Beyond Normal Accidents and High Reliability Organizations: The Need for an Alternative Approach to Safety in Complex Systems

Karen Marais, Nicolas Dulac, and Nancy Leveson
MIT

March 24, 2004

1 Introduction

Organizational factors play a role in almost all accidents and are a critical part of understanding and preventing them. Two prominent sociological schools of thought have addressed the organizational aspects of safety: Normal Accident Theory (NAT) [15, 23] and High Reliability Organizations (HRO) [7, 8, 19, 20, 22, 25, 26, 27]. In this paper, we argue that the conclusions of HRO researchers (labeled HRO in the rest of this paper) are limited in their applicability and usefulness for complex, high-risk systems. HRO oversimplifies the problems faced by engineers and organizations building safety-critical systems and following some of the recommendations could lead to accidents. NAT, on the other hand, does recognize the difficulties involved but is unnecessarily pessimistic about the possibility of effectively dealing with them. An alternative *systems* approach to safety in described, which avoids the limitations of NAT and HRO. While this paper uses the Space Shuttle, particularly the Columbia accident, as the primary example, the conclusions apply to most high-tech, complex systems.

2 Normal Accidents

Charles Perrow's initial formulation of what has come to be known as Normal Accident Theory (NAT) was developed in the aftermath of the accident at the Three Mile Island nuclear power plant in 1979 [14]. Perrow introduced the idea that in some technological systems, accidents are inevitable or "normal" [15]. He defined two related dimensions—interactive complexity and loose/tight coupling—which he claimed together determine a system's susceptibility to accidents.

Interactive complexity refers to the presence of unfamiliar or unplanned and unexpected sequences of events in a system that are either not visible or not immediately comprehensible. A *tightly coupled* system is one that is highly interdependent: Each part of the system is tightly linked to many other parts and therefore a change in one part can rapidly affect the status of other parts. Tightly coupled systems respond quickly to perturbations, but this response may be disastrous. Loosely coupled or decoupled systems have fewer or less tight links between parts and therefore are able to absorb failures or unplanned behavior without destabilization.

According to the theory, systems with interactive complexity and tight coupling will experience accidents that cannot be foreseen or prevented. Perrow called these *system accidents*. When the system is interactively complex, independent failure events can interact in ways that cannot be predicted by the designers and operators of the system. If the system is also tightly coupled, the cascading of effects can quickly spiral out of control before operators are able to understand the situation and perform appropriate corrective actions. In such systems, apparently trivial incidents can cascade in unpredictable ways and with possibly severe consequences.

Perrow made an important contribution in identifying these two risk-increasing system characteristics. His conclusion, however, that accidents are inevitable in these systems and therefore systems for which accidents would have extremely serious consequences should not be built is overly pessimistic. The argument advanced is essentially that the efforts to improve safety in interactively complex, tightly coupled systems all involve increasing complexity and therefore only render accidents more likely.

The flaw in the argument is that the only engineering solution he considers to improve safety is redundancy. He correctly argues that redundancy introduces additional complexity and encourages risk taking. He provides many examples of how redundant safety devices or human procedures may not only be ineffective in preventing accidents, but can even be the direct cause of accidents. A near meltdown at the Fermi demonstration reactor in Michigan in 1966, for example, occurred when a piece of zirconium installed inside the reactor as an additional safety feature broke off and stopped the flow of coolant to the reactor core. The core was partially melted and the reactor was permanently disabled.

While his basic argument about redundancy is very compelling, the use of redundancy is not the only way to increase safety and many of the alternatives do not involve increasing complexity. In fact, the most cost-effective methods often reduce complexity. Redundancy and the use of protection systems are among the least effective and the most costly approaches to designing for safety [11]. The most effective approaches involve eliminating hazards or significantly reducing their likelihood by means other than redundancy, for example, substituting non-hazardous materials for hazardous ones, reducing unnecessary complexity, decoupling, designing for controllability, monitoring, interlocks of various kinds, etc. Operations can also be made safer by eliminating and reducing the potential for human error. A simple example is the use of color coding and male/female adapters to reduce wiring errors. Leveson describes many non-redundancy approaches to system design for safety in [11].

While it is usually not possible to predict all accident scenarios in complex systems, thorough engineering analyses of system hazards can be used to prevent whole classes of potential accidents. And there are often ways to mitigate the consequences without requiring completeness in identifying potential causes. For example, it is not necessary to predict all potential causes of a ship sinking in order to provide life boats and other emergency measures.

Perrow himself provides a clue as to how to design safer systems in his identification of the two important factors in system accidents, i.e., reduce the interactive complexity and tight coupling in our designs. Interactively complex and tightly coupled designs are created because they often allow greater functionality and efficiency to be achieved, but simpler, decoupled designs can usually achieve the same basic goals. The problem boils down to tradeoffs and determining how much risk is acceptable in order to achieve goals other than safety. Such decisions go beyond engineering; Alvin Weinberg, former head of Oak Ridge National Laboratory, called these *trans-scientific questions* because they involve addressing

moral, ethical, philosophical, and political questions. While engineers and scientists can provide important information to help inform decision making about how much risk is acceptable, they cannot provide definitive answers to these trans-scientific questions. But the point is that it is possible to reduce or eliminate risk by making appropriate tradeoff decisions.

