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

Nicolas Dulac, Karen Marais, Nancy Leveson
Massachusetts Institute of Technology
2011

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) [Perrow84, Sagan] and High Reliability Organizations (HRO) [LaPorte96, LaPorte-Consolini91, Roberts90a, Roberts90b, Rochlin87, Weick87, Weick93, Weick99]. In this paper, we argue that the conclusions of HRO researchers 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 even 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 is described, which avoids the limitations of NAT and HRO. While this paper uses many aerospace examples, particularly the Columbia accident, the conclusions apply to most 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 [Perrow79]. Perrow introduced the idea that in some technological systems, accidents are inevitable or "normal" [Perrow84]. 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 potential for unfamiliar or unplanned and unexpected interactions between system components, leading to sequences of events 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. As can be seen from the definitions, tightly coupled system design leads to interactive complexity.

According to the NAT 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[1] can interact in unpredictable ways and the propagation of effects resulting from tight coupling can quickly spiral out of control. In these *system accidents*, events that could be easily contained and handled in a loosely coupled system, escalate quickly before operators are able to understand the situation and perform appropriate corrective actions.

Perrow made an important contribution in identifying these risk-increasing system characteristics. His conclusion, however, that accidents are inevitable in these systems and therefore systems for which

[1] Although Perrow focused on accidents resulting from failures, accidents can also result from non-failed, i.e., operational, system components interacting in unsafe ways, as captured in the systems approach to safety described later in this paper.

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 tightly coupled, interactively complex 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 [Safeware]. 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 [Safeware].

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. Tightly coupled, interactively complex 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 eliminate risk or to reduce it by making appropriate tradeoff decisions.

3. High Reliability Organizations

Some researchers have described a type of organization they call High Reliability, although the definition of this type of organization is unclear. Roberts [Roberts90a} 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*." [p.160]{Roberts90a}.

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 proper 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, 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 [LaPorte-Consolini91].

The most important flaw in this argument is that the systems they studied are not interactively complex nor are they 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 *de-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) 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 seems to support Perrow's arguments rather than contradict them, and they provide little assistance in understanding how to prevent accidents in systems that *are* tightly coupled.

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." [pp.29-30]{LaPorte-Consolini91} (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 *cannot* be thoroughly planned, understood, predicted, or guarded against. HROs are not interactively complex by the HRO researchers own descriptions. Therefore, 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. The real challenge is to reduce risk in today's tightly coupled, interactively complex high-tech and social systems where technical knowledge is not complete.

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 had 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.[2]

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.

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, 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 [LaPorte-Consolini91, Roberts90b], 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, has 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 and, in fact, some may even contribute to accidents in these 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

[2] 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.

of both safety and performance and consensus about the goals across the organization [LaPorte-Consolini91]; (2) promotion of a ``culture of reliability'' in simultaneously decentralized and centralized operations [Weick87]; (3) use of organization learning that maximizes learning from accidents, incidents, and near misses [LaPorte-Consolini91]; and (4) extensive use of redundancy [Rochlin87].

The problem is not that these characteristics are bad, but instead lies in the oversimplification of what are actually extremely difficult and complex engineering and management problems and the confusion between reliability and safety. The next four sections consider the practicality and usefulness of using each of these characteristics in 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 [LaPorte-Consolini91]. 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 the primary reason for 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 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. 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 shoot down of an Iranian commercial aircraft by the *USS Vincennes* indicate that combat conditions have a strong effect on high reliability performance.

For most organizations, safety is not their primary goal. 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” (effectively International Space Station completion) 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 than 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" [Barrett].

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 (i.e., implement these goals reliably) overly simplifies the difficulty in accomplishing the 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 [Barrett].

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 [Weick87]. 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 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.

