A New Approach to Risk Analysis with a Focus on Organizational Risk Factors

by

Karen Marais

B.Eng., Electrical and Electronical Engineering, University of Stellenbosch, 1994
B.Sc., Mathematics, University of South Africa, 1997

Submitted to the Department of Aeronautical and Astronautical Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2005

Signature of Author . . . .
Department of Aeronautics and Astronautics
February ***, 2005

Certified by . . . .
Nancy G. Leveson, Professor
Committee Chair, Department of Aeronautics and Astronautics

Certified by . . . .
John S. Carroll, Professor
Sloan School of Management

Certified by . . . .
Deborah J. Nightingale, Professor
Department of Aeronautics and Astronautics

Accepted by . . . .
Jaime Peraire, Professor
Department of Aeronautics and Astronautics
Chair, Committee on Graduate Students

A New Approach to Risk Analysis with a Focus on Organizational Risk Factors

by

Karen Marais

Submitted to the Department of Aeronautics and Astronautics
on ***, 2005 in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy
at the Massachusetts Institute of Technology

ABSTRACT

Preventing accidents in complex socio-technical systems requires an approach to risk management that continuously monitors risk and identifies potential areas of concern before they lead to hazards, and constrains hazards before they lead to accidents. This research introduces the concept of continuous participative risk management, in which risks are continuously monitored throughout the lifetime of a system, and members from all levels of the organization are involved both in risk analysis and in risk mitigation.

One aspect of effective risk management is accurate risk analysis that takes account of technical, human, and organizational factors. This research develops a new approach to risk analysis that improves on event-based models to include risks that do not depend only on component or subsystem failures, and incorporates both human and organizational factors. The approach enables the early identification of risk mitigation strategies, aids in the allocation of resources to best manage risk, and provides for the continuous monitoring of risk throughout the system lifecycle.

Organizational factors have been identified as a significant aspect of accidents in complex socio-technical systems. Properly managing and assessing risk requires an understanding of the impact of organizational factors on risk. Three popular theories of organizational risk, normal accidents theory (NAT), high reliability organizations (HRO), and normalization of deviance, are reviewed. While these approaches do provide some useful insights, they all have significant limitations, particularly as a basis for assessing and managing risk. This research develops the understanding of organizational risk factors by focussing on the dynamics of organizational risk. A framework is developed to analyze the strategic trade-offs between short and long-term goals and understand the reasons why organizations tend to migrate to states of increasing risk. The apparent conflict between performance and safety is shown to result from the different time horizons applying to performance and safety. Performance is measured in the short term, while safety is indirectly observed over the long term. Expanding the time horizon attenuates the apparent tension between performance and safety. By increasing awareness of the often implicit trade-offs between safety and performance, organizations can avoid decisions that unwittingly increase risk.

In addition to this general dynamic, several specific common patterns of problematic organizational behaviour in accidents in diverse industries are identified. While accidents usually differ in the technical aspects, the organizational behaviour accompanying the accident exhibits commonalities across industries. These patterns of behaviour , or archetypes, can be used to better understand how risk arises and how problematic organizational behaviours might be addressed in diverse settings such as the space industry and chemical manufacturing. NASA specific archetypes are developed based on historical accounts of NASA and investigations into the Challenger and Columbia accidents. The NASA archetypes illustrate several mechanisms by which the manned space program migrated towards high risk.

Thesis Supervisor: Professor Nancy G. Leveson
Title: Professor of Aeronautics and Astronautics and Engineering Systems

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

RISK IN MODERN SYSTEMS

> *A life without adventure is likely to be unsatisfying, but a life in which adventure is allowed to take whatever form it will, is likely to be short.*
>
> Bertrand Russell[1]

> *We often think, naively, that missing data are the primary impediments to intellectual progress—just find the right facts and all problems will disappear. But barriers are often deeper and more abstract in thought. We must have access to the right metaphor, not only to the requisite information.*
>
> Stephen Jay Gould, The Flamingo's Smile[2]

Socio-technical systems are becoming more complex in response to increasing performance and cost requirements. This increasing complexity makes it more difficult to ensure that the systems meet safety requirements. In particular, complex systems are susceptible to system accidents, which are caused by dysfunctional interactions between components, rather than by component failures alone. Such accidents are particularly difficult to predict or analyse.

In the past, safe design principles and operating procedures were developed over long periods of time, applying experience acquired by trial and error. Now, new types of systems are being developed and fielded so rapidly that learning solely from experience is not sufficient.

