MIT Open Access Articles

The Need for New Paradigms in Safety Engineering

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Leveson, N. G. "The Need for New Paradigms in Safety Engineering." Safety-Critical Systems: Problems, Process and Practice (2009): 3-20. Proceedings of the Seventeenth Safety-Critical Systems Symposium, Brighton, UK, 3-5 February 2009

As Published: http://dx.doi.org/10.1007/978-1-84882-349-5_1

Publisher: Springer-Verlag

Persistent URL: <http://hdl.handle.net/1721.1/59813>

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of Use: Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.

The Need for New Paradigms in Safety Engineering

Nancy G. Leveson
Massachusetts Institute of Technology

Introduction

Most of the safety engineering techniques and tools we use today were originally created for first mechanical and later electro-mechanical systems. They rest on models of accident causation that were appropriate for those types of systems, but not the majority of the systems we are building today. After computers and other new technology became important in most new systems, the primary approach to handling safety was to try to extend the traditional techniques and tools to include software. We have now attempted that for at least three decades with little real success. I believe that it is time to conclude that this approach may not lead to great success and that something else is needed [11, 13].

Software allows us to increase the complexity of the systems we build (in particular, *interactive* complexity and coupling) such that new types of accidents are occurring that do not fit the traditional accident causation model. These new accidents arise not from the failure of individual system components, but from dysfunctional interactions among components, none of which may have failed, i.e., they operated as specified in their requirements. The loss of the Mars Polar Lander was attributed to noise (spurious signals) generated when the landing legs were deployed during descent [6]. This noise was normal and expected and did not represent a failure in the landing leg system. The onboard software interpreted these signals as an indication that landing occurred (which the software engineers were told they would indicate) and shut the engines down prematurely, causing the spacecraft to crash into the Mars surface. The landing legs and the software performed correctly with respect to their specified requirements but the accident occurred because the system designers did not account for all interactions between the leg deployment and the descent-engine control software.

A model of accident causation and the engineering techniques built on it that consider only component failures will miss system accidents, which are the most common software-related accidents. In addition, the role of human operators is changing from direct control to supervisory positions involving sophisticated decision-making. Once again, the types of mistakes humans are making are different and are not readily explained or handled by the traditional chain-of-failure-events models. Finally, there is more widespread recognition of the importance of management, organizational, and cultural factors in accidents and safety: the traditional models, which were never derived to handle these factors, do so poorly if at all.

I believe that to make significant progress in safety engineering, we need to rethink the old models and create new accident causality models and engineering techniques and tools based on them that include not only the old accident causes but also the new types of accidents and accident causality factors. In this paper, I suggest one such model and some tools based on it, but it is not the only such model possible and other tools and techniques might be built on it or on other models. Our new model is based on system theory (rather than the reliability theory of the traditional models) and our experience with it has shown that it allows much more powerful accident analysis and root cause analysis, hazard analysis, design-for-safety techniques, and general approaches to risk management in complex, socio-technical systems.

STAMP: An Accident Causality Model Based on System Theory

Traditional accident causation models explain accidents in terms of a chain of events that leads up to the accident. The relationships assumed between events in the chain are direct and relatively simple. Using this model of causation, the most appropriate approaches to preventing accidents is to somehow “break the chain” by either preventing an event or by adding additional “and” gates in the chain to make the occurrence of the events in the chain less likely. Because the events usually included almost always involve component failures or human errors, the primary mechanism for increasing safety is to make the individual components more reliable or failure free. Such models are limited in their ability to handle accidents in complex systems, organizational and managerial (social and cultural) factors in accidents, human error, and the systemic causes of the events.

For the past seven years, I have been developing a new, more comprehensive model of accident causation, called STAMP (System-Theoretic Accident Model and Processes), that includes the old models

but expands them to better handle the levels of complexity and technical innovation in today's systems [8, 11]. STAMP extends the types of accidents and causes that can be considered by including non-linear, indirect, and feedback relationships among events. Accidents or unacceptable losses can result not only from system component failures but also from interactions among system components—both physical and social—that violate system safety constraints.

In systems theory, emergent properties associated with a set of components are related to *constraints* upon the degree of freedom of those components' behavior. *Safety constraints* specify the 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; and the public health system must prevent the exposure of the public to contaminated water. Accidents result from interactions among system components that violate these constraints—in other words, from a lack of appropriate constraints on component and system behavior.