3 High Reliability Organizations

The researchers most often associated with HROs are Todd La Porte, Gene Rochlin, Karlene Roberts, Karl Weick, and Paula Consolini. The first question is how High Reliability Organizations are defined and identified. Roberts [19] identifies HROs as the subset of hazardous organizations that enjoy a record of high safety over long periods of time:

> One can identify this subset by answering the question, 'how many times could this organization have failed resulting in catastrophic consequences that it did not?' If the answer is on the order of tens of thousands of times, the organization is 'high' reliability.' [19, p.160].

By this definition, it is difficult to think of any low reliability organizations: Any organization that did not have at least this type of safety record would be shut down immediately except in cultures or countries where frequent catastrophic consequences and death is acceptable. Such "failure" can be hypothesized to be possible every second of the day for most high-hazard activities and therefore a HRO might experience daily accidents and deaths, certainly something that would not be tolerated in our society. The only way to define safety or to compare organizations in terms of safety is to calculate accident rates based on particular activity over a common operational time period. In fact, the accident rate in aircraft carrier landings (a commonly cited example of an HRO) is relatively high compared to many other high-risk activities.

Assuming that such organizations can somehow be defined (in some other way than a circular argument that HROs are those organizations that exhibit the characteristics identified by the HRO researchers as good), HRO researchers seem to counter Perrow's hypothesis by suggesting that some interactively complex and tightly coupled systems operate with very few accidents. These conclusions are based on studies of two aircraft carriers, U.S. air traffic control, utility grid management, and fire fighting teams [8].

The most important flaw in this argument is that the systems they studied are not interactively complex nor tightly coupled according to Perrow's definitions. Air traffic control (ATC), for example, is as safe as it is precisely because the system design is deliberately decoupled in order to increase safety. The ATC system is carefully divided into non-interacting sectors and flight phases (enroute, arrival, and takeoff and landing) with the interfaces between the sectors and phases (for example, handoff of an aircraft between two air traffic control sectors) carefully limited and controlled. Loose coupling is also ensured by maintaining ample separation between aircraft so that mistakes by controllers can be remedied before they impact safety. Different parts of the airspace are reserved for different types of aircraft or aircraft operation (e.g., visual flight rules vs. instrument flight rules). Extra warning devices, such as TCAS, also exist to prevent accidents.

The functions of aircraft carrier landing and takeoff systems are similar to ATC (although much simpler) except that carriers operate in more extreme environmental conditions. Like ATC, the structure of aircraft carrier operations and systems reduce system

coupling and the availability of many different options to delay or divert aircraft, particularly during peacetime operation (which was when the HRO studies were done) introduces essential slack into the system. The fact that these relatively simple and loosely coupled systems are safe seem to support Perrow's arguments rather than contradict them.

The High Reliability Organization researchers themselves emphasize the low level of complexity in the example systems they studied:

> HROs struggle with decisions in a context of *nearly full knowledge of the technical aspects of operations* in the face of recognized great hazard ... The people in these organizations *know almost everything technical* about what they are doing—and fear being lulled into supposing they have prepared for every contingency ... This drive for technical predictability has resulted in *relatively stable technical processes* that have become quite well understood within each HRO. [8, pp.29-30] (emphasis added).

The fact that these systems allow perfect knowledge contradicts the definition of interactive complexity, which Perrow defined as system designs for which the interactions between components could *not* be thoroughly planned, understood, predicted, or guarded against. HROs are not interactively complex by the HRO researchers own descriptions. Their basic requirements for being an HRO omit interactively complex, tightly coupled systems by definition. If technical knowledge is complete, as required for HROs, it is relatively easy to lower risk through standard system safety and industrial safety approaches. Unfortunately, most complex systems, particularly high-tech and social systems, do not fall into this category.

The important factor here is *uncertainty*: technical, organizational, and social. It is uncertainty that makes engineering difficult and challenging and occasionally unsuccessful. The Space Shuttle is an example of a system at the other extreme from those studied by the HRO researchers. Technical uncertainty is inherent in any system as complex as the Shuttle. For example, although foam shedding from the external tank during liftoff has been identified as a problem for two decades, it is still not fully understood. All attempts to fix the problem have been unsuccessful. In complex systems, particularly those operating at the edge of technical innovation, there are always many technical uncertainties that cannot be resolved as required for HROs. At any time, the Shuttle has over 3000 waivers (allowing flight even though potential problems have not been completely resolved) out of a total of over 4000 Criticality 1 and 1R components.[1] Deciding which outstanding problems should be given priority is a difficult problem in itself. In addition, because many high-tech systems use new technology, understanding of the physical phenomena that may cause problems is often limited. NASA (and other organizations) cannot wait for complete understanding before launching technically complex systems. While the Shuttle is admittedly an extreme example, most high-tech systems have unresolved technical uncertainty. If it were necessary to resolve all uncertainty before use or operation, as required for HROs, most high-risk systems would need to be shut down and important functions provided in our society would come to a halt.

The uncertainties are not simply technical. The Shuttle, for example, exists in a highly uncertain political and budgetary environment. Budgets can change dramatically from year to year and even organizational goals, which may be dictated by politics, can change abruptly.

[1] Criticality 1 means the failure of the component could lead to the loss of the Shuttle. The category 1R indicates there is a redundant function that can potentially mitigate the failure.

