Moving Beyond Normal Accidents and High Reliability Organizations: A Systems Approach to Safety in Complex Systems

Nancy Leveson, Nicolas Dulac, Karen Marais, and John Carroll

Nancy Leveson
Massachusetts Institute of Technology, USA

Nicolas Dulac
Massachusetts Institute of Technology, USA

Karen Marais
School of Aeronautics and Astronautics, Perdue University, USA

John Carroll
Massachusetts Institute of Technology, USA

Abstract

In this century society faces increasingly large-scale accidents and risks emerging from our own wondrous technologies. Two prominent organizational approaches to safety, Normal Accident Theory and High Reliability Organizations, have focused attention on a variety of industries that deal with hazardous situations, developed concepts to explicate organizational structure and culture, and debated whether accidents are inevitable in complex systems. We outline these approaches and identify some limitations, including narrow definitions, ambiguity about key concepts, confusion of reliability and safety, and overly pessimistic or optimistic conclusions. We believe that the debate between NAT and HRO can become a more productive three-way conversation by including a systems approach to safety emerging from engineering disciplines. The more comprehensive systems approach clarifies the strengths and weaknesses of NAT and HRO and offers a more powerful repertoire of analytic tools and intervention strategies to manage and control post modern risk in complex, high-tech, systems with their potential for catastrophic disruptions and losses.

Keywords: accidents, safety, risk, reliability, complexity, socio-technical systems

1. Introduction

Although accidents have always been part of the human condition, in this century society faces increasingly large-scale accidents and risks emerging from our own wondrous technologies: nuclear power plants and nuclear weapons, aircraft and air traffic control, genetically modified organisms, new chemicals, and computer software underlying nearly everything. The same technologies that enable growth and promise global prosperity may also cause major disruptions and undesirable long-term consequences. To cope with and manage such postmodern risks (Beck 1992), we need to understand not only the technologies, but also the organizations and institutions that implement, sustain, and co-evolve with the technologies. In this paper, we discuss organizational factors underlying safety and contrast three different approaches based in both social science and engineering.

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

Organization Studies
30(02&03): 227–249
ISSN 0170–8406
Copyright © 2009
SAGE Publications
(Los Angeles,
London, New Delhi
and Singapore)

DOI: 10.1177/0170840608101478

Theory (NAT) (Perrow 1999; Sagan 1995) and High Reliability Organizations (HROs) (La Porte 1996; La Porte and Consolini 1991; Roberts 1990a,b; Rochlin et al. 1987; Weick 1987; Weick and Roberts 1993; Weick et al. 1999). Unfortunately, we believe that these approaches have talked around each other because they have failed to carefully define some key concepts and to recognize some important distinctions such as the difference between reliability and safety. We believe that the debate between NAT and HRO can become a more productive three-way conversation by including a systems approach to safety, emerging from engineering disciplines. The more comprehensive systems approach clarifies the strengths and weaknesses of NAT and HRO and offers a broader repertoire of analytic tools and intervention strategies to manage risk. This approach is of particular value in addressing the complex interdependencies and systemic causes associated with risks in postmodern society.

2. The NAT–HRO Debate

Charles Perrow initially formulated what has become known as NAT after the Three Mile Island nuclear power plant accident. His basic argument is that the interactive complexity and tight coupling in some technological systems, such as nuclear power plants, leads to unpredictability of interactions and hence *system accidents* that are inevitable or 'normal' (Perrow 1999) for these technologies.[1] For Perrow, accidents arise from incidents or localized failures that spread to disrupt or damage the larger system. In more interactively complex, tightly coupled systems, there is insufficient time and understanding to control incidents and avoid accidents. Indeed, efforts to avoid accidents in such systems, such as building in redundancy to compensate for local failures, can create increased complexity that may undermine the very goal being sought (Sagan 1995). Three Mile Island exemplifies the features of a normal accident: a small local problem and incorrect mental models that linked actions with defects, resulting in a rapidly emerging crisis that created considerable damage and nearly produced a disastrous off-site release of radiation.

In an optimistic rejoinder to Perrow's pessimism, Todd La Porte (La Porte and Consolini 1991) and Karlene Roberts (1990a) characterized some organizations as 'highly reliable' because they had a record of consistent safety over long periods of time. By studying examples such as air traffic control and aircraft carrier operations, they identified features that they considered the hallmark of HROs, including technical expertise, stable technical processes, a high priority placed on safety, attention to problems, and a learning orientation. Weick et al. (1999) later offered five hallmarks of HROs: preoccupation with failure, reluctance to simplify interpretations, sensitivity to operations, commitment to resilience, and deference to experience. In short, the HRO researchers asserted that organizations can become highly reliable and avoid system accidents by creating the appropriate behaviors and attitudes (Weick and Roberts 1993). In particular, bureaucratic rules are seen as stifling expert knowledge: according to HRO theory, safety has to be enacted on the front lines by workers who know the details of the technology and who may have to invent new actions or circumvent 'foolish' rules in order to maintain safety, especially during a crisis.

Over time, the ‘debate’ between these positions developed as a contest of concepts and illustrative examples. NAT argued that we cannot know everything about these complex and hazardous technologies, and therefore the accidents we see are normal and inevitable. HRO argued that some organizations appear to have very rare problems despite daunting hazards, so they must be doing something right. Does a plane crash mean that NAT is right or does the reduction in plane crashes over time mean that HRO is right? Sagan’s (1995) masterful collection of horrendous near-misses in nuclear weapons handling could be read as a story of how close we came and how lucky we are (NAT) or how robust and well defended the system really is (HRO).

As readers of this literature, we experience considerable frustration that there seems to be no systematic analytical approach to resolving the debate as a victory for one side or the other or some integration of both into a more comprehensive theory of safety. The more we read, the more we came to believe that the reason for this lay in the way the theories were articulated. Something was missing, and we believe that a systems approach to safety can provide a constructive critique of both theories and a way forward for this community. In the next sections, we discuss some major weaknesses in NAT and HRO and then suggest what a systems approach to safety, based more directly on engineering principles as well as social science concepts, would offer instead.

3. Are Accidents Normal?

Perrow’s provocative thesis that complex and tightly coupled technological systems face normal accidents prompted many responses, including the HRO work that we discuss below. Although he was not the first social scientist to study major accidents (e.g. Turner 1978), his work was the starting point for many others to enter this area. In this section, we present and critically analyze Perrow’s core concepts and definitions as well as some arguments that have been presented against them by members of the HRO community.

Perrow’s argument for the inevitability of accidents in some industries has two parts, both flawed. The first part classifies industries in terms of complexity and coupling and suggests that risk is greater in those industries with high complexity and tight coupling. The second part argues that the reason for the higher risk in these industries stems from the ineffectiveness of redundancy in preventing accidents. We explore and critique these issues in this and the following section.