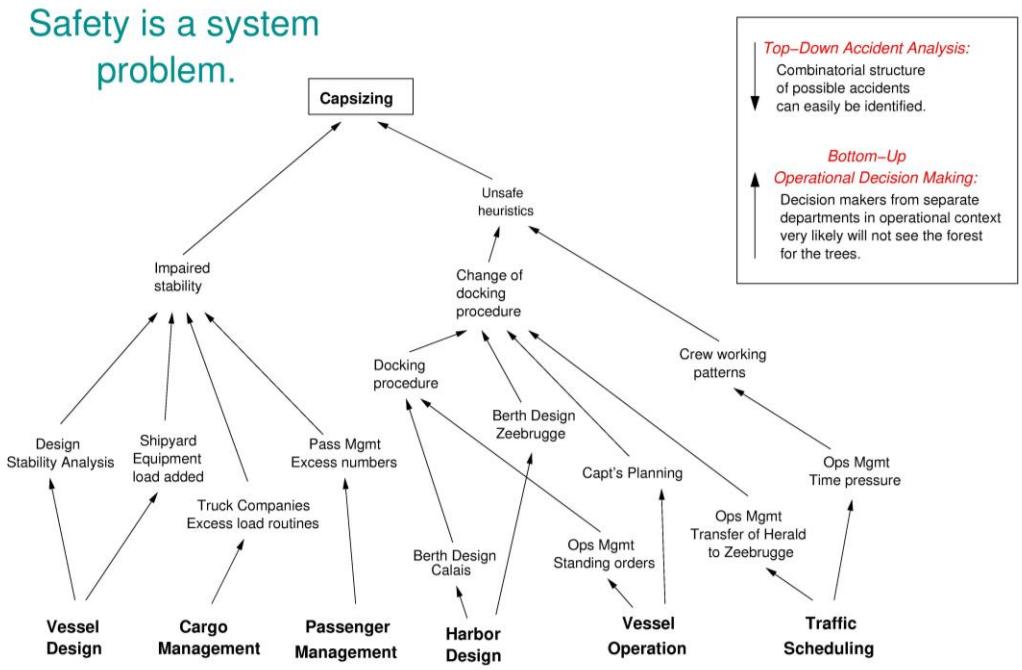

Figure 1. The Complex Interactions in the Zeebrugge Ferry Accident
(Adapted from Rasmussen, Safety Science, p.188)

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 [Rasmussen]. 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 (and has) led 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 [Weick87]. 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,'' and therefore usually more complete. In fact, some serious accidents have been the result of the substitution of incomplete and non-rigorous techniques for more expensive but more complete hazard analysis methods.

Organizational learning requires more sophisticated accident and incident analysis techniques than are currently used. Often a single "root cause" is hypothesized, usually involving a mechanical failure or human error. In-depth analysis of the systemic causes of accidents is required for effective organizational learning to occur. The difficulty of accomplishing this goal should not be underestimated. The use of sophisticated hazard analysis, design for safety, and safety assurance techniques that include the organizational and social factors in accidents and treat the entire socio-technical system will be more effective for complex systems than relying on *ad hoc* scenario generation. How this might be accomplished for organizational aspects of safety is outlined in Section 5 on a system's approach to safety.

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," [LaPorte96] where redundancy is defined as "`the ability to provide for the execution of a task if the primary unit fails or falters" [LaPorte-Consolini91]. 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) [Roberts90b]. 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 argues.

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. Perrow and Leveson have identified large numbers of accidents caused by common-cause failures.

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 unsafe 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 the context in which they are made, but not when combined at the system level. Redundancy will not help here.

The reliance on redundancy to ensure safety can lead to dangerous decision making and discounting of the need for additional safety measures. 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 [Safeware]. 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 [Safeware].

Redundancy is only one limited way to increase reliability (but not necessarily safety) in some special cases as noted in Section 2; 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 accidents. But the NAT 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 aspects, 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 [Rasmussen], Woods [Woods], and Leveson [Leveson-SafetyScience], most of whom come from systems engineering and human factors

backgrounds, 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.

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

In understanding the alternative systems approach to organizational safety, it is first necessary to recognize that reliability and safety are *different* properties and, in fact, are sometime conflicting. HRO researchers (as denoted even by the name) treat safety and reliability as equivalent. They 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 (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, 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 unsafe ways. As the interactive complexity grows in the systems we build, accidents caused by unsafe 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 [Safeware, Perrow]. At the same time, accidents have resulted precisely because the operators *did* follow the predetermined instructions provided to them in their training, such as at Three Mile Island. When the results of deviating from procedures are positive, operators are lauded but when the results are negative, they are punished for being unreliable.

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: To effectively deal with safety in complex systems requires simultaneous consideration of the 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. These factors must be considered from a top-down systems perspective that considers the interactions among the system components rather than a bottom-up component reliability approach. 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.

As noted earlier, system's approaches to safety have been proposed by several researchers. Leveson's STAMP (Systems-Theoretic Accident Modeling and Processes) approach goes the farthest toward a pure system's approach and differs significantly from all others with respect to assumptions about the causes of accidents and the analysis of social and cultural factors in accidents. STAMP is therefore used here to contrast between systems approaches and the HRO and NAT views of organizational aspects of safety.

5.1 Limitations of Traditional Accident Causality Models

One unique aspect of Leveson's approach is the use of a new, expanded model of causality. Traditionally in engineering and in organizational safety, accidents have been thought of being caused by chains (or trees, which are just a short hand for specifying multiple chains) of failure events and human errors, each directly related to the next. This model of accident causation forms the basis for almost all hazard analysis and for behavioral and organizational approaches to safety and influences the solutions that are proposed for preventing accidents.

If accidents are caused by chains of failure events, then the obvious way to prevent them is to "break" the chain by doing one or all of the following:

1. Make the individual failure events less likely by ensuring that the individual components are highly reliable and do not fail. In the case of humans as system components, the definition of failure is more difficult than often assumed, as discussed further below.
2. Add redundancy so that the events are less likely to lead to further failure events, or
3. Add barriers between the events so that bad effects are not generated further.

Note that the addition of a barrier simply creates a different potential failure chain that now includes the failure of a barrier. High reliability and redundancy are used to prevent barrier failure so that barrier approach does not add anything to the other two.