Major accidents rarely have a single root cause but result from an adaptive feedback function that fails to maintain safety as performance changes over time to meet a complex and changing set of goals and values. The accident or loss itself results not simply from component failure or human error (which are symptoms rather than root causes) but from the inadequate control (i.e., enforcement) of safety-related constraints on the development, design, construction, and operation of the entire socio-technical system.

System safety, then, can be reformulated as a system *control* problem rather than a component failure or reliability problem: accidents or losses occur when component failures, external disturbances, and/or dysfunctional interactions among system components are not handled adequately or controlled—where controls may be managerial, organizational, physical, operational, or manufacturing—such that required safety constraints on behavior are violated.

Note that the use of the term “control” does not imply a strict military command and control structure. Behavior is controlled not only by engineered systems and 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, i.e., the subset of organizational culture that reflects the general attitude about and approaches to safety and risk management.

Three important concepts in STAMP are hierarchical safety control structures, process models, and migration toward states of high risk.

Hierarchical Safety Control Structures: Hierarchies are a basic concept in systems theory. At any given level of a hierarchical model of complex systems, it is often possible to describe and understand mathematically the behavior of individual components when the behavior is completely independent of other components at the same or other levels. But emergent properties like safety do not satisfy this assumption and require a description of the acceptable interactions among components at a level higher than the components; these interactions are controlled through the imposition of constraints upon the component interactions at the level below.

Figure 1 shows an example of a hierarchical safety control structure for a typical U.S. regulated industry, such as aircraft. Each industry and company will, of course, have its own unique control structure. There are two basic hierarchical control structures in Figure 1—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 that component; together these responsibilities should result in enforcement of the overall system safety constraint.

Figure 1: An Example Hierarchical Safety Control Structure

Hierarchies, in system theory, are characterized by control and communication processes operating at the interfaces between levels. 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 1, company management in the development safety control structure may provide a safety policy, standards and resources to project management and in return, receive status reports, risk assessment, and incident reports as feedback about the status of the project with respect to the safety constraints.

To completely understand the cause of accidents and to prevent future ones, the system's hierarchical safety control structure must be examined to determine why the controls at each level were inadequate to maintain the constraints on safe behavior at the level below and why the events occurred—for example, why the designers arrived at an unsafe design (in the case of the space shuttle Challenger loss, there were political and other non-technical influences) and why management decisions were made to launch despite warnings that it might not be safe to do so (again, there were political and economic reasons).

When designing a new system or analyzing an existing system using STAMP as the foundation, required safety constraints are identified at the system level first and then a top-down iterative process is used to identify required safety constraints that must be imposed at each of the lower levels. The entire safety 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.

Process Models: Another important part of STAMP is the concept of process models. In basic system (and control) theory, the controller must contain a model of the system it is controlling. For human controllers,

this model is usually called the *mental model*. Accidents in complex systems often result from inconsistencies between the model of the process used by the controller and the actual process state. For example, the autopilot software thinks the aircraft is climbing when it really is descending and applies the wrong control law or the pilot thinks a friendly aircraft is hostile and shoots a missile at it. Part of the challenge in designing an effective safety control structure is providing the feedback and inputs necessary to keep the controller's model consistent with the actual state of the process. Similarly, an important component in understanding accidents and losses involves determining how and why the controller was ineffective; often this is because the process model used by the controller was incorrect or inadequate in some way. The reasons for such an inconsistency are used in the new hazard and risk analysis techniques built on STAMP.

Because STAMP is based on a firm mathematical foundation (systems and control theory), computational modeling and analysis of safety and risk becomes feasible: the process models, along with the feedback control loops, can be computationally modeled and analyzed. We have experimentally built computational models of complex systems (described below) to demonstrate feasibility and practicality.

Migration toward Accidents: Traditional models of accident causation and safety engineering techniques are not only limited in the types of causal factors they consider, primarily component failures, but they usually treat the system as static. This simplification, however, limits our ability to manage risk effectively.

Systems are continually changing under physical, social, and economic pressures. In STAMP, systems are not treated as static designs, but as dynamic processes that are continually adapting to achieve their ends and to react to changes in themselves and their environment. For safety, the original system design must not only enforce appropriate constraints on behavior to ensure safe operation (the enforcement of the safety constraints), but the system must continue to operate safely (safety constraints must continue to be enforced) as changes and adaptations occur over time, for example, operators change how they use the system once they become familiar with it, managers demand different performance such as increased throughput, or doctors spend less time talking with patients.