An important distinguishing characteristic in high-risk systems is the source of the risk. In aircraft carriers, the risk stems not from technical uncertainty or even organizational structure or social factors, but from the difficulty of the task being controlled. Landing a plane on an aircraft carrier is the most difficult task required of a naval aviator and lots of accidents result (despite the claims of some HRO proponents that there are no or few accidents), although they are usually blamed on the pilot. For the Shuttle, the task itself is not only high risk, but the technical, organizational, and social structures used to perform the task also contain significant uncertainty. The risk in the aircraft carrier stems from uncertainty about individual operator behavior. The equivalent for the Shuttle is uncertainty about the behavior of those employees manufacturing the components, preparing the Shuttles for launch, and operating in Mission Control. But much of the risk in high-tech systems like the Shuttle (and the source of both the Challenger and Columbia accidents) is uncertainty in the engineering and engineering decision-making environment, which does not exist in the systems that the HRO researchers studied.

Some HRO researchers also claim that their example systems operate in an environment that is constantly changing [8, 20], but there is little evidence of this and, in fact, the quote above (where HROs are described as having relatively stable technical processes) contradicts this assumption of constant change. Air traffic control has essentially remained the same for the past 30 years. On an aircraft carrier, the environment is actually quite stable, at least insofar as the types of changes in the problem environment are very limited. Over the nearly 75 years of aircraft carrier existence, only a few major changes have occurred; the greatest changes resulted from the invention of jet aircraft. The introduction of improvements in carrier aviation, such as the angled flight deck, the steam catapult, and the introduction of mirror landing systems, have occurred slowly and over long time periods.

But even if the HRO argument is flawed with respect to claims that the systems studied are tightly coupled and interactively complex, the suggestions they make for improving safety may still be useful and need to be considered. Unfortunately, these suggestions are not very helpful for complex, socio-technical systems.

Four primary organizational characteristics have been identified that supposedly will substantially limit accidents and "failures" and simultaneously result in high levels of performance: (1) prioritization of both safety and performance and consensus about the goals across the organization [8]; (2) promotion of a "culture of reliability" in simultaneously decentralized and centralized operations [25]; (3) use of organization learning that maximizes learning from accidents, incidents, and near misses [8]; and (4) extensive use of redundancy [22]. The next four sections consider the practicality and usefulness of each of these for organizations trying to cope with high-risk activities and systems.

3.1 Goal Prioritization and Consensus

La Porte and Consolini claim that in high-reliability organizations, the leaders prioritize both performance and safety as organizational goals, and consensus about these goals is unequivocal [8]. While this state of affairs is clearly desirable, it is almost never possible. Safety goals often do not coincide with performance goals (unless the sole or primary purpose of the system is to maintain safety) and in fact often they conflict. In addition, while organizations often verbalize consensus about safety goals (e.g., "safety is our number one priority"), performance and decision making often departs from these public pronouncements.

For most of the organizations studied by HRO researchers, including aircraft carrier

operations in peacetime, U.S. air traffic control, and fire fighting teams, safety is either the only goal or primary goal of the existence of the organization so prioritizing it is easy. For example, in peacetime aircraft carrier operations (which was when La Porte and Consolini observed them), military exercises are performed to provide training and ensure readiness. There are no goal conflicts with safety: The primary goal is to get aircraft landed and launched safely or, if that goal is not successful, to safely eject and recover the pilots. If conditions are risky, for example, during bad weather, flight operations can be delayed or canceled without major consequences.

In wartime, the carrier's goals are subordinated to the larger goals of the military operation. The peacetime primary goal of safely getting aircraft on and off the carrier must now be combined with additional goals from strategic planners, including speed of operations. Human safety, aircraft safety, and even carrier safety may no longer be the highest priority. Further complicating the situation is the stress of being in combat. While we could find no statistics on the accident rates comparing aircraft operations on carriers during peacetime and during military operations, we note that there is evidence that carrier operations do seem to be affected by combat stress. The grounding of the *USS Enterprise* and the accidental shootdown of an Iranian commercial aircraft by the *USS Vincennes* indicate that combat conditions have a strong effect on high reliability performance [21].

Safety is not the primary goal of most organizations. Usually, the mission of the organization is something other than safety, such as producing and selling products or pursuit of scientific knowledge. In addition, it is often the case that the non-safety goals are best achieved in ways that are not consistent with designing or operating for lowest risk. In that case, each employee reliably performing their job is not enough to ensure safety—the employees may behave in a highly reliable manner to satisfy the non-safety goals while compromising safety in the process. Management statements that safety is the primary goal are often belied by pressures on employees to bend safety rules in order to increase production or to meet tight deadlines. An example was the issuance of "countdown to Node 2 launch" screensavers to all NASA Shuttle employees before the Columbia accident. This action reinforced the message that meeting the ISS construction milestones was more important than other goals, despite management claims to the contrary.

When safety goals conflict with other goals, the resolution of conflicts will not always result in putting the safety goals first by everyone in the organization unless there are safeguards to require this. The internal and external pressures to focus on the mission goals rather then safety goals may be extreme. The accelerated Shuttle launch pressures, for example, arose as NASA was being pushed by agencies such as the Office of Management and budget to justify its existence:

> This need to justify the expenditure [on the Shuttle] and prove the value of manned space flight has been a major and consistent tension between NASA and other governmental entities. The more missions the shuttle could fly, the better able the program was to generate funding. Unfortunately, the accelerated launch schedule also meant that there was less time to perform required maintenance or do ongoing testing. The results of these tensions appears to be that budgetary and program survival fears gradually eroded a number of vital procedures as well as supplanted dedicated NASA staff with contractors who had dual loyalties [2].