While this traditional accident causality model worked well for the relatively simple electro-mechanical systems that existed at the time it originated, this model does not handle social and managerial factors in accidents and losses, software and other new types of technological artifacts that do not fail in a random fashion (and perhaps do not "fail" at all in the typical definition of engineering failure), the types of complex human decision-making involved in many new systems, and accidents or losses caused by non-linear interactions among components rather than individual component failure

(Perrow's system accidents). There is also an implicit assumption that systems are static but all systems, especially social aspects, change and evolve over time.

NAT and HRO (and nearly all other safety research and practice, including some of the proposed system approaches to safety) accept the chain-of-events view of accidents. Perrow's pessimistic conclusions about the inevitability of accidents are correct if only this overly simple model of accident causality and thus accident prevention is considered. HRO, as noted above, takes a reliability view of safety, and their proposals for accident prevention do not apply to tightly coupled systems with high interactive complexity and with high technical, social, and political uncertainty. Effectively dealing with accidents in complex, socio-technical systems requires an expanded view of accident causation. Human error and system change over time present particular difficulties for the traditional causality models [Leveson-Engineering a Safer World].

5.1.1 Human Error and Decision-Making

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

Mental models play a significant role here. At higher levels of cognitive control, experimentation is needed for operators to update procedures to handle changing conditions or to evaluate hypotheses while engaged in symbolic reasoning about the best response to unexpected situations. Actions that are quite rational and important during the search for information and test of hypotheses may appear to be unacceptable mistakes in hindsight, without access to the many details of a turbulent situation.

The ability to adapt mental models through experience in interacting with the operating system is what makes the human operator so valuable (see Figure 2). Based on current inputs, the operators' actual behavior may differ from the prescribed procedures. When the deviation is correct (the designers' models are less accurate than the operators' models at that particular instant in time), then the operators are considered to be doing their job. When the operators' models are incorrect, they are often blamed for any unfortunate results, even though their incorrect actions may have been reasonable given the information they had at the time.

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

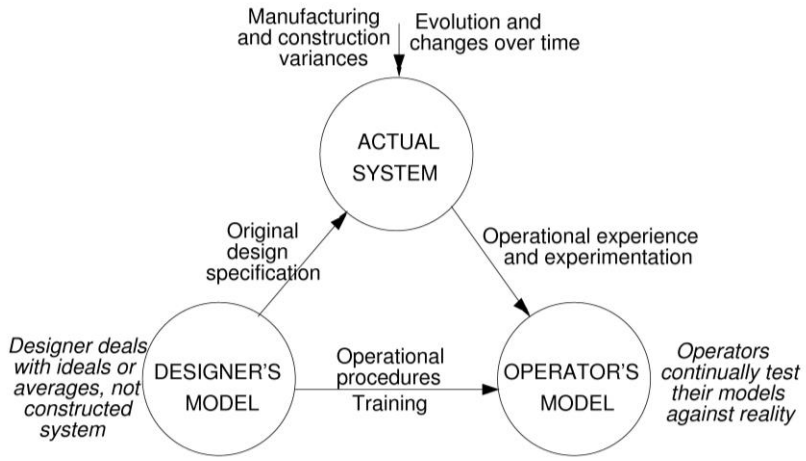

Figure 2. The role of mental models in operations. Designers deal with systems as ideals or averages and provide procedures to operators with respect to this ideal. Systems may deviate from the ideal through manufacturing and construction variances or through evolution over time. Operators must deal with the existing system and change their mental models (and operational procedures) using operational experience and experimentation.

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

Traditional decision theory research perceives decisions as discrete processes that can be separated from the context and studied as an isolated phenomenon. More recent research has taken a very different approach: Instead of thinking of operations as predefined sequences of actions, human interaction with a system is increasingly being considered to be a continuous control task in which separate "decisions" or errors are difficult to identify.

Edwards, in 1962, was one of the first to argue that decisions can only be understood as part of an ongoing process. The state of the system is perceived in terms of possible actions, one of these actions is chosen, and the resulting response from the controlled system acts as a background for the next action. Errors then are difficult to localize in the stream of behavior; the effects of less successful actions are a natural part of the search on the part of the operator for optimal performance.

Not only are separate decisions difficult to identify in this model of human control, but the study of decision making then cannot be separated from a simultaneous study of the social context, the value system in which it takes place, and the dynamic work process it is intended to control [Rasmussen]. This view is the foundation of *dynamic decision-making* and the new field of *naturalistic decision-making* [Klein, etc.] and is the approach taken in this research.

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

To accomplish this goal, a representation of human behavior at a higher level of functional abstraction than the level used for task analysis is needed. Such a representation must include the objectives of the individual actor in the actual context and the boundaries of acceptable performance, that is, the constraints on the work place. Rasmussen, Pejtersen, and Goodstein define one such framework for identifying the objectives, value structures, and subjective preferences governing behavior within the degrees of freedom faced by the individual decision maker and actor. Vicente also uses this framework for his Cognitive Work Analysis. In this approach, behavior is modeled directly in terms of the behavior-shaping constraints of the environment and the adaptive mechanisms of human actors in the environment. Leveson's STAMP takes a similar general approach but implements it differently, as described below.