The first part of Perrow’s argument involves classifying industries by the amount of complexity and coupling (see his coupling/complexity chart, Perrow 1999: 97). He puts systems like nuclear weapons, aircraft, and military early warning in the tightly coupled, highly interactive quadrant of his chart.[2] One would then expect that, if his theory was correct, these industries would experience high accident rates or, at the least, higher accident rates than those in the other quadrants, but they do not. For example, there has never been an accidental detonation of a nuclear weapon in the 60 plus years of their existence. Commercial aircraft have a remarkably low accident rate. At the same time, he puts manufacturing in the lowest quadrant, but many manufacturing plants (e.g. oil refineries and chemical plants) have high accident rates. Mining, which is relatively low on the chart, is historically a very dangerous industry.

Perrow's basic argument about complexity seems obvious and correct, that is, more complex systems are likely to have higher accident rates because the potential interactions in such systems cannot be thoroughly planned, understood, anticipated, and guarded against — they go beyond engineers' ability to understand and manage intellectually. Such systems are more likely to have undetected design errors because they cannot be thoroughly analyzed or tested before use. They also will be harder for operators to manage in a crisis situation. So why does this argument lead to incorrect results when applied to predicting losses in particular industries? There are two important problems in the argument: inappropriate comparisons between incomparable properties (apples and oranges) and misclassification of industries along his dimensions.

The first problem is inappropriate comparisons. Perrow is basically arguing that some design features and systems have higher inherent risk than others. Determining whether this is true requires first defining risk. *Risk* is the combination of the likelihood of an event and the consequences of that event. Perrow's coupling/complexity classification considers only likelihood, ignoring the concept of a 'hazard', which is the event (or condition) being avoided. Comparing systems that involve different hazards leads to inconsistencies in the theory. For example, Perrow notes: 'Complex systems are not necessarily high-risk systems with catastrophic potential: universities, research and development firms and some governmental bureaucracies are complex systems' (Perrow 1999: 86) and further observes that bakeries transform materials (where he defines transformation of materials as a characteristic of high-risk systems) but are not high-risk. He explains this contradiction by claiming that bakeries and other low-risk transformation systems have only linear interactions. In fact, the design of these low-risk systems need not be linear (and many are not) and they will still be very safe with respect to the hazard of explosion or some other high-energy event. Transformation processes are dangerous only when high energy or some toxins are a product of the transformation and that energy or toxin release, is inadequately controlled. Plants manufacturing toxic chemicals or refining oil at high temperatures are inherently more dangerous than those manufacturing teddy bears or bonbons. To compare risk requires classifying systems according to the types of losses and hazards involved. Any comparison of risk must include similar hazards, because comparing likelihood only makes sense when we ask 'likelihood of what?'

But, second, even if we consider only the entries in Perrow's complexity/coupling classification with similar hazard potential, many of those predicted to be high-risk actually have very low historical accident rates. The problem is that Perrow's categorization of systems is arbitrary and inconsistent with the actual design of real systems in these industries.

In the first place, Perrow's argument oversimplifies engineering design by not differentiating between different types of complexity and coupling. He provides only vague definitions of these properties and instead gives examples and, in the case of complexity, a long (and incomplete) laundry list of design features labeled as complex. While many of these features do increase the difficulty of engineering design and thus increase the risk of design errors and operational errors and thus accidents, engineers distinguish among many different types of complexity (interactive, structural, dynamic, etc.) and coupling (time coupling, control coupling, data or information coupling, structural coupling, etc.) in devising ways to protect against potential errors and eliminate or control hazards. The type of hazard involved and the type of complexity and coupling required to achieve the system goals will affect design tradeoffs and the ability of engineers to design protections into the system. More differentiation between hazards and between types of design features is required to make likelihood comparisons.

Second, Perrow classifies all the systems in a particular industry, such as manufacturing or aircraft, as having the same amount of interactive complexity and coupling, which is simply untrue and does not match the actual designs found in these industries. Perrow acknowledges that:

> 'One serious problem cannot be avoided, but should be mentioned. To some unknown extent it is quite possible that the degree of coupling and types of interactions have been inferred from a rough idea of the frequency of system accidents in the various systems rather than derived from analysis of the properties of the systems independent of the nature of their failures. That is, if there are few accidents caused by air traffic control, that "must" mean it is not highly complex and tightly coupled, and then evidence for that conclusion is sought. Since the analytical scheme evolved from the examination of many systems, there is no way to avoid this possible circularity. The scheme would have to be tested by examining systems not included here, as well as by collecting data based upon more rigorous concepts and varying the placement of the systems that are included here.' (Perrow 1999: 97)

The problem he notes in this footnote could be solved by categorizing systems not by their domain but by their actual engineered design features and by making the observations directly on the degree and types of interaction and coupling in the designs. That would require using more careful definitions of interactive complexity and coupling, and distinctions between the types of coupling and complexity involved, to categorize the systems, and the result would lead to very different conclusions for the specific systems. For example, Perrow puts space missions in the complex and tightly coupled category, but, in fact, spacecraft designers use very conservative, loosely coupled designs.

Some HRO researchers have argued against NAT by pointing to supposedly interactively complex and tightly coupled systems that operate with very few accidents. These conclusions are based on studies of two aircraft carriers, US air traffic control, utility grid management, and fire fighting teams (La Porte and Consolini 1991). The most important flaw in this argument is the same as Perrow's: misclassifying systems as tightly coupled without carefully defining that property. In fact, using engineering definitions, the design of most of the engineered systems they studied are *neither* interactively complex nor tightly coupled. Air traffic control (ATC), for example, is as safe as it is precisely because the system has been deliberately designed to be *loosely coupled* 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) 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). Proximity warning devices, such as TCAS and Ground Proximity Warning

Systems, also help maintain separation. Similarly, the design of aircraft carrier operations and systems reduces 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 and safety margins into the system.

The contradictions in both the Perrow and HRO sides of the debate arise from confusion between science and engineering. Scientists observe systems that already exist (natural systems) and try to infer the design from their observations. In contrast, engineers start from a blank slate and create an original design for each engineered system. Those new designs may (and do) have varying degrees of complexity, including linear or non-linear interactions (usually there is a mixture of both), and the components may have varying levels and types of coupling. Engineers usually have control over the degree and types of coupling and complexity in the designs they create. While nuclear reactions, for example, may have many of the characteristics Perrow associates with tight coupling, all designs for nuclear power plants do not have to have the same properties (level of complexity and coupling) just because they are producing power using nuclear reactions. Perrow's incorrect conclusions may stem from his familiarity with the US nuclear power field, where historically one basic design has been mandated. But this consistency in design is a political artifact and is neither necessary nor practiced in most other types of systems or even in nuclear power plants in other countries. Natural properties of the physical system being controlled must be differentiated from the engineered design of the man-made systems built to control or use those natural processes.

An important reason for not simply making all engineered systems linear and loosely coupled is that such designs are often more inefficient and therefore may not accomplish the goals or mission of the system in an acceptable way. Engineering design is a search for optimal or at least acceptable tradeoffs between the engineered system properties (e.g. weight and cost), physical limitations (limitations of the physical materials being used or the natural processes being controlled), and various system objectives (e.g. performance). These tradeoffs and the uncertainties involved will greatly impact the likelihood of accidents.