Consensus on prioritization of conflicting goals may waver when a company is forced to choose between operational safety goals (i.e., activities that lower risk) and their organization's continued existence. The goal of NASA's manned space flight organization is

to explore outer space and accrue scientific knowledge. Recognition that these activities involve risk is widely accepted. The problem is not only the simple one of prioritizing the safety goals—this would result in never launching any spacecraft (or producing chemicals, flying aircraft, generating electricity, etc.)—but making difficult tradeoffs and decisions about how much risk is acceptable and even how to measure the risk. Suggesting that NASA or any large organization should prioritize safety goals and operate reliably overly simplifies the difficulty in accomplishing these goals and is not very helpful in resolving the necessary tradeoffs and improving the complex decision-making involved. The flaw in the HRO argument is that safety is not black or white, but a matter of trying to determine how much risk exists in particular activities and decisions and how much risk is acceptable.

In a peacetime period, the Navy can afford to slow down carrier operations or keep a nuclear submarine in port for an extended time when safety goals are not met. But NASA and other organizations often have pressures, both internal and external, that limit their responses to goal conflicts. For example, the internal fight for primacy and survival by individual NASA centers, combined with external Congressional pressures to allocate functions and therefore jobs to centers in their own states, limits flexibility in designing programs.

A further complication is that individual employee goals may conflict with one or more of the organization's goals. Beliefs about the requirements for career advancement, for example, may lead employees to behave in ways that run counter to the organization's interests or to safety goals.

Organizations that contract out part of their functions have additional goal conflicts—the pressure to maintain supplier relationships is very great. NASA now has a large number of contractors working with and alongside civil servants:

> It is more difficult to come forward with negative information when you are employed by a firm that could lose its relationship with a prime customer; you also lose the place you have made within that customer organization. This is a situation full of mixed loyalties in which internal as well as external pressures come into play to affect actions. Analysis of these often intense pressures can provide insights into why gaps occurred in important functions such as information sharing and systems safety [2].

Even when an organization does attempt to clearly articulate and prioritize its internal goals, external forces (financial, political, or social) may limit the organization's options and so render these prioritizations ineffective.

We are not arguing that organizations cannot do a better job of conflict resolution and implement safeguards to counteract pressures to undervalue safety, but the arguments of some HRO researchers do not take into account the extreme difficulty of achieving this goal and of having everyone in the organization reach consensus on prioritizing safety goals above other goals, not just in speech, but in practice.

3.2 Simultaneously Decentralized and Centralized Operations

The second characteristic of so-called High Reliability Organizations is that organization members are socialized and trained to provide uniform and appropriate responses to crisis situations [25]. This field-level response to crises is the "decentralized response" that forms such a large part of HRO philosophy. The other side, "simultaneous centralization," refers to the maintenance of clear chains of command in crisis situations. For example, La Porte