5.1.2 Adaptation

Another important requirement for improved accident models is that they include adaptation. Systems and organizations continually experience change as adaptations are made in response to local pressures and short-term productivity and cost goals. People adapt to their environment or they change their environment to better suit their purposes. A corollary of this propensity for systems and people to adapt over time is that safety defenses are likely to degenerate systematically through time, particularly when pressure toward cost-effectiveness and increased productivity is the dominant element in decision-making. The critical factor here, as noted by Rasmussen, is that such adaptation is not a random process—it is an optimization process depending on search strategies—and thus should be predictable and potentially controllable.

For an accident model to handle system adaptation, it must consider the processes involved in accidents and not simply events and conditions: Processes control a sequence of events and describe system and human behavior as it changes and adapts over time rather than considering individual events and human actions. To talk about the cause or causes of an accident makes no sense in this view of accidents. Deterministic causal models are inadequate to explain the organizational and social factors in highly adaptive socio-technical systems. Instead, accident causation must be viewed as a complex process involving the entire socio-technical system including legislators, government agencies, industry associations and insurance companies, company management, technical and engineering personnel, operators, etc.

STAMP: An Expanded View of Accident Causality

In Leveson's STAMP (System-Theoretic Accident Modeling and Processes) model of accident causation, the traditional conception of accidents as chains (or trees) of directly related failure events and human errors is expanded. 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. STAMP allows more complex relationships among events (in addition to the simple relationships of the traditional causality models) to be considered, It 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 [Leplat]. 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.

The theoretical foundation of STAMP is systems theory, rather than reliability theory. Systems theory is the mathematical foundation for systems engineering, with roots that go back to the 1930's, especially von Bertalanffy in biology. It was a response to limitations of the classic analysis techniques in coping with increasingly complex systems [Checkland]. 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 [Ramo]. These system properties derive from the relationships between the parts of systems: how the parts interact and fit together [Ackoff]. Thus, the systems approach concentrates on the analysis and design of the whole as distinct from the parts.

STAMP attempts to integrate all aspects of risk, including organizational, social, and technical. Systems are viewed in STAMP as interrelated components that are kept in a state of dynamic equilibrium by feedback loops of information and control. A socio-technical system is not treated as just a static design, but as a dynamic process that is continually adapting to achieve its ends and to react to changes in itself and in its environment. The original design must not only enforce constraints on behavior to ensure safe operations, but it must continue to operate safely as changes and adaptations occur over time.

Accidents then are viewed as the result of flawed processes involving interactions among people, societal and organizational structures, engineering activities, and physical system components. The process leading up to an accident can be described in terms of an adaptive feedback function that fails to maintain safety as performance changes over time to meet a complex set of goals and values. The accident itself results not simply from component failure (which is treated as a symptom of the problems) but from inadequate control of safety-related constraints on the development, design, construction, and operation of the socio-technical system.

Safety in this model is treated as a *control* problem: Accidents occur when component failures, external disturbances, and/or unsafe interactions among system components are not adequately handled. In the Space Shuttle Challenger accident, for example, the O-rings did not adequately control the propellant gas release by sealing a tiny gap in the field joint. In the Mars Polar Lander loss, the software did not adequately control the descent speed of the spacecraft— it misinterpreted noise from a Hall effect sensor as an indication the spacecraft had reached the surface of the planet.

Accidents such as these, involving engineering design errors, may in turn stem from inadequate control of the development process, i.e., 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.

While events reflect the *effects* of unsafe interactions and inadequate enforcement of safety constraints, the inadequate control itself is only indirectly reflected by the events—the events are the *result* of the inadequate control. The control structure itself, therefore, must be carefully designed and evaluated to ensure that the controls are adequate to maintain the constraints on behavior necessary to control risk. 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.

STAMP is constructed from three fundamental concepts: constraints, hierarchical levels of control, and process models. These concepts, in turn, give rise to a classification of control flaws that can lead to accidents. Each of these is described only briefly here.

Constraints

The most basic component of STAMP is not an event, but a *constraint*. 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 theory, systems are viewed as hierarchical structures where each level imposes constraints on the activity of the level below it—that is, constraints or lack of constraints at a higher level allow or control lower-level behavior. Safety-related constraints specify those relationships among system variables that constitute the non-hazardous or safe system states—for example, the power must never be on when the access to the high-voltage power source is open, the descent engines on the lander must remain on until the spacecraft reaches the planet surface, and two aircraft must never violate minimum separation requirements.

Instead of viewing accidents as the result of an initiating (root cause) event in a chain of events leading to a loss, accidents are viewed as resulting from interactions among components that violate the system safety constraints. The control processes that enforce these constraints must limit system behavior to the safe changes and adaptations implied by the constraints. Accidents result when they do not.

Hierarchical Control Structures

The second basic construct in STAMP is hierarchical control structures. Preventing accidents requires designing a control structure, encompassing the entire socio-technical system, that will enforce the necessary constraints on development and operations. Figure 3 shows a generic hierarchical safety 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. Feedback during operations is critical here. For example, the safety analysis process that generates constraints always involves some basic assumptions about the operating environment of the process. When the environment changes such that those assumptions are no longer true, the controls in place may become inadequate.