The contribution Perrow makes by identifying complexity and coupling as characteristics of high-risk engineering design is substantial and important. The arguments about the normalcy of accidents in particular application domains, however, are flawed and rely on inadequate definitions, which accounts for the lack of correlation between the classification of systems as complex or coupled and their historical accident rates. The HRO proponents' counter-arguments, based on the same inadequate definitions, are equally flawed. Accidents in particular industries are not inherently normal or non-normal — risk depends on the specific design features selected and the technical and social uncertainties involved in that particular system. A goal of the systems approach, described in Section 8, is to provide risk management tools that decision makers (engineers, managers, regulators) can use to understand and control risk in engineered designs and operations and to assist in evaluating alternative social and organizational policies and structures.

4. Engineering Design and Redundancy

Even if Perrow's classification of all systems within particular industries as having the same risk is flawed, his conclusion that accidents are inevitable in complex systems could still hold. The second part of his argument is essentially that the efforts to improve safety in tightly coupled, interactively complex systems all involve increasing complexity and therefore only render accidents more likely.

Perrow is correct that redundancy is limited in its effectiveness in reducing risk. Redundancy introduces additional complexity and encourages risk taking. Perrow 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. The decision to launch the Challenger Space Shuttle on its fatal flight, for example, was partly based on over-reliance on redundant O-rings. The failure of the primary O-ring led to the failure of the secondary O-ring (Rogers 1986), that is, the failures in the redundant components were not independent. 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, and this overconfidence contributed to the incorrect decision making. Common-cause and common-mode failures and errors, both technical and human, can defeat redundancy. Redundancy itself makes systems more complex and therefore more difficult to understand and operate.

While Perrow's basic argument about redundancy is very compelling, the flaw in his larger argument is that the use of redundancy is not the only way to increase safety, and many of the alternatives do not involve increasing complexity and may even reduce it. Redundancy and the use of protection systems are among the least effective and the most costly approaches to designing for safety (Leveson 1995). The most effective (and usually the least costly) 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,[3] designing for controllability, monitoring, using 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 (1995) describes many non-redundancy approaches to system design for safety.

The role of redundancy in increasing the safety of socio-technical systems is a point of disagreement between NAT and HRO. HROs have been described as being 'characterized especially by flexibility and redundancy in pursuit of safety and performance' (La Porte 1996), where redundancy is defined as 'the ability to provide for the execution of a task if the primary unit fails or falters' (La Porte and Consolini 1991). According to Roberts (1990b), 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). 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 air boss.

Once again, the problem seems to be that the proponents of each viewpoint (NAT and HRO) are arguing about completely different types of systems and are oversimplifying the causes of accidents. Perrow is arguing about the potential for design errors in complex, tightly coupled systems. He is correct that redundancy does not protect against system design errors and, in fact, redundancy under such circumstances can actually increase the risk of an accident. The HRO examples of the effective use of redundancy are in loosely coupled systems where the redundancy is protecting against accidents caused by individual, random component failures rather than system design errors. If the system designs are loosely coupled, redundancy can reduce accidents caused by component failure. Many, if not most, causes of accidents in interactively complex and tightly coupled systems, however, do not involve random component failure, particularly organizational, cultural, and human factors, and redundancy will not prevent those accidents.

The emphasis on redundancy in some HRO literature arises from misunderstandings (by both Perrow and the HRO researchers) about the cause of accidents, that is, both groups assume they are caused by component failures. This confusion of component reliability with system safety leads to a focus on redundancy as a way to enhance reliability, without considering other ways to enhance safety. We explore this important distinction in the next section.

5. Reliability vs Safety

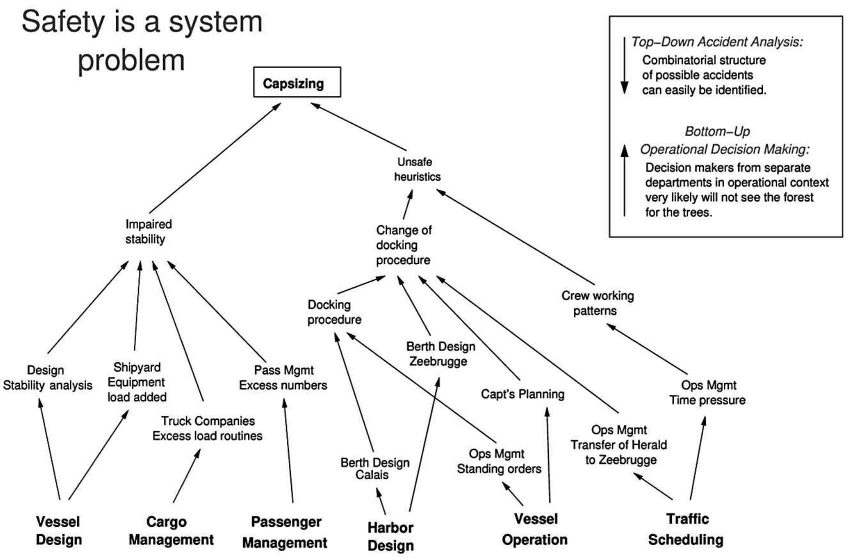
Safety and reliability are *different* properties. One does not imply nor require the other — a system can be reliable and unsafe or safe and unreliable. In some cases, the two system properties are conflicting, that is, making the system safer may decrease reliability and enhancing reliability may decrease safety. To fully understand the differences and even potential conflicts between reliability and safety requires defining terms. Reliability in engineering is defined as the probability that a component satisfies its specified behavioral requirements over time and under given conditions. Safety can be defined as freedom from unacceptable losses (accidents). Note that the reliability of nuclear power plants with the same design as Chernobyl is very high, i.e. the calculated mean time between failures is 10,000 years.

HRO theory (as denoted even by the name) treats safety and reliability as equivalent. The papers 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.

Perrow also seems to assume that accidents require failures in his definitions of accident and incident. He defines an *accident* as a failure in a subsystem, or the system as a whole, that damages more than one unit and in so doing disrupts the ongoing or future output of the system (i.e. the output ceases or decreases to the extent that prompt repairs will be required). An *incident* is defined as a failure involving damage that is limited to parts or a unit, whether the failure disrupts the system or not.[4]

These assumptions are not true. In complex systems, accidents often result from interaction among perfectly functioning (reliable and non-failed) components. For example, the loss of the Mars Polar Lander was attributed to noise

Figure 1.
The Complex Interactions in the Zeebrugge Ferry Accident (Adapted from Rasmussen 1997: 188)


Safety is a system problem
Capsizing
Unsafe heuristics
Impaired stability
Change of docking procedure
Docking procedure
Crew working patterns
Berth Design Zeebrugge
Design Stability analysis
Shipyard Equipment load added
Pass Mgmt Excess numbers
Capt's Planning
Ops Mgmt Time pressure
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.