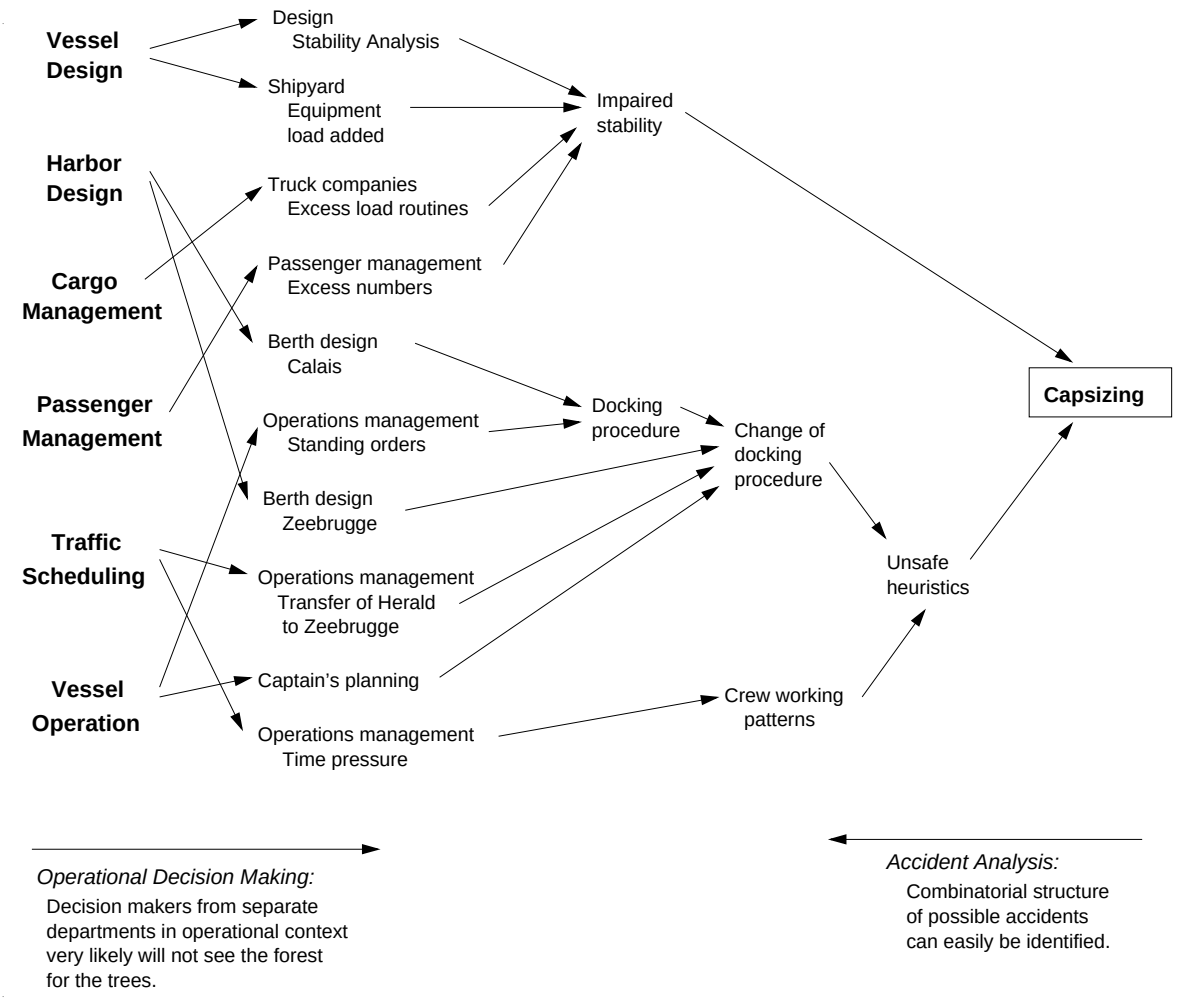

Figure 1: The Complex Interactions in the Zeebrugge Ferry Accident (Adapted from Rasmussen, Safety Science, p.188)

and Consolini argue that while the operation of aircraft carriers is subject to the Navy's chain of command, even the lowest-level seaman can abort landings. Clearly, this local authority is necessary in the case of aborted landings because decisions must be made too quickly to go up a chain of command. Overtraining of emergency responses is a standard practice in the training of operational personnel working in potentially dangerous, time-critical conditions. Note also that low-level personnel on aircraft carriers may only make decisions in one direction, that is, they may only abort landings. The actions governed by these decisions and the conditions for making them are relatively simple.

More interesting cases arise when decision-making is not time critical. La Porte and Consolini state that all personnel, regardless of rank, are trained to own a problem when they see it until it is solved or until someone who can solve the problem takes responsibility for it. This approach works only because the systems they studied were loosely coupled. In systems that are interactively complex and tightly coupled, taking individual action and acting alone may lead to accidents when local decisions are uncoordinated with other local or global decisions. Figure 1 shows an analysis by Rasmussen of the Zeebrugge ferry accident [17]. 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 decision-making advocated for HROs can lead to major accidents in complex socio-technical systems.

3.3 Organizational Learning

A third characteristic of HROs claimed by some proponents of this theory is that they use sophisticated forms of organizational learning. The argument is made that limiting learning to trial and error is not practical in these organizations. Instead, HROs use "imagination, vicarious experiences, stories, simulations, and other symbolic representations of technology and its effects" as substitutes for trial-and-error learning [25]. This process sounds exactly like what engineers do in hazard analysis, although the approaches engineers use are more rigorous than simply using "stories and vicarious experiences."

More interesting is the claim by some that HROs try to maximize learning from accidents, incidents, and near misses [8]. While it is difficult to argue against learning from mistakes, the costs of implementing effective organizational learning are high and the problems of competition for resources arises again. In addition, the difficulty of implementing effective organizational learning should not be underestimated. Finally, trial and error is not a very efficient way to learn, particularly for complex systems where the factors that can be involved in accidents may be very large. A billion dollar satellite was lost when only past errors that had led to accidents were considered; the human error that led to the loss had never been identified as the cause of an accident before and no safeguards were implemented.

The most important point here is that learning from accidents is not the only or even the most effective way to lower risk in high-tech systems. The organizations studied by HRO researchers are characterized by unchanging or very slowly changing designs and technology, which makes learning from accidents and incidents more effective. Organizations like NASA and the military that operate at the edges of technological innovation do not always have past experience from which to learn. Lessons learned on old technologies are also often inapplicable to newer ones. For example, digital systems are changing the nature of accidents and even changing the types of errors made by operators [24]. Experience with older, electro-mechanical systems does not apply to these new system designs and technology.

The average time to translate a basic technical discovery into a commercial product in the early part of the 20th century was 30 years. Today our new technologies get to market in two to three years and may be obsolete in five. We no longer have the luxury of (1) carefully testing systems and designs to understand all the potential behaviors and risks before commercial or scientific use or (2) depending on learning from our mistakes.

In addition, the potential losses stemming from accidents are increasing with the cost and potential destructiveness of the systems we are building and operating. Our new scientific and technical discoveries have not only created new or increased hazards (such as radiation exposure and chemical pollution), but have provided the means to harm increasing numbers of people as the scale of our systems increases and to impact future generations through environmental pollution and genetic damage. Financial losses and lost potential for scientific advances are also increasing in an age where, for example, a spacecraft may take 10 years and up to a billion dollars to build. Learning from accidents (the *fly-fix-fly* approach to safety) needs to be supplemented with increasing emphasis on preventing the

first accident through the use of hazard analysis, design for safety, and safety assurance techniques.