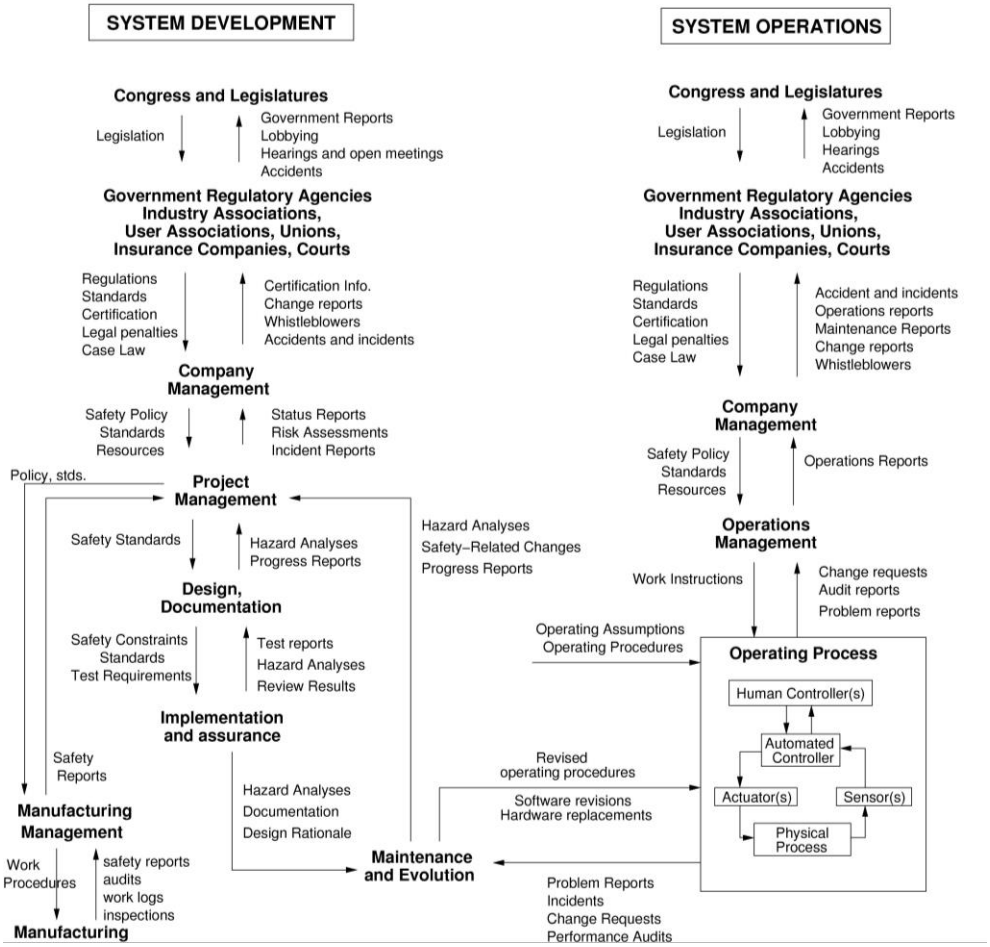

Figure 3. General Form of a Model of Socio-Technical Control.

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. 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. As an example, Figure 4 shows the changes over time that led to a water contamination accident in Canada where 2400 people became ill and 7 died (most of them children). Only a small part of the overall STAMP model is shown (see [24] for the complete model). Each component of the water quality control structure played a role in the accident.