Accidents in complex systems often involve a migration of the system and changes in the safety control structure toward a state where a small deviation (in the physical system or in human operator 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 led to a loss. Union Carbide and the Indian government blamed the Bhopal MIC (methyl isocyanate) release, one of the worst industrial accidents in history in terms of human death and injury, on the improper cleaning of a pipe at the chemical plant. The maintenance worker, however, was in fact only a minor and somewhat irrelevant player in the loss [12]. Instead, degradation in the safety control structure occurred over time and without any particular single decision to do so but rather as a series of independent decisions that moved the plant toward a situation where any slight error would lead to a major accident.

"The stage for an accidental course of events very likely is prepared through time by the normal efforts of many actors in their respective daily work context, responding to the standing request to be more productive and less costly. Ultimately, a quite normal variation in someone's behavior can then release an accident. Had this 'root cause' been avoided by some additional safety measure, the accident would very likely be released by another cause at another point in time. In other words, an explanation of the accident in terms of events, acts, and errors is not very useful for design of improved systems [17]."

Degradation of the safety-control structure over time may be related to *asynchronous evolution* [7], 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 safety control aspects, may be neglected or inadequate. Asynchronous evolution may also occur when one part of a properly designed system deteriorates. In both these cases, the erroneous expectations of users or system components about the behavior of the changed or degraded subsystem may lead to accidents. One factor in the E. coli contamination of the water supply of a small town in Ontario, Canada, was the privatization of the government water testing laboratory without establishing feedback loops from the private labs to the government overseers of the water system to detect when conditions were degrading [11]. A factor in the loss of contact with the SOHO (Solar Heliospheric Observatory) spacecraft in 1998 was the failure to communicate to the operators that a functional change had been made in a procedure to perform gyro spin-down [14]. A factor in the friendly fire shootdown of a U.S. Army

Blackhawk helicopter by a U.S. Air Force fighter over northern Iraq in 1996 was that the Air Force had upgraded their radio technology while the Army had not, thus violating the safety constraint that U.S. forces would be able to communicate over their radios [11].

Applying STAMP to System Safety Problems

Using this basic model of accident causation as the foundation, powerful new approaches to system safety can be developed, just as techniques such as Fault Tree Analysis and Failure Modes and Effects analysis were constructed atop the basic chain of events model. Because the foundations of STAMP are mathematical, computation models and tools can be used to support these new techniques.

Basic structural control models, such as shown in Figure 1, will be used in most new tools. To augment the static structural models, formal models can be used of changes over time and the physical and social influences that can lead to these changes. One such modeling technique we have found useful is system dynamics [20]. The field of system dynamics, created at MIT in the 1950s by computer pioneer Jay 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. 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 [20].

“All too often, well-intentioned efforts to solve pressing problems create unanticipated ‘side effects.’ Our decisions provoke reactions we did not foresee. Today’s solutions become tomorrow’s problems. The result is policy resistance, the tendency for interventions to be defeated by the response of the system to the intervention itself. From California’s failed electricity reforms, to road building programs that create suburban sprawl and actually increase traffic congestion, to pathogens that evolve resistance to antibiotics, our best efforts to solve problems often make them worse. At the root of this phenomenon lies the narrow, event-oriented, reductionist worldview most people live by. We have been trained to see the world as a series of events, to view our situation as the result of forces outside ourselves, forces largely unpredictable and uncontrollable...System dynamics helps us expand the boundaries of our mental models so that we become aware of and take responsibility for the feedbacks created by our decisions.” —John Sterman [21]

System behavior is modeled in system dynamics by using feedback (causal) loops, stocks and flows (levels and rates), and the non-linearities created by interactions among system components. In this view of the world, behavior over time (the dynamics of the system) can be explained by the interaction of positive and negative feedback loops [18]. The combined STAMP and system dynamics models can be used to devise and validate fixes for technical and organizational safety problems and to design systems with lower risk.

STPA: A New Hazard Analysis Technique.

To create new approaches to both technical system hazard analysis and organizational risk analysis based on STAMP, we have identified a set of factors that can lead to violation of safety constraints, such as inadequate feedback to maintain accurate mental (process) models. These factors are derived from basic control theory. We used these factors in creating a new hazard analysis technique called STPA (STAMP Analysis) [3, 11], which we have applied to both technical system hazard analysis and organizational risk analysis, as described below.