(spurious signals) generated when the landing had legs were deployed during descent. 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 had 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) as did the descent engines and all the other spacecraft components. The accident occurred because designers failed to account for all interactions between the leg deployment and the descent-engine control software (Albee et al. 2000).

The same phenomenon occurs at the organizational and social levels above the physical system, as illustrated by Rasmussen's analysis of the Zeebrugge ferry mishap (Rasmussen 1997) shown in Figure 1. In this accident, those independently making decisions about vessel design, harbor design, cargo management, passenger management, traffic scheduling, and vessel operation (shown at the bottom of Figure 1) were unaware of how their design decisions might interact with decisions made by others and lead to the ferry accident. 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 (portrayed by intersecting upward arrows in the figure). 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 rather than the component level. More discussion of this distinction can be found in the description of the systems approach to safety in Section 8.

Accidents like the Mars Polar Lander, where the cause lies in dysfunctional interaction of non-failing, reliable components (i.e. the problem is in the overall system design) illustrate reliable components in an unsafe system. There can also be safe systems with unreliable components if the system is designed and operated so that component failures do not create hazardous system states. Redundancy, in fact, is only one of many ways to protect against unreliable components leading to accidents (Leveson 1995).

Even at the system level, reliability and safety are not equivalent and, in fact, they often conflict: increasing system reliability may decrease system safety and increasing system safety may decrease system reliability. One of the challenges of engineering is to find ways to increase system safety without decreasing system 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. In fact, almost all systems have multiple and sometimes conflicting goals, so that achieving all goals in a highly 'reliable' manner is impossible. There are often extreme pressures to reliably achieve the non-safety goals in such a way that risk is increased.

While, in some systems, safety is part of the mission or reason for existence, e.g. ATC 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. The mission is not to be safe in terms of not exposing bystanders to toxins or not polluting the environment. These are constraints on how the mission can be achieved. The best way to ensure the safety and environmental constraints are satisfied is not to build or operate the system at all. The (non-existent) plant is 'unreliable' with respect to its mission, but it is safe. Alternatively, a particular plant may very reliably produce chemicals while poisoning those around it. The plant is reliable but unsafe. There are always multiple goals and constraints for any system — the challenge in engineering and management decision making is to make tradeoffs among multiple requirements and constraints when the designs and operational procedures for best achieving the requirements conflict with the constraints.

6. Decentralization of Safety-Related Decision Making

HRO theorists have asserted that professionals at the front lines can use their knowledge and judgment to maintain safety (or reliability). They claim that during crises, decision making in HROs migrates to the front-line workers who have the necessary judgment to make decisions (Weick et al. 1999). The problem is that the assumption that front-line workers will have the necessary knowledge and judgment to make decisions is not necessarily true. While examples exist of operators ignoring prescribed procedures that would have been unsafe in particular circumstances, and as a result preventing an accident (Leveson 1995; Perrow 1999), in other instances, operators ignoring prescribed procedures have frequently caused losses that would not otherwise have occurred. The information required to distinguish between these two cases is usually available only in hindsight and not when the decisions need to be made.

Decentralized decision making is, of course, required in some time-critical situations. But like all safety-critical decision making, the decentralized decisions must be made in the context of system-level information and from a total systems perspective in order to be effective in reducing accidents. The most common way to accomplish this (in addition to decoupling system components so that decisions do not have system-wide repercussions) is to specify and train standard emergency responses. Safe procedures are determined at the system level and operators are usually socialized and trained to provide uniform and appropriate responses to crisis situations.

There are situations, of course, when unexpected conditions occur (Perrow's system accidents) and avoiding losses requires the operators to violate the specified (and in such cases unsafe) procedures. If the operators are expected to make decisions in real time and not just follow a predetermined procedure, then they usually must have system-level information about the situation in order to make safe decisions (if, again, the components have not been decoupled in the overall system design in order to allow independent safe decisions).

As an example, La Porte and Consolini (1991) 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. But note that low-level personnel on aircraft carriers may only make decisions in one direction: they may only abort landings, that is, change to an inherently safe state with respect to the hazard involved. System-level information is not necessary for this special case where there is a safe state that has no conflicts with other critical goals. The actions governed by these decisions and the conditions for making them are relatively simple. Aircraft carriers are usually operating in areas containing little traffic (i.e. decoupled from the larger system) and thus localized decisions to abort are almost always safe and can be allowed from a larger-system safety viewpoint.

In a high-traffic area, such as a go-around by a pilot at a busy airport, the situation is not so clear. While executing a go-around when a clear danger exists if the pilot lands is obviously the right decision; there have been recent near misses when a pilot executed a go-around and came too close to another aircraft that was taking off on a perpendicular runway. The solution to this problem is not at the decentralized level —the individual pilot or controller lacks the system-level information to avoid hazardous system states — but at the system level, where the danger has to be reduced by instituting different landing and takeoff procedures, building new runways, redistributing air traffic or by making other system-level design changes. We still want pilots to be able to execute a go-around if they feel it is necessary, but unless the system is designed to prevent collisions, the action decreases one hazard while increasing another one.

7. Generalization from Special Cases

In the HRO literature, HROs are identified as the subset of hazardous organizations with good safety records over long periods of time (Roberts 1990a). But selecting on the dependent variable does not guarantee that the practices observed in organizations with good safety records are the reason for that success or that these practices can be applied elsewhere with similar results.

Indeed, the systems and organizations often cited in the HRO literature have such good safety records because they have distinctive features that make the practices they use to improve safety rates difficult or impossible to apply in other organizations. For example, La Porte and Consolini have characterized HRO organizations in the following manner:

> '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.' (La Porte and Consolini 1991: 29–30; emphases added)

While these properties certainly help to engineer and operate safer systems and they do exist in the systems that were studied, they do not apply to most systems.

The first property identified for HROs is that they have nearly full knowledge of the technical aspects of operations. If technical knowledge is complete, however, it is relatively easy to lower risk through standard system safety and industrial safety techniques. As Perrow noted, the challenges arise in complex systems when the interactions between components *cannot* be thoroughly planned, understood, predicted, or guarded against, i.e. when full knowledge does not exist. In fact, complete technical knowledge does not exist in most high-risk systems, and society is usually unwilling to defer the benefits of these systems until that knowledge can be obtained, perhaps only after decades of research. Most systems must operate under uncertainty (technical, organizational, economic, and market), and the level of uncertainty is an important dimension of risk. To avoid accidents, and indeed losses of all kinds, the system must be able to cope with uncertainty, usually in ways that will and should differ depending on the specific characteristics of the system involved. The systems approach to organizational safety presented later embodies this philosophy.