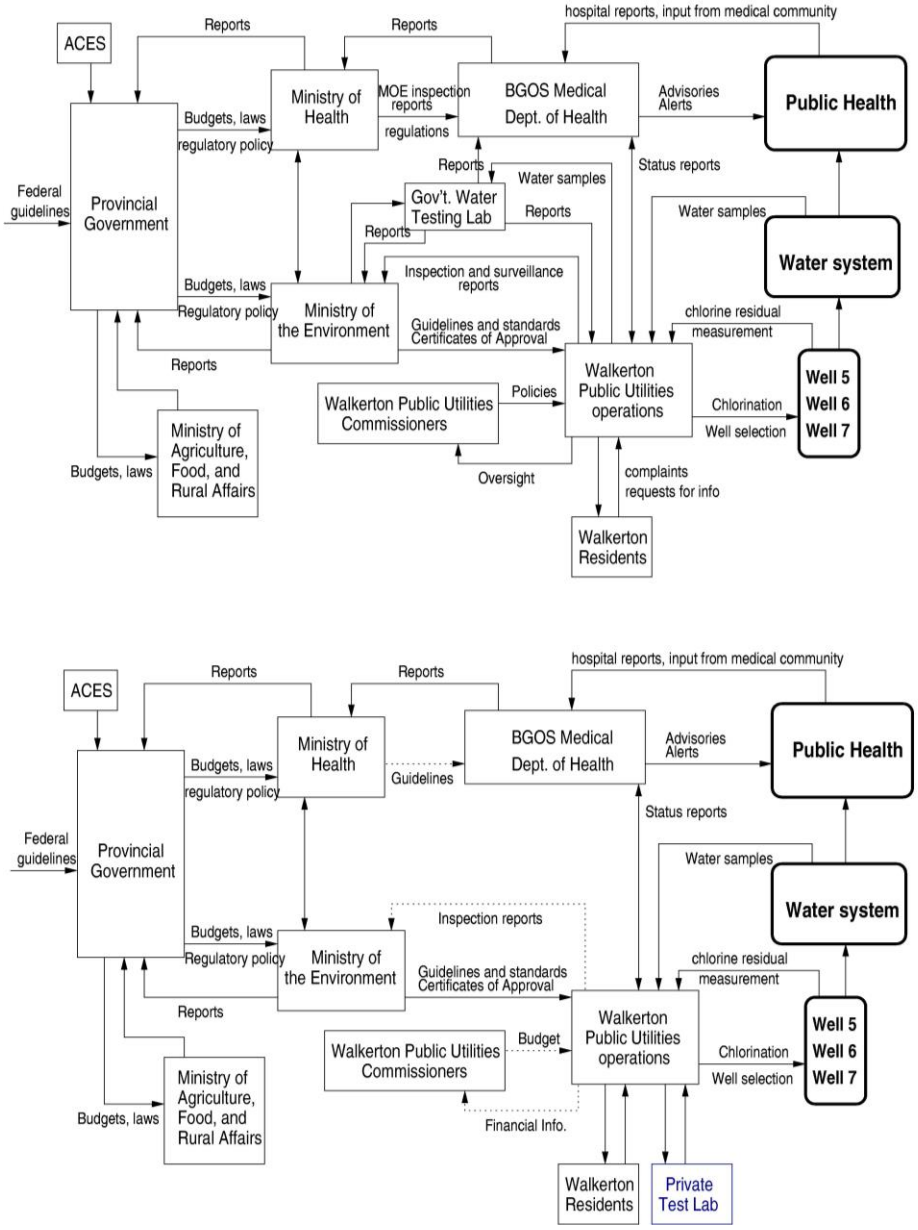

Figure 4. The Safety Control Structure in the Walkerton Water Contamination Accident. Note: The structure is drawn in the form commonly used for control loops. Lines going into the left of a box are control lines. Lines from or to the top or bottom
of a box represent information, feedback, or a physical flow. Rectangles with sharp corners are controllers while rectangles with rounded corners represent controlled physical processes.

The model at the top shows the control structure for water quality in Ontario Canada as designed. The figure at the bottom shows the control structure as it existed at the time of the accident. One of the important changes that contributed to the accident was the elimination of a government water-testing laboratory. The private companies that were substituted were not required to report instances of bacterial contamination to the appropriate government ministries. Essentially, the elimination of the feedback loops made it impossible for the

government agencies and public utility managers to perform their oversight duties effectively. Note that the goal here is not to identify individuals to blame for the accident but to understand why they made the mistakes they made (none were evil or wanted children to die) and what changes are needed in the culture and water quality control structure to reduce risk in the future.

Often degradation of the control structure involves *asynchronous evolution* where one part of a system changes without the related necessary changes in other parts. Changes to subsystems may be carefully designed, but consideration of their effects on other parts of the system, including the control aspects, may be neglected or inadequate. Asynchronous evolution may also occur when one part of a properly designed system deteriorates. Modeling tools and analysis tools built on STAMP can be used to help decision makers understand the effects of asynchronous changes to the control structure and thus assist in evaluating the impact on risk of potential organizational changes and policy decisions—to act as a proverbial "canary in the coal mine" in detecting unsafe change that increases risk to unacceptable levels.

Figure 4 shows static models 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* models can be used for this goal. The field of system dynamics, created at MIT in the 1950s by Forrester, is designed to help decision makers learn about the structure and dynamics of complex systems, to design high leverage policies for sustained improvement, and to catalyze successful implementation and change. System dynamics provides a framework for dealing with dynamic complexity, where cause and effect are not obviously related. Like the other STAMP models, it is grounded in the theory of non-linear dynamics and feedback control, but also draws on cognitive and social psychology, organization theory, economics, and other social sciences [37]. Traditional system dynamics models lack the basic structural components of the STAMP hierarchical control structure model—combining these structural and dynamic models has the potential to provide more complete information about the system than either alone and therefore provide the greater range of modeling power necessary to include organizational, social, and individual behavior in the modeling and understanding of organizational safety and risk [Dulac Diss, System Dynamics 2007 paper].

Figure 5 shows a simplified system dynamics model of causation in the Columbia Space Shuttle loss. A complete description of system dynamics modeling is beyond the scope of this paper, but briefly, the model is composed of interacting control loops. The arrows connect variables that can affect other variables and the + or – show the polarity of the relationship, with + denoting that the two variables move in the same direction and – representing variables that move in opposite directions.