Uses for STAMP

A major advantage of this new approach to system safety engineering is that it can handle very complex systems and both the technical and social (organizational and cultural) aspects of accident understanding and prevention. STAMP can be applied to accident and incident investigation; hazard analysis (i.e., investigating an accident before it occurs); enhanced preliminary hazard analysis that allows safety to be considered during early system architecture selection; design for safety (at the technical and social level); risk analysis of organizational safety policies and designs and identification of leading indicators of migration toward increasing organizational risk; and programmatic risk analysis of the tradeoffs between

safety, performance, schedule, and budget.

Accident and Incident Investigation and Analysis

All current accident models and accident analysis techniques suffer from the limitation of considering only the events underlying an accident and not the entire accident *process*. The events preceding the loss event, however, reflect only the *results* of dysfunctional interactions and inadequate enforcement of safety constraints. The inadequate control itself is only indirectly reflected by the events.

A focus on proximal events often makes it appear that accidents are the result of an unfortunate coincidence of factors that come together at one particular point in time and lead to the loss. This belief arises from too narrow a view of the causal time line. As argued above, systems are not static. Rather than accidents being a chance occurrence of multiple independent events, they tend to involve a migration to a state of increasing risk over time. A point is reached where an accident is inevitable (unless the high risk is detected and reduced) and the particular events involved are somewhat irrelevant: if those events had not occurred, something else would have led to the loss. This concept is reflected in the common observation that a loss was “an accident waiting to happen.” The proximate cause of the Challenger Space Shuttle was the foam coming loose from the external tank and damaging the re-entry heat control structure. But many potential problems that could have caused the loss of the Shuttle had preceded this event and an accident was avoided by luck or unusual circumstances. The economic and political pressures had led the Shuttle program to drift to a state where any slight deviation could have led to a loss [9].

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

Instead of decomposing behavior into events over time, systems theory (and STAMP) focuses on systems taken as a whole. It assumes that some system properties can only be treated adequately in their entirety, taking into account all facets relating the social to the technical aspects [16]. These system properties derive from the relationships among the parts of the system: how the parts interact and fit together [1]. Thus, the system approach concentrates on the analysis and design of the whole as distinct from the components or the parts and provides a means for studying emergent system properties, such as safety [11]. Using this approach as a foundation, new types of accident analysis (both retroactive and proactive) can be devised that go beyond simply looking at events and can identify the processes and systemic factors behind the losses and also the factors (reasons) for migration toward states of increasing risk. This information can be used to design controls that prevent hazardous states by changing the design to prevent or control the hazards and migration and, in operational systems, detect the increasing risk before a loss occurs

To completely understand the cause of accidents and to prevent future ones, the system’s hierarchical safety control structure must be examined to determine why the controls at each level were inadequate to maintain the constraints on safe behavior at the level below and why the events occurred. The goal is not to assign blame—blame is the enemy of safety¹—but to determine why well-meaning people acted in ways that contributed to the loss. If the hierarchical safety control structure has not already been documented, then that should be done and then used to identify and understand the safety control inadequacies in the engineered system (the physical system), the aspects of the design and the environment that affected the loss, and the systemic factors that contributed to the loss.

The first step in the accident analysis is to understand the physical factors involved in the loss, including:

- The limitations of the physical system design. For the sinking of the Herald of Free Enterprise

¹ In the Herald of Free Enterprise loss, for example, many of the individuals at Townsend Thoresen (the owner of the ferry) were prosecuted for manslaughter, as was the operating company. Such reactions do not increase safety. See recent work on *Just Culture*, e.g., Dekker [2].

- [19], for example, the ferry's loading ramp was too low to reach the upper car deck at high tide,
- The failures and dysfunctional interactions among the physical system components, e.g., the Assistant Bosun did not close the doors to the Herald's car deck, and
- The environmental factors, e.g., the high spring tides in Zeebrugge where the sinking occurred, that interacted with the physical system design.

Most accident analyses include this information, although they may omit dysfunctional interactions and look only for component failures.

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

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

To understand why people behave the way they do, we must examine their mental models and the environmental factors affecting their decision making. All human decision-making is based on the person's mental model of the state and operation of the system being controlled. For example, the Herald's First Officer assumed that the Assistant Bosun had closed the doors, the Assistant Bosun may have thought that someone else would notice that the doors were open and close them, and the Captain thought the doors had been closed. Preventing inadequate control actions in the future requires not only identifying the flaws in the controllers' mental models (including those of the management and government components of the hierarchical safety control structure) but also why these flaws existed. For example, the Captain's inadequate mental model (thinking the doors were closed) resulted from lack of feedback about the state of the doors. All of them thought that leaving the doors open would not cause a loss of the ferry because a year earlier one of the Herald's sister ships sailed from Dover to Zeebrugge with bow doors open without incident, i.e., they had inadequate knowledge about the potential ferry hazards.