We are not arguing that organizational learning is not important, only that the difficulty should not be underestimated and that sophisticated hazard analysis (that includes organizational and social factors in accidents) and system safety engineering methods (including engineering the safety culture) may be more effective for high-tech and rapidly or constantly changing organizations and systems than relying on past experience and ad hoc scenario generation.

3.4 Extensive Use of Redundancy

A fourth characteristic often cited about HROs is the extensive use of redundancy. HROs are "characterized especially by flexibility and redundancy in pursuit of safety and performance," [7] where redundancy is defined as "the ability to provide for the execution of a task if the primary unit fails or falters" [8]. According to Roberts, HROs use technical redundancy, where parts are duplicated (e.g., backup computers) and personnel redundancy, where personnel functions are duplicated (e.g., more than one person is assigned to perform a given safety check) [20]. On aircraft carriers, for example, control for setting the arresting gear ultimately rests in the hands of at least three people, with oversight from the carrier's airboss.

The role of redundancy in increasing the safety of socio-technical systems is a major point of disagreement between Normal Accident Theory (NAT) and HRO. Once again, the problem seems to be that the proponents of each are arguing about completely different types of systems. Interactive complexity, tight coupling, and working in environments of uncertainty and imperfect knowledge limit the effectiveness of redundancy and, in fact, redundancy under these circumstances can actually increase the risk of an accident, as Perrow so persuasively argued.

The systems studied by HRO researchers are interactively *not* complex, *not* tightly-coupled, and, according to their own accounts, are characterized by low levels of uncertainty. In these relatively simple, decoupled systems, redundancy can be effective in preventing a single component failure (or sometimes multiple component failures) from leading to an accident. Even in these cases, however, there are limitations. For example, common-mode failures, where supposedly independent redundant components fail due to the same cause, limit the effectiveness of redundancy in protecting against component failure. A Lockheed L-1011 flying out of Miami in 1984 lost oil pressure in all three engines simultaneously because both mechanics did not put O-rings on three newly installed engine oil plugs. Inadequate preventive maintenance is one type of common error that will affect all components, including the supposedly redundant backups.

Redundancy depends on an assumption of random failure of the components for its effectiveness. But many, if not most, causes of accidents in interactively complex and tightly-coupled systems do not involve random component failure. In fact, engineers designing complex systems have long known about and tried to protect designs against accidents where no components fail, i.e., the accident is caused by dysfunctional interactions among components. The same is true for human components and human decision making. As illustrated by the Zeebrugge accident analysis described in the previous section, individual decisions may be reasonable in their context but not when combined at the system level. Redundancy will not help here.

The use of redundancy can, in fact, lead to dangerous decision making when false reliance

is based on it and the need for additional safety measures is discounted. The decision to launch the Challenger Space Shuttle on its fatal flight was partly based on overreliance on redundant O-rings. The failure of the primary O-ring led to the failure of the secondary O-ring [11]. Redundancy does not provide protection against underlying design errors, only random failures. Worse, the overconfidence provided by the redundancy convinced the decision makers that the Shuttle would survive a cold-weather launch even if the primary O-ring failed.

When systems contain software, redundancy is not useful in protecting against software commands that can lead to accidents. In fact, most software-related accidents can be traced back to errors in the software requirements, i.e., a misunderstanding about what the software was supposed to do under some circumstances. In these accidents, the software did exactly what the programmers intended it to do—it did not "fail." In addition, software redundancy management systems are so complex that they often introduce errors and can lead to system failures themselves.

Redundancy is only one limited way to increase reliability (but not necessarily safety) in some special cases; under other circumstances it can be the cause of or contributor to accidents. Complex, socio-technical systems need more sophisticated approaches to increasing reliability and safety for the non-random, technical, and organizational factors involved in accidents.

4 Limitations of Both NAT and HRO

Both the NAT and HRO theories oversimplify the cause of accidents. HRO underestimates the problems of uncertainty. NAT recognizes the difficulty of dealing with uncertainty but underestimates and oversimplifies the potential ways to cope with uncertainty. Both seem to believe that redundancy is the only way to handle risk.

The contribution of Perrow to understanding accidents in complex systems by identifying interactive complexity and tight coupling as critical factors should not be discounted. His top-down system view of accidents versus the bottom-up, component reliability view of the HRO theorists is critical in understanding and preventing future accidents. But the theory is incomplete and leads to more pessimism than required with respect to designing and operating complex high-risk systems. While the HRO theorists do offer more suggestions, most of them are inapplicable to complex systems or oversimplify the problems involved.

Organizational sociologists in general have made an important contribution to safety by emphasizing the organizational aspects of accidents. At the same time, they have underemphasized or oversimplified the engineering parts, for example, focusing only on simple redundancy, not considering accidents where component failure is not the cause, or studying only systems that are relatively simple and loosely coupled and then drawing conclusions from them to apply to all systems.

Another group of researchers, including Rasmussen [17], Hollnagel [6], Woods [28], and Leveson [12], most of whom come from a systems engineering and human factors background, have been advocating an alternative, *systems* approach to safety. The primary differences between a systems approach and the HRO and standard engineering approaches are: (1) top-down systems thinking rather than a bottom-up, reliability engineering focus and (2) a focus on the integrated socio-technical system as a whole and the relationships between the technical, organizational, and social aspects.