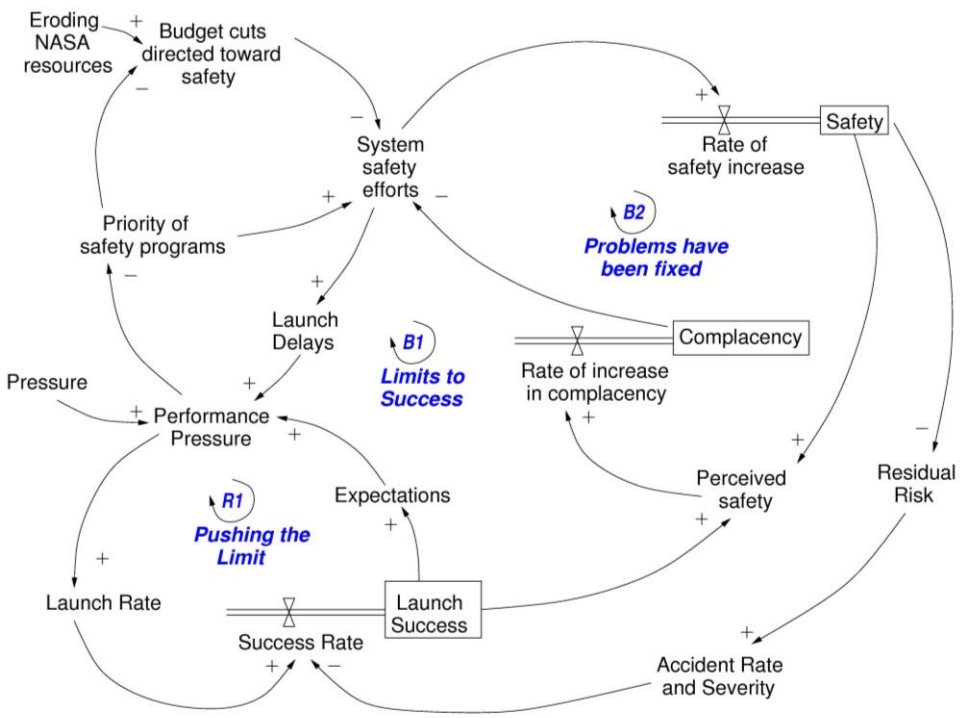

Figure 5. A Simplified System Dynamics Model of Causal Factors in the Space Shuttle Columbia loss.

The control loop in the lower left corner of Figure 5, labeled R1 or *Pushing the Limit*, shows how as external pressures increased, performance pressure increased, which led to increased launch rates and thus success in meeting the launch rate expectations, which in turn led to increased expectations and increasing performance pressures. This, of course, is an unstable system and cannot be maintained indefinitely—note the larger control loop, B1, in which this loop is embedded, is labeled *Limits to Success*. The upper left loop represents part of the safety program loop. The external influences of budget cuts and increasing performance pressures that reduced the priority of safety procedures led to a decrease in system safety efforts. The combination of this decrease along with loop B2 in which fixing problems increased complacency, which also contributed to reduction of system safety efforts, eventually led to a situation of (unrecognized) high risk. There is one other important factor shown in the model: increasing system safety efforts led to launch delays, another reason for reduction in priority of the safety efforts in the face of increasing launch pressures.

One thing not shown in the diagram is the delays in the system. While reduction in safety efforts and lower prioritization of safety concerns may lead to accidents, accidents usually do not occur for a while so false confidence is created that the reductions are having no impact on safety and therefore pressures increase to reduce the efforts and priority even further as the external performance pressures mount.

Combined STAMP structural models and system dynamics models can be used to devise and validate fixes for the problems and to design systems with lower risk [Dulac Diss]. For example, one way to eliminate the instability of the model in Figure 5 is to anchor the safety efforts by, perhaps, externally enforcing standards in order to prevent schedule and budget pressures from leading to reductions in the

safety program. Other solutions are also possible. Alternatives can be evaluated for their potential effects and impact on risk using a more complete models. This analysis procedure has been applied to real systems [ITA report, ESMD report].

Process Models

Besides constraints and hierarchical levels of control, a third basic concept in STAMP is that of process models. *Any* controller—human or automated—must contain a model of the system being controlled [5]. For humans, this model is commonly known as a mental model. Figure 6 shows a typical control loop where an automated controller is supervised by a human controller.

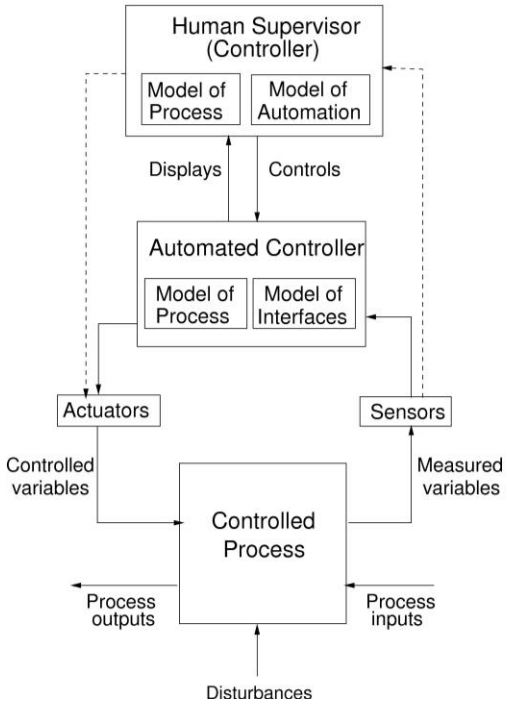

Figure 6. Classic Control Loop with Process Models Denoted

Automated controllers must have a model of the process being controlled and a model of its interfaces with other parts of the system. A human operator or supervisor also needs a model of the process being controlled, a model of the automation (if any), and a model of the interfaces. These process models must contain (1) the current state of the system being controlled, (2) the required relationships between system variables, and (3) the ways the process can change state. The models also include information about the context in which decisions are made and the influences and pressures that can lead to poor risk-related decision-making.