[1] [Russell, 1968]

[2] [Gould, 1985]

Risk management is the process of maintaining an acceptable level of risk throughout a system‘s lifetime. Risk analysis, or the identification and assessment of risk and mitigation options, is a crucial part of this process. Existing risk assessment methodologies do not provide a good understanding of the risks associated with complex socio-technical systems. These methodologies are appropriate for simple systems where ‘mechanical’ failures prevail. But they are fundamentally limited when it comes to complex socio-technical systems because they are all event-based and do not adequately capture emergent behaviour.

1.1 Origins of Risk

Risk arises from uncertainty about the future behaviour of a system. In the absence of uncertainty there would be no risk. We would know with complete certainty what would go wrong, when it would happen, and what the consequences would be. With sufficient investment of time and resources, we could then make completely informed decisions about whether or not to accept these future events. Unfortunately, the very characteristics of systems that make them useful also give rise to uncertainty and hence to risk. One goal of risk management, and system development and operation in general, is to reduce uncertainty as much as possible, and to make sure that the remaining uncertainty is identified and understood.

1.1.1 Uncertainty

The literature provides innumerable and often conflicting definitions of uncertainty. For the purpose of this thesis it is sufficient to define uncertainty as an inability to predict the future behaviour of a system. This inability is seen as arising from a lack of information.

Uncertainty can be grouped into three types based on the type of information deficiency. The first two are uncertainties of fact, that is, it is not known how close to the truth the measurements or beliefs are. The last one is stochastic, that is, the uncertainty is due to the random part of a system or process.

- Measurement uncertainty: Any measurement is limited in its accuracy. This uncertainty can be reduced, but not eliminated, by improving the measurement process.
- Epistemic or lack-of-information uncertainty: We cannot be certain about what we do not know. For example, it is not guaranteed that a specific risk does not exist simply because it has not been convincingly proven that it does exist. Lack-of-information uncertainty can be reduced by expending more resources to gain more information.
- Stochastic or aleatory uncertainty: This uncertainty results from the inherently random part of a system or process. Note that some events may appear random because we do not understand the underlying mechanism of the process generating the events. In this case, uncertainty that appears to be stochastic is in fact due to a lack of information[1]. Alternatively, measurement error may give rise to an apparently random distribution.

In risk analysis, there is uncertainty at all levels. Predictions based on the models that underlie all risk assessments are subject to both uncertainty of fact and stochastic uncertainty. Uncertainty of fact arises because a model is always an incomplete representation of a system, and the parameters used to define the model contain measurement and epistemic uncertainty. Stochastic uncertainty in the system being modeled can be represented by means of probability distributions. But these distributions are again only best-guess approximations of the underlying stochastic process.

There is uncertainty about the coverage of the analysis: Are we aware of all the risks? There is uncertainty about the likelihood of the risks: Do we have a good understanding of the likelihood of a risk being realised? And there is uncertainty about the consequences attached to the risks: How well do we know what will happen if a risk is realised, and how much we will care about the consequences?

[1] For example, chaos theory has shown that many apparently random phenomena are actually driven by an highly structured underlying process. The phenomena appear random because the underlying process is highly non-linear and is not visible to the observer.

1.1.2 Complexity

Complexity is an important source of uncertainty and hence risk. Complexity makes it difficult to understand or predict a system‘s behaviour, thus giving rise to epistemic uncertainty. The more complex a system is, the more opportunity there is for components to interact in unforeseen and possibly undesirable ways.

Systems become complex for various reasons. One reason is attempts to satisfy increasingly demanding requirements on their performance. Complexity allows systems to satisfy multiple and changing requirements. Although there are approaches to system design that can decrease complexity while retaining functionality [Leveson, 1995], there is a minimum level of complexity, or essential complexity, that arises from trying to satisfy a system‘s requirements.

Complexity also arises from the sheer scale of systems. Chaos theory tells us that any real system is at least partially non-linear and that no matter how accurately such a system is characterised, this non-linear nature implies that we cannot accurately predict all the behaviour of the system all of the time. We may not even be able to identify or characterise all the non-linear aspects of a system. For example, the number of possible pairwise interactions between system components increases combinatorially with the number of components. A relatively modest system with 100 components has $\binom{100}{2}$, or 495, possible pairwise interactions. Anticipating the nature of all the interactions between components rapidly becomes impossible as the number of components increases beyond the trivial.

