wrong places?

The answer to all of these questions, as argued above, is yes. Our current methods and attempts to learn from events are limited because they are based on the assumptions questioned in this paper. As long as we continue to base our accident analysis and learning from events on assumptions that no longer hold for today's systems, we should not be surprised that our accident analysis and prevention efforts are of limited usefulness.

What is needed instead? An argument has been presented that sophisticated models of causality (not more notations for the basic chain-of-events model) based on systems thinking and systems theory presents an opportunity to perform more powerful accident analysis and hence learning from events. Increasing emphasis on proactive analysis is also necessary. The author has proposed one such model of causality (called system-theoretic accident model and processes (STAMP)) (Leveson, 2004, 2009)) and some ways to analyze and prevent accidents using this model. Others are possible. But they must reflect the reality of today's complex sociotechnical systems and not oversimplify the causes of accidents.

One question remains regarding costs. Does it cost more to use more sophisticated and expanded models of causality based on systems thinking? Our experience is that the answer to this question is No. The application of systems thinking requires only a different interpretation of the information usually collected in an accident or incident investigation. We have not found that creating the hierarchical control structure and analyzing it (informally) is difficult or time-consuming. In addition, the structure only has to be created once and then simply reused and updated as changes are made. The model can also be used to evaluate the impact on safety of any proposed changes to ensure that they are not inadvertently reducing safety. Even better would be to create and use such a model in the design of the system so that accidents are prevented from occurring at all. If an incident or accident did occur, then one important activity in response should be determining where the original analysis was flawed and improving both it and the modeling and analysis procedures used.

The alternative to better accident analysis is to continue to have preventable accidents, the cost of which usually dwarfs the cost of more sophisticated hazard analysis. Every major accident has precursors that might have been used to prevent the major loss (Leveson, 1995). The Bhopal chemical plant catastrophe and the Three Mile Island nuclear power plant incident, for example, were both preceded by nearly identical events that, for various reasons, did not lead to a loss and were not properly investigated or analyzed and changes were not made as a result.

Another argument we have heard is that it is too expensive to analyze all the accidents and incidents that occur in some industries and companies. How does one select the incidents requiring in-depth analysis? That question is the wrong one to ask. Given the number of incidents and accidents that have identical systemic causes, simply investigating one or two in-depth could potentially eliminate dozens of incidents. For example, a superficial accident analysis might blame the occurrence of a loss on the flawed design of a relief valve and such relief valves might be replaced throughout the plant or company resulting in prevention of incidents due to that particular flawed valve design. Understanding why the flawed design was used, such as inadequate design, hazard analysis, review, testing, or other development or management practices, and improving those practices could have a potentially much greater impact on reducing a much larger number of future losses due to design inadequacies of all kinds.

References

Ackoff, R.L., 1971. Towards a system of systems concepts. Management Science 17 (11), 661–671.

Brehmer, B., 1992. Dynamic decision making: human control of complex systems. Acta Psychologica 81, 211–241.

Checkland, Peter, 1981. Systems Thinking, Systems Practice. John Wiley & Sons, New York.

Dekker, Sidney, 2007. Just Culture: Balancing Safety and Accountability. Ashgate.

Dulac, Nicholas, Brandon Owens, Nancy Leveson, Betty Barrett, John Carroll, Joel Cutcher-Gershenfeld, Stephen Friedenthal, Joseph Laracy, Joseph Sussman, 2007. Demonstration of a Powerful New Approach to Risk Analysis for NASA Project Constellation, MIT CSRL Final Report, March. <http://sunnyday.mit.edu/ESMD-Final-Report.pdf>.

Edwards, W., 1962. Dynamic decision theory and probabilistic information processing. Human Factors 4, 59–73.

Fujita, Y., 1991. What shapes operator performance? JAERI Human Factors Meeting, Tokyo, November.

iSixSigma, 2008. <http://www.isixsigma.com/dictionary/5_Whys-377.htm>.