Accidents, 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 STAMP involves creating the process models, examining the ways that 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.

Note that no assumptions or hypotheses are being made about how the human mind works, only the information needed to make appropriate risk-related decisions. Using this knowledge, it is possible to design organizational structures and communication channels that provide decision-makers with the information they need for decision-making. When there are multiple controllers and decision makers, i.e., distributed decision-making, accidents may involve unexpected side effects of decisions or actions and conflicts between independently made decisions, 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 [18]. 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.

5.3 Applying STAMP to Organizational Safety

To summarize, STAMP treats socio-technical systems 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 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. These processes involve non-linear, indirect, and feedback relationships among components and events and not just directly related chains of failure events. More specifically, accidents occur when the designed hierarchical safety control structure and control actions do not enforce the system safety constraints due to:

- Unhandled environmental disturbances or conditions
- Unhandled or uncontrolled component failures
- Unsafe interactions among components
- The hierarchical safety control structure degrades over time (asynchronous evolution) and thus provides inadequate enforcement of system safety constraints
- Control actions are inadequately coordinated among multiple controllers of the same process or of the boundaries between processes.

With this model, understanding and preventing accidents in complex systems requires:

- Identifying the system safety constraints necessary to prevent accidents;
- Designing the system (the organization or socio-technical system being considered) to enforce the constraints, including understanding how the safety constraints could be violated and building in protections against these 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 to safety built on a formal foundation creates the possibility of formally modeling and engineering the safety culture and organizational aspects of safety, including the

entire socio-technical system—physical, organizational, and social. The models can be simulated and analyzed and the results used for enhanced decision-making about risk [ITA, ESMD]. For example, 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, process models representing the view of the process by those controlling it, and 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 and a new STAMP-based hazard analysis method called STPA . 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 to identify when risk is increasing to unacceptable levels.

The practicality of the approach has been demonstrated by applying it to a number of real systems, including a risk analysis of the organizational structure of the Space Shuttle program; tradeoffs among safety, budget, schedule, and performance risks in the new NASA space exploration mission organization; safety in the pharmaceutical industry; and safety of out-patient surgery at Boston's Beth Israel Deaconess Hospital. Other examples and demonstrations are currently in progress.

6. Conclusions

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. One particular alternative approach was described in this paper, but others are possible and should be investigated.

6. References

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

{Barrett} Barrett, B. Personal Communication, March 2004.

{Carroll} 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.

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

{CAIB} Gehman, Harold W. (Chairman). Columbia Accident Investigation Board Report, Volume 1. NASA and GAO, August 2003.

{LaPorte96} La Porte, Todd R. High Reliability Organizations: Unlikely, Demanding, and At Risk. *Journal of Contingencies and Crisis Management*, 63(4), 1996.

{LaPorte-Consolini91} 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.

{LaPorte-Rochlin96} La Porte, Todd R. and Rochlin, Gene. A Rejoinder to Perrow. *Journal of Contingencies and Crisis Management*, 2(4), 1994.

{Leplat} 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.

{Safeware} Leveson, Nancy G. *Safeware: System Safety and Computers*. Addison-Wesley, 1995.

{Leveson-SafetyScience} Leveson, Nancy G. A New Accident Model for Engineering Safer Systems. *Safety Science*, 42(4), April 2004, pp. 237–270.

{Walkerton} 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.

{Perrow79} 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.

{Perrow84} Perrow, Charles. *Normal Accidents: Living with High-Risk Technologies*. Princeton University Press, 1999.

{Ramo-268} 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.

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

{Rasmussen-Svedung} Rasmussen, Jens, and Svedung, Inge. *Proactive Risk Management in a Dynamic Society*. Swedish Rescue Services Agency, 2000.

{Roberts90a} Roberts, K.H (1990a). Managing high reliability organizations. *California Management Review*, 32(4), pp. 101–114.

{Roberts90b} Roberts, K.H. (1990b) Some characteristics of one type of high reliability organization. *Organization Science*, 1(2), pp. 160–176.

{Rochlin91} 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.

{Rochlin87} Rochlin, Gene I., La Porte, Todd R., and Roberts, Karlene H. The Self-Designing High Reliability Organization. *Naval War College Review*, Autumn, 1987.

{Sagan} Sagan, Scott. *The Limits of Safety*. Princeton University Press, 1995.

{Weick87} Weick, Karl E. Organizational Culture as a Source of High Reliability. *California Management Review*, 29(2), Winter 1987, pp. 112–127.

{Weick93} 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.

{Weick99} Weick, Karl E., Sutcliffe, K., and Obstfeld, D. Organizing for High Reliability. *Research in Organizational Behavior*, 21, 1999, pp. 81–123.

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