The second property of HROs in the quote above is that they have relatively stable technical processes and thus opportunities to learn from operating experience. Unfortunately, this property is violated when new technology is introduced and process and product changes are made to improve efficiency, production, or other important goals. Air traffic control has essentially remained the same for the past 30 years. But this stability (which stems not from a desire to avoid changes but from inability to successfully and safely introduce new technology) has led to potential gridlock in the skies and has stymied attempts to introduce efficiency into the system and increase capacity. While technical stability has improved accident rates, it is not a practical or desirable goal for most organizations, particularly profit-making organizations that must compete on innovation, efficiency, quality, and other attributes.

In another classic HRO example, landing on aircraft carriers, the environment has been quite stable, at least insofar as the types of changes have been 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 mirror landing systems, has occurred slowly and over long time periods. The time dimension of design changes is yet another important dimension of risk and provides tension between the desire to maintain low risk and the desire to introduce changes to achieve other goals such as increased productivity. Occasionally they overlap —the changes are being introduced purely to increase safety — but, even then, uncertainty about the efficacy of the changes in reducing risk itself has an impact on the operational risk of the enhanced design.

In summary, an important problem with HRO theory is that the practices were observed in systems with low levels of uncertainty and stable technical processes. For most systems in competitive industries where technological innovation and advances are necessary to achieve the system mission and goals, these features do not exist or are not practical. The practices the HRO researchers observed in these special cases may not apply to other systems or may be much more difficult to implement in them.

HRO practices have been identified by observing organizations where safety goals are buffered from conflicts with other goals because of the nature of the mission. For example, 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 (La Porte and Consolini 1991). While this state of affairs is clearly desirable, it is much easier to achieve if safety is indeed the paramount goal of the organization. For many of the organizations studied by HRO researchers, including aircraft carrier landing operations in peacetime, US air traffic control, and fire fighting teams, safety is either a primary goal or the primary reason for the existence (i.e. the mission) 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.

For most organizations, however, the mission is something other than safety, such as producing and selling products or pursuing 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. 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 computer screensavers to all NASA Shuttle employees, before the Columbia accident, that counted down by seconds to the deadline for completion of the International Space Station. This action reinforced the message that meeting the ISS construction milestones was more important than other goals, despite management pronouncements to the contrary.

On an aircraft carrier during 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 and potentially contradictory goals from strategic planners, including speed of operations. Human safety, aircraft safety, and even carrier safety may no longer be the highest priority.

Analogously, NASA and most profit-making 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. In healthcare, where the risks themselves can conflict and often require trading one risk for another, prioritization of safety over other goals makes no sense. The problem in healthcare involves trading one risk for another, that is, the risk in not getting a particular treatment versus the risks inherent in the treatment itself, such as adverse side-effects. There are also other difficult healthcare tradeoffs such as the ordering of actions (triage) or saving many people versus saving a few.

The problem is not simply 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. For this, sophisticated risk analysis and risk management procedures and tools to support decision making are required, along with social technologies to reach consensus among stakeholder groups, some of which are less powerful or more vulnerable (Beck 1992), and to avoid losing social trust or societal support (Kasperson 1986). In contrast to the HRO argument that safety must be primary, as if safety is a yes or no, black white decision, managing system safety is a continuous process of trying to determine how much risk exists in particular activities and decisions, how much risk is acceptable, and how to achieve multiple system goals and requirements.

The only organization we have found that seems to have been successful in operationalizing total prioritization successfully is the SUBSAFE program in the nuclear navy (and perhaps carrier landing operations during peacetime, although that is not entirely clear with respect to conflicts between safety and training goals). The SUBSAFE program focuses only on submarine hull integrity to preclude flooding and on the operability and integrity of critical systems to control and recover from a flooding casualty. There are few conflicts or tradeoffs here — loss of a submarine due to flooding is always disastrous to the mission goals. Other aspects of submarine safety use traditional system safety engineering techniques and procedures, and losses have occurred, but no loss involving a lack of hull integrity has occurred in the 45 years of the program's existence.

In addition to its limited focus, SUBSAFE operates in an environment that differs in significant ways from most other environments. For example, the US Navy is non-profit, it operates under a strict command and control structure, it is unlikely to go out of existence due to market pressures and competition, and failure to achieve mission goals is out of the public eye (unlike NASA). None of these factors take away from the astoundingly successful design and operation of the SUBSAFE program (within its unique environment) and much can be learned from this success, but simply reproducing the SUBSAFE program without significant changes in a different environment may not be practical and may lead to less success. In general, observing a few special cases and assuming the practices observed will ensure safety in all organizations oversimplifies the complex problems involved.

8. The Top-Down, Systems Approach to Organizational Safety

Organizational sociologists have made important contributions to safety. Perrow drew attention to the critical factors of interactive complexity and tight coupling in accidents. But NAT 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 and more optimism about the potential for achieving acceptable levels of safety in complex organizations, most of their suggestions, as argued above, are inapplicable to interactively complex, tightly coupled, high-tech systems with complex goal structures. Both approaches use vague and sometimes shifting definitions, oversimplify the cause of accidents, and confuse reliability with safety.

Another group of researchers, including Rasmussen (1997), Woods and Cook (2002), Dekker (2005), Hollnagel (2004), and Leveson (2004), most of whom come from system engineering and human factors backgrounds, have advocated an alternative, *systems* approach to technical and organizational safety. The primary characteristics of a systems approach are: (1) top-down systems thinking that recognizes safety as an emergent system property rather than a bottom-up, summation of reliable components and actions; (2) focus on the integrated socio-technical system as a whole and the relationships between the technical, organizational, and social aspects; and (3) focus on providing ways to model, analyze, and design specific organizational safety structures rather than trying to specify general principles that apply to all organizations. There are many potential ways to achieve safety goals. The goal in organizational safety should be to create technical and organizational designs requiring the fewest tradeoffs between safety and other system goals while considering the unique risk factors (including uncertainty) and risk characteristics involved in the organizational mission and environment.

While systems approaches to safety have been proposed by several researchers, Leveson's STAMP (Systems-Theoretic Accident Modeling and Processes) approach (Leveson 2004) goes the farthest toward a pure systems' approach and differs the most from NAT and HRO with respect to assumptions about the causes of accidents and the analysis of social and cultural factors in accidents. We present STAMP as a contrast between systems approaches and the HRO and NAT views of organizational aspects of safety.

8.1 Basic STAMP Theory

As argued in Section 5, safety is an *emergent* or system property, rather than a component property. In systems theory, complex systems are viewed as a hierarchy of organizational levels, each level more complex than the one below. The levels are characterized by *emergent* properties that are irreducible and represent constraints on the degree of freedom of components at the level below.

Determining whether a nuclear power plant is acceptably safe, for example, is not possible by examining a single valve in the plant. Conclusions can be reached about the reliability of the valve, where reliability is defined as the probability that the behavior of the valve will satisfy its specification over time and under given conditions. But the 'safety of the valve' is meaningless: safety can only be determined by the relationship between the valve and the other plant components, that is, in the context of the whole.