The impact of the operating environment (including environmental conditions, cultural values, etc.) must also be identified. For example, the problematic ferry design features were influenced by the competitive ferry environment in which the ferry was to operate.

The accident report blamed a "disease of sloppiness and negligence at every level of the corporation's hierarchy" [19]. But this superficial level of analysis (management sloppiness and negligence) is not useful in preventing future accidents—it simply provides someone to blame and to prosecute. It does not eliminate the underlying pressures that led to the poor decision making nor the inadequate design of the hierarchical safety control structure. Without changes that respond to those factors, similarly flawed and risky decision-making is likely again in the future, although the actual accident details may be very different. We have used system dynamic models to understand the complex environmental, social, and economic factors contributing to poor decision making in order to provide policy and other changes to improve risk-related decision making in the future [5, 10].

A complete accident/incident analysis based on STAMP usually finds dozens of causal factors

contributing to the accident process and points to many changes that could prevent future losses. Leveson provides several examples of such analyses of major accidents [11].

Early System Architectural Trades

Ideally, safety should be a part of the early decision making used in conceptual system design. However, effectively evaluating safety-related risk early enough to inform the early trade studies is not possible with current technology. We have created a new approach to preliminary hazard analysis (PHA) that can be performed prior to system architecture selection and thus can influence key decisions that will be difficult and costly or impossible to change later in the system lifecycle [4]. After an architecture is selected, the information generated in these early analyses can be used to design hazards out of the system during the detailed design process as the original analyses are revised and refined.

Risk in a PHA is usually evaluated using a matrix with various categories representing severity along one dimension and likelihood along the other. While severity (the consequences of the worst possible loss related to the hazard being considered) is easily determined, there is no way to determine likelihood before any system design or even architecture has been selected, especially in systems where new technology or new functions are included. Our new analysis technique uses the hazard mitigation of potential candidate architectures to estimate hazard likelihood. Hazards that are more easily mitigated in the design and operations are less likely to lead to accidents, and similarly, hazards that have been eliminated during system design cannot lead to an accident. The goal of the new analysis process is to assist in selecting an architecture with few serious hazards and inherently high mitigation potential for those hazards that cannot be eliminated, perhaps because eliminating them would reduce the potential for achieving other important system goals.

We chose mitigation potential as a surrogate for likelihood for two reasons: (1) the potential or eliminating or controlling the hazard in the design has a direct and important bearing on the likelihood of the hazard occurring (whether traditional or new designs and technology is used) and (2) mitigatability of the hazard can be determined before an architecture or design is selected— indeed, it helps in the design selection process.

The new process has been demonstrated in a MIT/Draper Labs project to perform an early concept evaluation and refinement for the new NASA space exploration mission (return humans to the Moon and then go on to Mars). The goal was to develop a space exploration architecture that fulfills the needs of the many stakeholders involved in manned space exploration. Because safety is an important property to many of the stakeholders, using it to influence early architectural decisions was critical as most of the architectural decisions would be very costly or impossible to change later in the development process.

The hazard-based safety_risk analysis developed is a three-step process:

1. Identify the system-level hazards and associated severities
2. Identify mitigation strategies and associated impact
3. Calculate safety/risk metrics for a given transportation architecture

The first two steps are performed only once, at the beginning of the process. They may have to be repeated if the architectural design space changes or if additional hazards are identified. The third step is repeated in order to evaluate as many candidate architectures and variations as necessary.

Hazard mitigation metrics are defined and used to evaluate and rank potential architectures. By systematically selecting and de-selecting options in the architecture description, it is possible to perform a first-order assessment of the relative importance of each architectural option in determining an Overall Residual Safety-Risk Metric.

Hundreds of architectures were evaluated for their inherent hazard mitigation potential. An automated tool was created to perform multiple evaluations based on the needs of the team responsible for designing the manned space architecture. The analysis started at the very beginning of the conceptual design phase and the methodology proved flexible and extensible enough to carry the team from Day 1 of conceptual design up to the beginning of the detailed design phase, at which point, a more detailed hazard analysis methodology such as STPA [3,11] will be necessary and safety-driven design of the system and its components can be started (see below).