JPL Special Review Board, 2000. Report on the Loss of the Mars Polar Lander and Deep Space 2 Missions, NASA Jet Propulsion Laboratory, 22 March.

Klein, Gary, A., Judith Orasano, R., Calderwood, Caroline, E., Zsambok (Eds.), 1993. Decision Making in Action: Models and Methods. Ablex Publishers.

La Porte, R., Todd, R., 1996. High reliability organizations: unlikely, demanding, and at risk. Journal of Contingencies and Crisis Management 63 (4).

Leveson, Nancy G., 1995. Software. Addison-Wesley, Reading, MA.

Leveson, N.G., 2004. A new accident model for engineering safer systems. Safety Science 42 (4), 237–270.

Leveson, Nancy, Nicolas Dulac, Betty Barrett, John Carroll, Joel Cutcher-Gershenfeld, Stephen Friedenthal, 2005. Risk Analysis of NASA Independent Technical Authority, NASA Final Report, June. (<http://sunnyday.mit.edu/ITA-Risk-Analysis.doc>)

Leveson, N.G., 2007. Technical and managerial factors in the NASA challenger and Columbia losses: looking forward to the future. In: Fleishman, Handelsman (Eds.), Controversies in Science and Technology, From Chromosomes to the Cosmos, vol. 2. Mary Ann Liebert, Inc..

Leveson, Nancy G., 2009. System Safety Engineering: Back to The Future (tentative title). Unpublished draft. <http://sunnyday.mit.edu/book2.html>

Pavlovich, J.G., 1999. Formal Report of the Investigation of the 30 April 1999 Titan IV B/Centaur TC-14/Milstar-3 (B32) Space Launch Mishap. US Air Force.

Perrow, Charles, 1999. Normal Accidents: Living with High-Risk Technologies. Princeton University Press.

Ramo, Simon, 1973. The systems approach. In: Ralph, F., Miles, Jr. (Eds.), Systems Concepts: Lectures on Contemporary Approaches to Systems. New York, John F. Wiley & Sons, pp. 13–32.

Rasmussen, Jens, 1990. Human error and the problem of causality in analysis of accidents. In: Broadbent, D.E., Reason, J., Baddeley, A. (Eds.), Human Factors in Hazardous Situations. Clarendon Press, Oxford, pp. 1–12.

Rasmussen, Jens, 1997. Risk management in a dynamic society: a modelling problem. Safety Science, vol. 27(2/3), Elsevier Science Ltd., pp. 183–213.

Rasmussen, Jens, Annelise Mark Pejtersen, Goodstein, L.P., 1994. Cognitive System Engineering. John Wiley & Sons.

Roberts, K.H., 1990. Managing high reliability organizations. California Management Review 32 (4), 101–114.

Rochlin, Gene, I., La Porte, Todd, R., Roberts, Karlene, H., 1987. The Self-Designing High Reliability Organization, Naval War College Review, Autumn.

Sheen, Barry, 1987. Herald of Free Enterprise Report, Marine Accident Investigation Branch, Department of Transport (originally Report of Court No 8074 Formal Investigation, HMSO, London).

Vicente, R., Kim, J., 1995. A Field Study of Operator Cognitive Monitoring at Pickering Nuclear Generating Station, Technical Report CEL 9504, Cognitive Engineering Laboratory, University of Toronto.

Weick, Karl, E., 1987. Organizational Culture as a Source of High Reliability. California Management Review, 29(2), pp. 112–127 (Winter).

Weick, Karl, E., Sutcliffe, K., Obstfeld, D., 1999. Organizing for high reliability. Research in Organizational Behavior 21, 81–123.

Woods, David, D., 1984. Some results on operator performance in emergency events. In: Whitfield, D., (Ed.), Ergonomic Problems in Process Operations. Institute of Chemical Engineering Symposium, Series, vol. 90.

Zsambok, R., Caroline, E., Gary Klein (Eds.), 1997. Naturalistic Decision Making. Lawrence Erlbaum Associates.