4.1 A Top-Down, Systems Approach to Safety

First, it is important to recognize the difference between reliability and safety. HRO researchers talk about a "culture of reliability" where it is assumed that if each person and component in the system operates reliably, there will be no accidents. Even Perrow seems to assume that accidents require failures of components. This assumption is simply not true. In complex systems, accidents often result from interaction *among* perfectly functioning components. The loss of the Mars Polar Lander was attributed to noise (spurious signals) generated when the landing legs were deployed during descent. The onboard software interpreted these signals as an indication that landing occurred and shut the engines down prematurely, causing the spacecraft to crash into the Mars surface. The landing legs and the software performed correctly, but the accident occurred because designers failed to account for all interactions between the leg deployment and the software descent engine control software .

The same applies to organizational decision making as illustrated by Rasmussen's analysis of the Zeebrugge ferry mishap. Each 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.

Highly reliable systems are not necessarily safe and highly safe systems are not necessarily reliable. Reliability and safety are *different* qualities and should not be confused. In fact, these two qualities 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. For example, some ways to reduce the accident rate on aircraft carriers would be to slow down the landing rates, only allow landing in the most perfect weather and the most ideal conditions, and only allow the most experienced pilots to make the landings. Clearly these operational conditions would conflict with the achievement of other goals, such as training for combat.

Reliability in engineering is defined as the probability that a component satisfies its specified behavioral requirements over time and under given conditions. 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 [11]. At the same time, accidents have resulted precisely because the operators *did* follow the predetermined instructions provided to them in their training. When the results of deviating from procedures are positive, operators are lauded but when the results are negative, they are punished for being unreliable. The HRO researchers correctly point out the need for operators sometimes to break the rules in order to prevent an accident, but incorrectly label their behavior as reliable. Such behavior is in fact *not* reliable with respect to following the specified rules or training; it is unreliable but safe. The distinction becomes extremely important when multiple, perhaps conflicting, goals are involved.

If the goal is to increase safety, then we should be talking about enhancing the *safety culture*, not the reliability culture. The safety culture is that part of organizational culture that reflects the general attitude and approaches to safety and risk management. Aircraft carriers do have a very strong safety culture and many of the aspects of this culture observed by the HRO researchers can and should be copied by other organizations, but labeling these

characteristics as "reliability" is misleading and can lead to misunderstanding what is needed to increase safety in complex, tightly coupled systems.

A systems approach to safety recognizes that safety is a property of the system as a whole, not a property of individual system components: The socio-technical system must be treated as an integrated whole using a top-down rather than a bottom-up perspective. This fact, in turn, implies that effectively tackling safety problems will require researchers and practitioners to step outside their traditional boundaries and take a broad view of the problems. The MIT Engineering Systems Division, for example, bridges the engineering and management disciplines to address challenges in the architecture, launch and deployment, operation, and sustainment of complex engineered systems. The emphasis in ESD is on the simultaneous consideration of social *and* technical aspects of systems, including social structures and cultures, social interaction processes, and individual factors such as capability and motivation as well as engineering design and analysis techniques.

In a systems view of safety, the traditional conception of accidents as chains (or trees) of directly related failure events and human errors is abandoned. Chain-of-events models encourage limited notions of linear causality and cannot account for the indirect, non-linear, and feedback relationships common for accidents in complex systems. A systems-theoretic approach to understanding accident causation allows more complex relationships between events to be considered and also provides a way to look more deeply at why the events occurred. Accident models based on systems theory consider accidents as arising from the interactions among system components and usually do not specify single causal variables or factors [10]. Whereas industrial (occupational) safety models focus on unsafe acts or conditions and reliability engineering emphasizes failure events and the direct relationships among these events, a systems approach takes a broader view of what went wrong with the system's operation or organization to allow the accident to take place.

Systems theory dates from the thirties and forties and was a response to the limitations of classic analysis techniques in coping with the increasingly complex systems being built [4]. The systems approach assumes that some properties of systems can only be treated adequately in their entirety, taking into account all facets and relating the social to the technical aspects [16]. These system properties derive from the relationships between the parts of systems: how the parts interact and fit together [1]. Thus, the systems approach concentrates on the analysis and design of the whole as distinct from the parts.

Safety is an *emergent* or system property, not a component property. Determining whether a plant is acceptably safe is not possible by examining a single valve in the plant (although conclusions can be reached about the valve's reliability). Safety can be determined only by the relationship between the valve behavior and the other plant components and often the external environment of the plant—that is, in the context of the whole. A component and its specified behavior may be perfectly safe in one system but not when used in another.

A basic concept in systems theory is that of *constraints*. Constraints are limitations on behavior or activities. Safety constraints are those limitations on behavior that prevent unsafe behavior or unsafe interactions among system components. For example, in a wartime situation, a safety constraint on aircraft operations in order to prevent friendly fire accidents is that the pilots must always be able to identify the nationality of the other aircraft or ground troops in the area around them. A large number of components of the military "system" each play a role in ensuring that this constraint is satisfied.

In systems approaches, socio-technical systems are treated as dynamic processes that are continually adapting to achieve their ends and to react to change in themselves and

their environment. The processes must not only satisfy appropriate safety constraints in their original design, but continue to do so as changes and adaptations occur over time. Accidents, using a systems approach, are viewed as the result of flawed processes involving interactions among system components, including people, societal and organizational structures, engineering activities, and physical system components. Understanding and preventing accidents requires:

- Identifying the system safety constraints necessary to prevent accidents;
- Designing the system to enforce the constraints, including understanding how the safety constraints could be violated and building in protections against these dysfunctional (unsafe) behaviors;
- Determining how changes in the processes over time, due to internal or external pressures and influences, could increase risk and building in protection to identify when the system is changing or adapting in a way that increases risk to an unacceptable level. The results of this activity can be used to define metrics and various forms of performance auditing to detect when risk is increasing.