Details are beyond the scope of this paper. The interested reader is referred to [4].

Safety-Driven Design

Ideally, hazard analysis should precede or at least accompany system design in order to avoid the

problems associated with changing design decisions after they have been made. The problem is that most of the existing hazard analysis techniques require a detailed design before they can be applied, because they rely on identifying potential component failures and their impact on system hazards. STPA is based on control rather than failure analysis and can be applied to hazards before a design is developed. The development of the design and the hazard analysis can go hand-in-hand, starting with the requirements for control of the high-level hazards and then refinement of the analysis as design decisions are made.

To demonstrate this safety-driven design process on a real system, we designed a spacecraft for outer planets exploration for NASA JPL (Jet Propulsion Laboratory) using safety-driven design procedures[15].

Safety Assessment of Complex Systems

Most current safety assessment techniques are impractical on very complex “systems of systems” but STAMP-based methods will work. We have applied the new approach to assessing the vulnerability of the U.S. Missile Defense System to inadvertent launch. The latter is a vast system of systems, including radars, launch platforms, early warning systems, interceptors, etc., some of which have been used for decades and others of which are new. While some of these components had been analyzed using traditional safety analysis techniques, an analysis of the hazards at the integrated system level was needed. STAMP-based methods were tried after nobody could figure out how to apply any existing techniques to such a complex system. The assessment was successfully completed on the integrated system and, in fact, the analysis found so many paths to inadvertent launch that deployment and testing was delayed six months while these vulnerabilities were fixed. STAMP and STPA have now been adopted by the government for all future missile defense system analysis.

Organizational Risk Analysis

STAMP can go beyond physical system design. New approaches to organizational risk analysis based on STAMP involve creating a model of the social and organizational control structure and identifying the safety constraints each component is responsible for maintaining, a model of the social 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. To model the social dynamics and pressures, we use system dynamics as described earlier.

We have completed a demonstration of applying STAMP to organizational and cultural risk analysis in the U.S. manned space program, specifically the current Space Shuttle operations program [10] Our models start with Congress and the White House and continue down through the NASA management structure to the engineering project offices and the actual operations (in the case of the Space Shuttle). In this analysis, we identified system-level requirements to reduce poor engineering and management decision-making leading to an accident, identified gaps and omissions in the operational program design and changes made after the Columbia accident, and performed a rigorous risk analysis to evaluate proposed policy and structure changes and to identify leading indicators and metrics of migration toward states of unacceptable risk over time.

Programmatic Risk Analysis

STAMP-based modeling and analysis can be used to create sophisticated programmatic risk management tools. While looking at safety alone is important, practical risk management requires understanding the tradeoffs among safety, performance, schedule, and budget risks. In another demonstration project for NASA, we showed how STAMP-based methods could be used for programmatic risk analysis in the new NASA Space Exploration Initiative (to return humans to the Moon and go on to Mars) [5]. Again, the models included the entire socio-technical system from Congress and the Executive Branch down to engineering processes and management. A major differences between this demonstration and the one for the current Space Shuttle program described above is that this project involves development as well as future operations. A second difference is that we modeled and analyzed performance, budget, and schedule risks along with safety and showed how the results could be used for management decision making. For example, we found that attempting to speed up development resulted in surprisingly little improvement in schedule (less than 2 percent) 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 less required changes and rework when problems are discovered late. Although this result is probably not surprising to safety engineers, it was to managers and provided, in addition, a mathematical analysis of the differences and rationale. Another example result, in the area of workforce planning, was that the development of the Space Shuttle replacement (called Orion) would not be possible within the time frame anticipated unless Congress relaxed hiring constraints on NASA.

Conclusions

STAMP is not the only possible expanded model of accident causation that could be devised. The purpose of this paper is not to sell STAMP, but to encourage those working in this field to expand beyond the techniques and models created for simple electro-mechanical systems whose underlying assumptions no longer match the majority of the systems we are building today.

By creating new models, we will be able to provide much more powerful safety engineering techniques and tools. This hypothesis is supported by our experience with STAMP as described above. To make significant progress, we need to get beyond the limiting assumptions about accidents and accident causality of the past and build on new foundations that better reflect the types of systems and engineering prevalent today.