In a systems-theoretic view of safety, the emergent safety properties are controlled or enforced by a set of safety *constraints* related to the behavior of the system components. Safety constraints specify those relationships among system variables or components that constitute the non-hazardous or safe system states. For example, the power must never be on when the access door to the high-power source is open; pilots in a combat zone must be able to identify targets as hostile or friendly; the public health system must prevent the exposure of the public to contaminated water; or the air traffic control system must maintain minimum separation between aircraft. Accidents result from interactions among system components that violate these constraints — in other words, from a lack of appropriate and effective constraints on component and system behavior.

The problem of ensuring safety, then, can be stated as a *control* problem, rather than a component failure problem: accidents occur when component failures, external disturbances, and/or dysfunctional interactions among system components are not adequately controlled or handled. While it is true that the O-ring failed in the Challenger Space Shuttle accident, that was only a part of the larger problem that the O-ring did not control (prevent) propellant gas release by sealing a gap in the external tank field joint. In the Mars Polar Lander loss (where no components 'failed'), the software did not adequately control the descent speed of the spacecraft — it misinterpreted noise from a Hall-effect sensor as an indication that the spacecraft had reached the surface of the planet and turned off the descent engine prematurely.

Losses such as these, involving engineering design errors, may in turn stem from inadequate control of the development process, that is, risk is not adequately managed in design, implementation, and manufacturing. Control is also imposed by the management functions in an organization — the Challenger and Columbia accidents, for example, involved inadequate controls in the launch-decision process and in the response to external pressures — and by the political system within which the organization exists. Note that the use of the term 'control' does not imply a strict military command and control structure. Behavior is controlled not only by direct management intervention, but also indirectly by policies, procedures, shared values, and other aspects of the organizational culture. All behavior is influenced and at least partially 'controlled' by the social and organizational context in which the behavior occurs. Engineering this context can be an effective way of creating and changing a safety culture.

The hierarchical safety control structure (i.e. the organizational and physical control structure) must be able to enforce the safety constraints effectively. Figure 2 shows an example of a hierarchical safety control structure for a typical US regulated industry, such as aircraft. Each industry and company (and each national governance system) will, of course, have its own unique control structure.

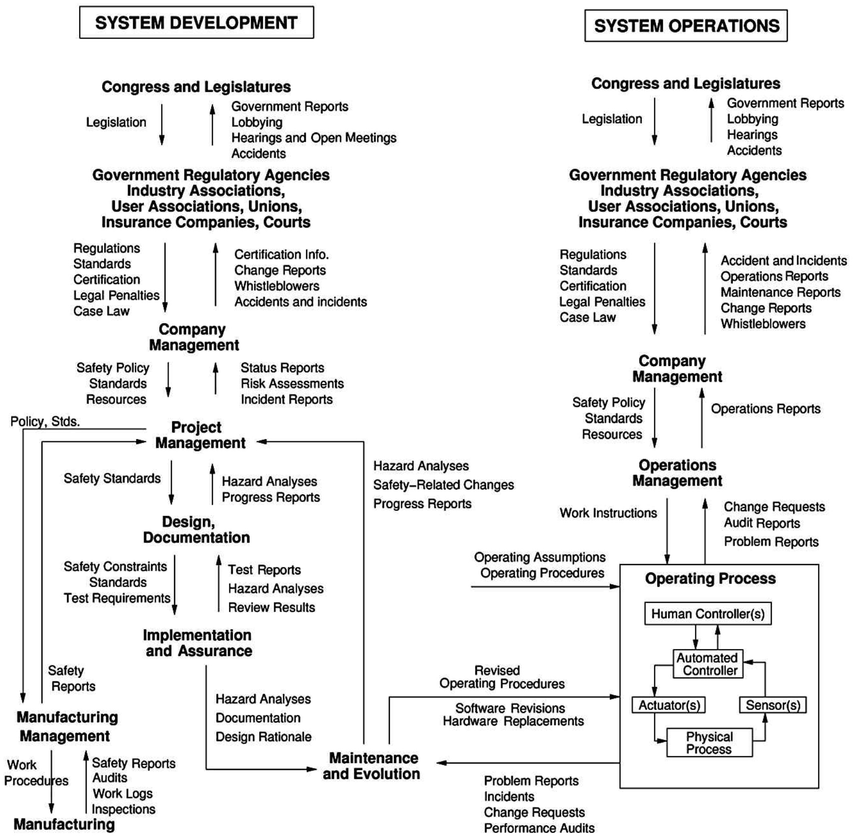
Accidents result from inadequate enforcement of constraints on behavior (e.g. the physical system, engineering design, management, and regulatory behavior) at each level of the socio-technical system. There are two basic hierarchical control structures in Figure 2 — one for system development (on the 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 operation 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.

Hierarchies, in systems theory, are characterized by control and communication processes operating at the interfaces between levels (Checkland 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 Figure 2, company management has a role in the development of the safety control structure by providing a safety policy, standards, and resources to project management and, in return, receiving status reports, risk assessments, and incident reports as feedback about the status of the project with respect to the safety constraints.

Another important concept in systems theory is *process models*. Any controller — human or automated — must contain a model of the system being controlled (Conant and Ashby 1970). For humans, this model is commonly known as a mental model. Accidents, particularly those arising from dysfunctional interactions among components, frequently result from inconsistencies between the model of the process used by the controllers and the actual process state; for example, the Mars Lander software thinks the lander has reached the surface and shuts down the descent engine; the Minister of Health has received no reports about water quality problems and believes the state of water quality in the town is better than it actually is and makes decisions on that basis; or a NASA Space Shuttle mission manager believes that foam shedding is a maintenance or turnaround issue only and underestimates the consequences of a foam strike on the Shuttle. Part of the modeling efforts using a systems approach to safety involves creating the process models, examining the ways they can become inconsistent with the actual state (e.g. missing or incorrect feedback), and determining what feedback loops are necessary to maintain the safety constraints and how to implement them.

When there are multiple controllers and decision makers, i.e. distributed decision making, accidents may involve unexpected side-effects of decisions or

Figure 2.
General Form of a Model of Socio-Technical Control



actions and conflicts between independently made decisions (see Figure 1), often the result of inconsistent process models. For example, two decision makers may both think the other is making the required control action, or they may implement control actions that conflict with each other. Communication plays an important role here. Leplat suggests that accidents are most likely in *boundary* or *overlap* areas where two or more controllers control the same process (Leplat 1987). One potential use for STAMP models is to determine what communication channels and other system design features are necessary to provide adequate safeguards for distributed decision making.

The safety control structure often changes over time, which accounts for the observation that accidents in complex systems frequently involve a migration of the system toward a state where a small deviation (in the physical system or in human behavior) can lead to a catastrophe (Rasmussen 1997). The foundation for an accident is often laid years before. One event may trigger the loss, but if that event had not happened, another one would have. The control structure must be carefully designed and evaluated to ensure that the controls are adequate to maintain the constraints on behavior necessary to control risk, including preventing migration toward states of higher risk or detecting them before a loss occurs.