Systems tend to become more complex as they age. Changes in the system‘s design resulting from changing requirements (e.g. adding functionality) or a changing operating environment (e.g. unavailability of obsolete computer chips) will generally increase the complexity of a system. In the case of socio-technical systems, bureaucracy tends to increase over time [McCurdy, 1993], thus increasing the system’s complexity (and bureaucracy). While efforts can be made to roll this acquired complexity back, it is usually difficult or impossible to strip a system back to its essential complexity.

Systems will in general therefore be more complex than is strictly necessary, but reducing the complexity may not always be feasible or necessary. Complexity does not always result in sufficiently inefficient, risky, or otherwise undesirable behaviour to make a reduction in complexity necessary.

Much has been written about different types of complexity and how they should be measured. See, for example [Moses, 2002]. For the purpose of this thesis, the following types and definitions of complexity are sufficient.

Dynamic Complexity. Systems that exhibit unexpected or unpredictable behaviour over time are said to be dynamically complex. Systems are not static: they change over time, and different aspects of a system change at different rates. This change is driven by various factors, such as changing user requirements, changes in the environment, and the physical ageing of components. The rates of changes are not necessarily constant—a slow rate of change may suddenly become rapid. For example, a crack in an aircraft wing that has been gradually growing may suddenly and rapidly propagate, causing a structural failure. The behaviour of a system at some future time will be different from its behaviour at the present. While it is expected that any prediction about future behaviour is subject to error, unpredictable or rapid rates of change can make predictions meaningless.

Dynamic complexity can arise in even structurally simple systems. For example, the Beer Distribution Game [Senge, 1990], which simulates a simple supply chain, illustrates how delays in the supply chain can lead to wild fluctuations in production and inventory. Lacking an appreciation for the impact of supply chain delays, players are unable to maintain production and inventory levels at the optimum level. Once the delay effect has been pointed out, players fare better, but a change in the delay at any point in the chain can again wreak havoc. Knowing why a system exhibits dynamic complexity does not guarantee that its behaviour can be predicted.

Dynamic complexity arises as a result of latency and non-linearity in systems. Latency is a delay between action and response. Such delays can result in dynamic complexity, as demonstrated by the Beer Distribution Game. In control systems, latency can drive a system into unstable, or dynamically complex behaviour [Ogata, 1990]. A system can exhibit a range of latencies. Zero latency means that a response immediately follows the action. In some cases zero latency may be desirable, for example, activating the “dead-man” switch on a machine should immediately cause operation to cease. In other cases, a delay between action and response may be desirable. For example, in cases where actions with potentially devastating consequences are accidently taken, a time delay may provide the opportunity to devise a compensating strategy.

Large delays between action and effect can make the effect appear unrelated to the original action. This lack of apparent connection can create the impression that the action has no effect. In the case of desirable effects, it may be unclear how to obtain the same effect again.

When a system output is not proportional to the corresponding input, the input-output relationship is said to be non-linear. Non-linearity makes it difficult to predict system behaviour, because a small change in input, or in system elements, can have a large effect on system behaviour. Non-linear behaviour can result from the physics of the system. For example, “clipping” occurs in an amplifier when the amplitude of the amplified signal is larger than the output capability of the amplifier. This gives rise to the familiar “distortion” on audio speakers when the volume is turned up too high. Non-linear behaviour also arises from positive feedback loops. For example, the irritating squeak on public address systems occurs when the output of the speakers feeds back into the microphones and is repeatedly amplified.

Non-linearity can also be observed in the characteristics of a system. Systems with a large number of different elements are said to be combinatorially complex, reflecting the non-linear increase in the number of possible element interactions as the number of elements increases. In system modelling, “state explosion” refers to the exponential increase in the possible number of states as the number of state space variables increases.

Interactive Complexity. systems are interactively complex when there are many unfamiliar or unplanned and unexpected sequences of events in a system that are either not visible or not immediately comprehensible [Perrow, 1999a; Leveson, 1995]. A similar view comes from organizational theory, where an organization is seen as being complex to the extent that its parts mutually sustain each other [March and Simon, 1993].

Social Complexity. Organizations that have a large number of stakeholders are socially complex. When these stakeholders are very different the social complexity increases further.

1.2 Accidents: Definitions and Models

To manage risk in modern complex socio-technical systems, an understanding of how accidents happen is necessary. This section first discusses the nature of accidents in modern systems. Next, an accident modelling technique based on systems theory is reviewed. The section concludes with some definitions of accidents and related concepts that will be used in this thesis.