We are currently exploring the limits of STAMP and applying it to a large variety of risk management problems including safety in pharmaceutical testing, hospitals, the process industry, and the air transportation system as well as non-safety problems such as corporate fraud and security of national infrastructure systems.

References

1. Ackoff, R.L. Towards a system of systems concepts, *Management Science*, 17(11):661–671, July 1971.
2. Dekker, Sidney *Just Culture: Balancing Safety and Accountability*, Ashgate, 2007.
3. Dulac, Nicholas and Nancy Leveson, An Approach to Design for Safety in Complex Systems, *International Conference on System Engineering (INCOSE)*, Toulouse, France, June 2004.
4. Dulac, Nicolas and Nancy G. Leveson. "Incorporating Safety into Early System Architecture Trade Studies," Int. Conference of the System Safety Society, August 2005.
5. Dulac, Nicolas, Brandon Owens, Nancy Leveson, Betty Barrett, John Carroll, Joel Cutcher-Gershenfeld, Stephen Friedenthal, Joseph Laracy, and Joel Sussman, Demonstration of a Powerful New Approach to Risk Analysis for NASA Project Constellation, CSRL Final Project Report, March 2007 (can be downloaded from <http://sunnyday.mit.edu/papers.html>)
6. JPL Special Review Board, Report on the Loss of the Mars Polar Lander and Deep Space 2 Missions, NASA Jet Propulsion Laboratory, 22 March 2000.
7. Leplat, Jacques Occupational accident research and systems approach. In Jens Rasmussen, Keith Duncan, and Jacques Leplat (eds.) *New Technology and Human Error*, pages 181–191, John Wiley & Sons, New York, 1987.
8. Leveson, Nancy G., A New Accident Model for Engineering Safer Systems, *Safety Science*, 42(4), April 2004, pp. 237–270.
9. Leveson, Nancy G. "Technical and Managerial Factors in the NASA Challenger and Columbia Losses: Looking Forward to the Future," in Handelsman and Fleishman (eds.), *Controversies in Science and Technology, Vol. 2: From Chromosomes to the Cosmos*, Mary Ann Liebert, Inc., 2007.

10. Leveson, Nancy, Nicolas Dulac, Betty Barrett, John Carroll, Joel Cutcher-Gershenfeld, Stephen Friedenthal, *Risk Analysis of NASA Independent Technical Authority*, CSRL Final Report, June 2005 (can be found on <http://sunnyday.mit.edu/papers.html>)
11. Leveson, Nancy, *System Safety Engineering: Back to the Future*, unfinished manuscript, <http://sunnyday.mit.edu/book2.html>.
12. Leveson, Nancy, Mirna Daouk, Nicolas Dulac, and Karen Marais, Applying STAMP in Accident Analysis, *Second Workshop on the Investigation and Reporting of Accidents*, Williamsburg, September 2003.
13. Leveson, Nancy, Karen Marais, Nicolas Dulac, and John Carroll, Beyond Normal Accidents and High Reliability Organizations: The Need for an Alternative Approach to Safety in Complex Systems, to appear in *Organization Studies*.
14. NASA/ESA Investigation Board. *SOHO Mission Interruption*. NASA, 31 August 1998
15. Owens, B., Herring, M., Leveson, N.G., Ingham, M., Weiss, K.A., "Application of a Safety-Driven Design Methodology to an Outer Planet Exploration Mission," in *IEEE Aerospace Conference*, Big Sky, Montana, March 2008, pp. 174-193.
16. Ramo, Simon. The systems approach. In Ralph F. Miles Jr., editor, *Systems Concepts: Lectures on Contemporary Approaches to Systems*, pp. 13–32, John F. Wiley & Sons, New York, 1973.
17. Rasmussen, J. Risk Management in a Dynamic Society: A Modelling Problem. *Safety Science*, vol. 27, No. 2/3, Elsevier Science Ltd., pages 183-213, 1997.
18. Senge, P.M. *The Fifth Discipline: The Art and Practice of the Learning Organization*, New York: Doubleday Currency, 1990.
19. Sheen, Barry. *Herald of Free Enterprise Report*, Marine Accident Investigation Branch, Department of Transport (originally **Report of Court No 8074 Formal Investigation**, HMSO, London 1987)
20. Sterman, J. *Business Dynamics: Systems Thinking and Modeling for a Complex World*, New York: McGraw Hill, 2000.
21. Sterman, J, "All models are wrong: reflections on becoming a systems scientist," *System Dynamics Review*, 18(4):501-531, Winter 2002.