Using systems and control theory, safety-related control flaws can be classified and provide the foundation for designing safer systems, both technical and social. Figure 2 shows an example of a static model of the safety control structure. But understanding why accidents occurred (and how to prevent them in the future) also requires understanding *why* the structure changed over time, in order to build in protection against unsafe changes. System dynamics (Sterman 2000) or other dynamic models can be used to model and understand these change processes.

8.2 Applying STAMP to Organizational Safety

Using systems theory as a foundation, existing organizational safety control structures can be evaluated and improved or can be designed from scratch. An important part of the process is understanding the system performance and safety requirements and constraints and any potential conflicts that must be resolved. A STAMP-based risk analysis involves creating:

- a model of the organizational safety structure, including the static safety control structure and the safety constraints that each component is responsible for maintaining;
- a model of the dynamics and pressures that can lead to degradation of this structure over time;
- the process (mental) models required by those controlling it and the feedback and communication requirements for maintaining accurate process models;
- a model of the cultural and political context in which decision making occurs.

We then apply a set of factors we have identified that can lead to violation of safety constraints, such as inadequate feedback, to maintain accurate mental (process) models. The information that results from this modeling and analysis effort can be used to assess the risk in both the current organizational culture and structure and in potential changes, to devise policies and changes that can decrease risk and evaluate their implications with respect to other important goals, and to create metrics and other performance measures and leading indicators to identify when risk is increasing to unacceptable levels. Because the models used have a mathematical foundation, simulation and mathematical analysis are possible.

The practicality of the approach has been demonstrated by applying it to a number of real and complex systems, including a risk analysis of the organizational structure of the Space Shuttle program after the Columbia loss (Leveson et al. 2005); tradeoffs among safety, budget, schedule, and performance risks in the new NASA space exploration mission organization (Dulac et al. 2007); unmanned spacecraft design (Owens et al. 2008); a safety assessment of the new US missile defense system; safety in the pharmaceutical industry; and safety of out-patient surgery at Boston's Beth Israel Deaconess Hospital (Dierks et al. 2008).

As an example of the use of a systems approach at the organizational and cultural level, we performed a risk analysis of a proposed new organizational structure for safety-related decisions in the Space Shuttle program after the *Columbia* loss. In this analysis, we identified the NASA organizational requirements to reduce poor engineering and management decision making leading to an accident, identified gaps and omissions in the new organizational design, and performed a rigorous programmatic risk analysis to evaluate the proposed policy and structure changes and to identify leading indicators and metrics of migration toward states of unacceptable risk over time. In a second application of the approach, to the new NASA Space Exploration Mission (to return humans to the Moon and go on to Mars), we demonstrated how tradeoffs among safety, performance, schedule, and budget can be evaluated. The analysis included the entire socio-technical system from Congress and the Executive Branch down to engineering processes and management. In this effort, we found, for example, that attempting to speed up development resulted in surprisingly little improvement in schedule (less than 2%) primarily because of resulting increases in rework, but the attempted schedule reduction had a very high negative impact on the safety of the resulting design. At the same time, early emphasis on safety led to improvements in both schedule and budget, due, again, to fewer required changes and rework when problems are discovered late. Although this result is probably not surprising to safety engineers, it was a surprise to managers, who found the mathematical analysis of the differences and rationale for evaluating alternatives to be very compelling.

9. Conclusions

The two prevailing organizational approaches to safety, Normal Accidents and HROs, have made important contributions to theory by focusing attention on a variety of industries that deal with hazardous situations, by developing concepts such as complexity and coupling, and by focusing attention on the role of organizational factors and safety culture in accidents. Yet both approaches limit the progress that can be made toward achieving highly safe systems by too narrowly defining the problem and the potential solutions. In this paper we have outlined some of the limitations of these approaches: for example, overly pessimistic or optimistic conclusions, confusion of reliability and safety, and ambiguity about some key concepts. We then described an alternative approach, based on systems theory, that we believe can provide more powerful ways to manage and control postmodern risk in complex, high-tech systems with their potential for catastrophic disruptions and losses.

Our approach offers new directions for both organization theory and safety management. Sociologists and engineers need to be working more closely together with more shared definitions and assumptions. We believe that more systematic and integrative theory will emerge from such efforts (despite their difficulties). In particular, the opportunities for multi-level theory, connecting institutional, organizational, group, and individual actions in a systems approach, seem very rich. For empirical research and the practice of system safety, our work offers new measures (including ways to design leading indicators) and a rich set of analytic techniques, including system dynamics modeling. Some particularly fertile opportunities for research could be found in documenting the safety constraints and organizational practices of organizations such as aircraft carriers, air traffic control, and hospitals as they change over time in response to changing environments (such as wartime and peacetime), new technologies, and varied regulations. Differences across nations and industries can be analyzed more systematically. Researchers can be involved in helping design and evaluate the changes being undertaken. In our postmodern world, researchers must rise to these challenges.

Notes

[1] Perrow defined *system accidents* as those caused by unanticipated interactions of failures. He differentiated these from *component failure accidents* caused by failures linked in anticipated sequences. The only difference between these two types of accidents, according to his definition, is that one is anticipatable and the other is not.

[2] Note that although Perrow considered these two properties to be independent, in engineering they are not. Perrow defined *complex interactions* as those arising from unfamiliar sequences or from unplanned and unexpected sequences that are either not visible or not immediately comprehensible. Perrow does not provide a definition of coupling, but in engineering *coupling* is usually defined in terms of degree and type of interdependence among system components, that is, whether and how the behavior of one component can impact the behavior of other components. Using these definitions, it can be seen that coupling is a system design feature that leads to interactive and other types of complexity. The amount and type of coupling (interdependencies among components) determines the amount and types of complexity of the component interactions during system operation and hence the visibility and comprehensibility of the component interactions.

[3] Note that Perrow's argument about risk being increased by interactive complexity and coupling is understood in engineering and reflected in general engineering practice and design for safety by the use of methods that reduce complexity and coupling.

[4] Note that these definitions omit the concept of *loss* or the consequences of the events beyond the physical components of the system. In particular, human death and injury are not included. Thus, if hundreds of people die but no parts of the system itself are damaged, then an accident has not occurred. Clearly this definition does not match common understanding of what an accident entails, nor the engineering definition (which is usually defined as an unplanned and unacceptable loss; Leveson 1995).

References

Albee, Arden et al.
2000 'Report on the loss of the Mars Polar Lander and Deep Space 2 Missions'. JPL Special Review Board, *JPLD-18709*, 29 March.

Beck, Ulrich
1992 *Risk society: Towards a new modernity*. Los Angeles: Sage.

Checkland, Peter
1981 *Systems thinking, Systems practice*. New York: John Wiley and Sons.

Clarke Lee
1993 'Drs. Pangloss and Strangelove Meet Organizational Theory: High Reliability Organizations and Nuclear Weapons Accidents', *Sociological Forum*, 8:675–689.

Conant, R. C., and W. R. Ashby
1970 'Every good regulator of a system must be a model of that system'. *International Journal of System Science* 1: 89–97.