The use of a systems approach creates the possibility of modeling and engineering the safety culture and organizational aspects of safety, including the entire socio-technical system—physical, organizational, and social. Some modeling techniques have been proposed by Rasmussen and Svedung [18] and Leveson [12, 13], but are beyond the scope of this paper.

5 Summary

The two prevailing organizational approaches to safety, Normal Accidents and HROs, both limit the progress that can be made toward achieving highly safe systems by too narrowly defining the problem and the potential solutions. We believe that a systems approach to safety will allow building safety into socio-technical systems more effectively and provide higher confidence than is currently possible for complex, high-risk systems.

References

[1] Russell L. Ackoff. Towards a system of systems concepts. *Management Science*, 17(11):661–671, July 1971.

[2] Barrett, B. Personal Communication, March 2004.

[3] Carroll, John S. Organizational Learning Activities in High-Hazard Industries: The Logics Underlying Self-Analysis. *Journal of Management Studies*, 35(6), November 1998, pp. 699–717.

[4] Peter Checkland. *Systems Thinking, Systems Practice*. John Wiley & Sons, New York, 1981.

[5] Gehman, Harold W. (Chairman). Columbia Accident Investigation Board Report, Volume 1. NASA and GAO, August 2003.

[6] Hollnagel, Erik. Understanding Accidents—from Root Causes to Performance Variability. in J.J. Persensky, B. Hallbert, and H. Blackman (Eds.), *New Century, New Trends: Proceedings of the 2002 IEEE 7th Conference on Human Factors and Power Plants*, 2002.

[7] La Porte, Todd R. High Reliability Organizations: Unlikely, Demanding, and At Risk. *Journal of Contingencies and Crisis Management*, 63(4), 1996.

[8] La Porte, Todd R. and Consolini, Paula. Working in Practice But Not in Theory: Theoretical Challenges of High-Reliability Organizations. *Journal of Public Administration Research and Theory*, 1, 1991, pp. 19–47.

[9] La Porte, Todd R. and Rochlin, Gene. A Rejoinder to Perrow. *Journal of Contingencies and Crisis Management*, 2(4), 1994.

[10] Jacques Leplat. Occupational accident research and systems approach. In Jens Rasmussen, Keith Duncan, and Jacques Leplat, editors, *New Technology and Human Error*, pages 181–191, John Wiley & Sons, New York, 1987.

[11] Leveson, Nancy G. *Safeware: System Safety and Computers.* Addison-Wesley, 1995.

[12] Leveson, Nancy G. A New Accident Model for Engineering Safer Systems. *Safety Science*, 42(4), April 2004, pp. 237–270.

[13] Leveson, Nancy G., Daouk, Mirna, Dulac, Nicolas, and Marais, Karen. A Systems-Theoretic Approach to Safety Engineering: A Case Study. Monograph of the ESD Symposium, March 2004.

[14] Charles Perrow. The President's Commission and the Normal Accident. in David L. Sills, C.P. Wolf, and Vivien B. Shelarski (Eds.), *The Accident at Three Mile Island: The Human Dimension*, Westview Press, 1982.

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

[16] Simon Ramo. The systems approach. In Ralph F. Miles Jr., editor, *Systems Concepts: Lectures on Contemporary Approaches to Systems*, pages 13–32, John F. Wiley & Sons, New York, 1973.

[17] Rasmussen, Jens. Risk Management in a Dynamic Society: A Modelling Problem. *Safety Science*, 27(2/3), 1997, pp. 183–213.

[18] Rasmussen, Jens, and Svedung, Inge. *Proactive Risk Management in a Dynamic Society.* Swedish Rescue Services Agency, 2000.

[19] Roberts, K.H (1990a). Managing high reliability organizations. *California Management Review*, 32(4), pp. 101–114.

[20] Roberts, K.H. (1990b) Some characteristics of one type of high reliability organization. *Organization Science*, 1(2), pp. 160–176.

[21] Rochlin, Gene. Iran Air Flight 655 and the USS Vincennes: Complex, Large-Scale Military Systems and the Failure of Control. in La Porte, T.R. (Ed.), *Social Responses to Large Technical Systems: Control or Anticipation*, NATO ASI Series, Kluwer Academic Publishers, 1991.

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

[23] Sagan, Scott. *The Limits of Safety*. Princeton University Press, 1995.

[24] Sarter, N. D. and D. Woods "How in the world did I ever get into that mode?": Mode error and awareness in supervisory control. *Human Factors 37*, 5–19.

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

[26] Weick, Karl E. and Roberts, Karlene H. Collective Mind in Organizations: Heedful Interrelating on Flight Decks. *Administrative Science Quarterly*, 38(3), September 1993, pp. 357–381.

[27] Weick, Karl E., Sutcliffe, K., and Obstfeld, D. Organizing for High Reliability. *Research in Organizational Behavior*, 21, 1999, pp. 81–123.

[28] Woods, David D., and Cook, Richard I. Nine Steps to Move Forward from Error. *Cognition, Technology, and Work*, 4(2), 2002, pp. 137–144.