1.2.1 Accidents in Complex Socio-Technical Systems

Accidents are rarely caused by simple, random component failure. Component failures that appear random at first sight often turn out to be caused instead by inadequate maintenance or by using the component in a way for which it was not designed. Similarly, accidents are usually not caused by simple human error either: humans often take actions that contribute to accidents because those actions appear rational given their understanding of the situation or because the system design encourages incorrect behaviour.

Further, even in high-technology systems using unprecedented designs and technologies, accidents are rarely caused by unforeseen or unforeseeable physical phenomena. For example, the Space Shuttle Challenger exploded shortly after take-off when the O-rings failed to seal the solid rocket booster joints. It was well-known that rubber loses its elasticity in cold temperatures, like those on the day of the launch, but somehow this knowledge did not translate into delaying the launch.

Component failure and human error are inadequate explanations for accidents. Investigations into accidents have shown that the causes of accidents are much more complex, involving factors at all levels of the system, from component failure to organizational and social factors. For example, the investigation into the Space Shuttle Columbia accident placed extensive blame on organizational factors [Gehman, 2003, p. 177]:

> The organizational causes of this accident are rooted in the Space Shuttle Program's history and culture, including the original compromises that were required to gain approval for the Shuttle Program, subsequent years of resource constraints, fluctuating priorities, schedule pressures, mischaracterizations of the Shuttle as operational rather than developmental, and lack of an agreed national vision. Cultural traits and organizational practices detrimental to safety and reliability were allowed to develop, including: reliance on past success as a substitute for sound engineering practices (such as testing to understand why systems were not performing in accordance with requirements/specifications); organizational barriers which prevented effective communication of critical safety information and stifled professional differences of opinion; lack of integrated management across program elements; and the evolution of an informal chain of command and decision-making processes that operated outside the organization's rules.

Perrow famously noted that in complex socio-technical systems accidents are “normal”, that is, the complexity of these systems makes accidents almost inevitable [Perrow, 1999a]. He also popularized the concept of system accidents. Perrow’s original formulation of system accidents referred to accidents that involved some degree of component failure. The current understanding of system accidents is that they are accidents that are caused not by the failure of one or more components, but by dysfunctional interactions between components (some of which may have failed). Such accidents can occur despite every component performing as required.

Any attempt to understand accidents or manage risk requires an underlying model of how accidents happen. Many accident models have been proposed, from very simple chain-of-event models (see [Leveson, 2004a] for a review) to models based on a systems perspective [Leveson, 2004a; Rasmussen, 1997]. While chain-of-events models are intuitively appealing, they do not provide an adequate explanation of how accidents occur in modern socio-technical systems. For example, they cannot incorporate the effects of feedback. More sophisticated models are necessary to provide a more complete understanding of accidents in these systems. In this thesis the STAMP model proposed by [Leveson, 2004a] is used.

1.2.2 A Control Theoretic View of Safety

STAMP (Systems Theory Accident Modelling and Processes) [Leveson, 2004a] is a new accident modelling technique that describes each level of the socio-technical structure of a system in terms of levels of control over safety. Accidents are seen as occurring as a result of an ineffective control structure not enforcing safe behaviour at each level of the system development and operation structures. The lack of control at each level leads to component failures, dysfunctional interactions among components, and unhandled environmental disturbances at a lower level. The most basic concept in the new model is not an event, but a constraint. The cause of an accident, instead of being understood in terms of a series of failure events, is viewed as the result of a lack of constraints imposed on the system design and operations. This definition of accidents fits both classic component failure accidents and system accidents, but it allows effectively incorporating software, human-decision making, adaptation, and social and managerial factors into the analysis. Accidents can then be understood in terms of why the controls that were in place did not prevent or detect maladaptive changes, which safety constraints, if any, were violated at each level, and why the constraints were inadequate or were not adequately enforced.

1.2.3 Definitions

Although the meaning of “accident” is intuitively clear, when performing a risk assessment a rigorous definition is necessary. The following definition from the safety literature is apt [Leveson, 1995]:

Accident. An accident is defined as an undesired and unplanned, but not necessarily unexpected, event that results in at least a specified level of loss.

Note that this definition does not restrict accidents to be events leading to injury or loss of life. The level of loss at which a particular event is referred to as being an accident is subjective, and depends on social, economic, technical and other factors. Safety is then defined as follows [Leveson, 1995]:

Safety. Freedom from accidents or losses.

These definitions for accidents and safety can be applied to all types of losses, not only those losses that result in injury or loss of life. For example, if a telecommunications satellite does not achieve the desired transmission rate, it has violated the constraint that it achieve a specified minimum transmission rate. In this case the loss is the loss of income associated with the satellite not delivering the required service. In project management, cost and schedule overruns may be viewed as violations of financial and time constraints by the project management system. The associated losses may include cancellation of the project or loss of income due to late delivery.