Dekker, Sydney W. A.
2005 *Ten questions about human error: A new view of human factors and system safety*. Mahwah, NJ: Lawrence Erlbaum Associates.

Dierks, Meghan M., Nicolas Dulac, Nancy Leveson, Margaret Stringfellow
2008 'System dynamics approach to modeling risk in complex healthcare settings'. *Proceedings of the System Dynamics Conference*. Athens.

Dulac, Nicolas; Brandon D. Owens, and Nancy G. Leveson
2007 'Modeling risk management in the development of space exploration systems'. International Association for the Advancement of Space Safety Conference, Chicago, May 2007.

Hollnagel, Erik
2004 *Barriers and accident prevention*. Hampshire, England: Ashgate.

Kasperson, Roger
1986 'Six propositions on public participation and their relevance for risk communication'. *Risk Analysis* 6: 275–281.

La Porte, Todd R.
1996 'High reliability organizations: Unlikely, demanding, and at risk'. *Journal of Contingencies and Crisis Management* 63/4.

La Porte, Todd R., and Paula Consolini
1991 'Working in practice but not in theory: Theoretical challenges of high-reliability organizations'. *Journal of Public Administration Research and Theory* 1: 19–47.

Leplat, Jacques
1987 'Occupational accident research and systems approach' in *New technology and human error*. Jens Rasmussen, Keith Duncan, and Jacques Leplat (eds), 181–191. New York: John Wiley & Sons.

Leveson, Nancy G.
1995 *Safeware: System safety and computers*. Addison-Wesley.

Leveson, Nancy G.
2004 'New accident model for engineering safer systems'. *Safety Science* 42/4: 237–270.

Leveson, Nancy G., Nicolas Dulac, Joel Cutcher-Gershenfeld, Betty Barrett, John Carroll, David Zipkin, and Stephan Friedenthal
2005 'Modeling, analyzing, and engineering safety culture'. First International Conference of the Association for the Advancement of Space Safety, Nice, October 2005.

Owens, Brandon D., Margaret Herring, Nicolas Dulac, Nancy G Leveson, Michel Ingham, and Kathryn Weiss
2008 'Application of a safety-driven design methodology to an outer planet exploration mission'. IEEE Aerospace Conference, Big Sky, Montana, March 2008.

Perrow, Charles
1999 *Normal accidents: Living with high-risk technologies*. Princeton, NJ: Princeton University Press.

Rasmussen, Jens
1997 'Risk management in a dynamic society: A modelling problem'. *Safety Science* 27/2–3: 183–213.

Roberts, H. Karlene
1990a 'Managing high reliability organizations'. *California Management Review* 32/4: 101–114.

Roberts, H. Karlene
1990b 'Some characteristics of one type of high reliability organization'. *Organization Science* 1/2: 160–176.

Rochlin, Gene I., Todd R. La Porte, and Karlene H. Roberts
1987 'The self-designing high reliability organization'. *Naval War College Review*, Autumn.

Rogers, William P.
1986 *Report of the Presidential Commission on the space shuttle Challenger accident*. Washington DC: Government Printing Office.

Sagan, Scott D.
1995 *The limits of safety*. Princeton, NJ: Princeton University Press.

Sterman, J. John
2000 *Business dynamics: Systems thinking and modeling for a complex world.* New York: McGraw-Hill.

Turner, Barry A.
1978 *Man-made disaster*. London: Wykeham.

Weick, Karl E.
1987 'Organizational culture as a source of high reliability'. *California Management Review* 29/2: 112–127.

Weick, Karl E., and Karlene H. Roberts
1993 'Collective mind in organizations: Heedful interrelating on flight decks'. *Administrative Science Quarterly* 38/3: 357–381.

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

Woods, David D., and Richard I. Cook
2002 'Nine steps to move forward from error'. *Cognition, Technology, and Work* 4/2: 137–144.

Nancy G. Leveson

Nancy G. Leveson is Professor of Aeronautics and Astronautics, and Professor of Engineering Systems at MIT. She taught previously at the University of California and the University of Washington. She received a BA in Mathematics, an MS in Management, and PhD in Computer Science from UCLA. Dr. Leveson is an elected member of the National Academy of Engineering. Her research has focused on system engineering, system safety, accident analysis, human-computer interaction, software engineering, safety culture, risk management, security, and systems thinking in aerospace, defense, nuclear power, transportation systems, healthcare, pharmaceuticals, and most other safety-critical systems as well as social problems such as food safety, corporate fraud, and financial risk.
Address: MIT, 77 Massachusetts Avenue, Cambridge, MA 02139, USA.
Email: leveson.nancy@gmail.com

Nicolas Dulac

Nicolas Dulac is a Strategy Consultant at the Boston Consulting Group. He received a B.Eng. in Mechanical Engineering from McGill University, and a SM. and PhD. from the Department of Aeronautics and Astronautics at MIT. His graduate and postdoctoral research work focused on the application of system dynamics and systems theory to increase the safety of socio-technical systems such as the US space exploration program, the pharmaceutical and food supply industries, and inpatient surgical units.
Address: Exchange Place, 31st Floor, Boston, Massachusetts MA 02109, USA.
Email: ndulac@gmail.com

Dr. Karen Marais

Dr. Karen Marais is an Assistant Professor in the School of Aeronautics and Astronautics at Purdue University. Her research interests include safety analysis and risk assessment of complex socio-technical systems in general, and aerospace systems in particular. In addition, Dr. Marais conducts research on the environmental impact of aviation within the FAA PARTNER Centre of Excellence, and is currently growing her research platform to include maintenance and operations of engineering systems. She received her PhD. from the Department of Aeronautics and Astronautics at MIT in 2005. She also holds a Master's degree in Space-based Radar from MIT. Prior to graduate school, she worked in South Africa as an electronic engineer. She holds a B.Eng. in Electrical and Electronic Engineering from the University of Stellenbosch and a BSc in Mathematics from the University of South Africa. Dr Marais is the author or co-author of some twenty technical publications including six journal publications and one book chapter. She is currently working on an introductory textbook on reliability and risk analysis, to be published by Cambridge University Press.
Address: School of Aeronautics and Astronautics, Purdue University, West Lafayette, Indiana IN 47907, USA.
Email: kmarais@purdue.edu

John S. Carroll

John S. Carroll is Morris A. Adelman Professor of Management at the MIT Sloan School of Management and the MIT Engineering Systems Division and Co-Director of the MIT Lean Advancement Initiative. He taught previously at Carnegie-Mellon University, Loyola University of Chicago, and the University of Chicago. He received a BS in Physics from MIT and a PhD in Social Psychology from Harvard. His research has focused on decision-making and learning in organizational contexts. Current projects examine organizational safety issues in high-hazard industries such as nuclear power, aerospace, and health care, including self-analysis and organizational learning, safety culture, leadership, communication, and systems thinking.
Address: MIT Sloan School of Management, 50 Memorial Drive, E52-536, Cambridge, MA 02142, USA.
Email: jcarroll@mit.edu