Note that accidents are often said to result from some kind of ‘failure’. The term failure is used to refer to anything from component failure (e.g. a light bulb burning out) to software ‘failure’ (i.e. where the software does something that contributes to an accident) to human error (e.g. an operator pressing the wrong button). Failure is used in so many ways that it has become almost meaningless. In this thesis failure will be used only in the sense of mechanical failure, that is, when something has broken and cannot start functioning again without being repaired.

1.3 Aspects of Risk

While there are many competing definitions of risk, there seems to be a common understanding of the general concept of risk as the possibility that human actions or other events lead to consequences that harm aspects of things humans value [Klinke and Renn, 2002]. There is general agreement that probability and consequence are two dimensions of risk. But does risk have other dimensions? Who defines these dimensions and how they are measured? This section explores this problem. First, the definitions of risk and associated concepts that will be used in this thesis are given. Second, the standard ‘scientific’ or realist view of risk is discussed. Then, an alternative view of risk, as a concept constructed by society, is discussed. This view introduces additional non-physical dimensions of risk that affect how risk is perceived.

1.3.1 Definitions

In defining risk it is useful to begin with the concept of a hazard, which is used in system safety to identify system states that could lead to an accident under particular conditions:

Hazard. A hazard is a state or set of conditions of a system that, together with other conditions in the system‘s environment, will lead inevitably to an accident [Leveson, 1995].

In the STAMP approach, hazards are states that violate the system safety constraints. They arise as a result of inadequate enforcement of control actions.

Note that hazards are defined relative to a system. What are considered ‘environmental’ conditions and what are considered system states depend on where the system boundaries are drawn. From the point of view of attempting to manage risk, the boundaries should be drawn to include those parts of the system over which the system developers have, or should have, some control. For example, there is no point in defining climate to be part of a system because at this time there is no way reliably to control the weather.

A hazard has two important properties: the *severity* or *damage*, and the *likelihood* of the hazard occurring [Leveson, 1995]. These two quantities together define the *hazard level*. Often neither quantity is fully defined. The likelihood may be unknown or may be only subjectively defined. The potential damage associated with the hazard may be known only partially. For example, some consequences of accidents associated with the hazard may not be known beforehand.

In the case of air traffic control, a hazard can be defined as two aircraft being closer together than the minimum separation distance [Leveson, 1995]. The constraint is that aircraft may not violate the minimum separation distance. The control requirement is to maintain minimum separation. The damage associated with this hazard is a mid-air collision, which could lead to complete or partial loss of one or both aircraft. There is also a possibility that people or property on the ground could be damaged by falling debris. The likelihood of the hazard is the likelihood of two aircraft violating the minimum separation distance.

The hazard *exposure*, or *duration*, is the length of time that the hazardous state exists [Leveson, 1995]. In the air traffic control example, the duration is the period of time that the two aircraft violate the minimum separation distance. Generally, the longer a hazardous state persists, the greater the likelihood that the necessary conditions will be present for an accident to occur.

A violation of the minimum separation distance is in itself not guaranteed to cause an accident. If the hazardous condition is noted sufficiently quickly the aircraft can execute evasive action and restore minimum separation. There are various ways in which this state could lead to a mid-air collision. For example, the two pilots could only become aware of each other when it was too late to execute evasive action, perhaps because of restricted visibility. Or one or more of the aircraft could execute an inappropriate evasive action and fail to avoid the collision. The likelihood that the hazardous condition is not noted or corrected before a mid-air collision occurs is the likelihood of the hazard leading to an accident.

Hazards should be defined in a pragmatic way. There is no point in defining hazards in such a way that nothing can be done about them. For example, continuing the air traffic control example, defining the hazard as two aircraft being in the same airspace is not useful. Air traffic volumes make such a requirement impractical. However, defining a minimum separation distance based on the capabilities of air traffic controllers, aircraft, and their pilots reduces the likelihood of mid-air collisions.

The risk associated with a hazard, then, is the possible hazard damage (consequences), combined with the likelihood of an accident occurring:

Risk. Risk is the likelihood and consequences of an accident occurring in a system.

In risk assessment, the damage associated with a hazard is usually referred to as the consequences or costs of the risk. These consequences may be foreseen or unforeseen, that is, all the consequences associated with a risk are not necessarily known beforehand. Figure 1.1 illustrates the relationship between hazard and risk.

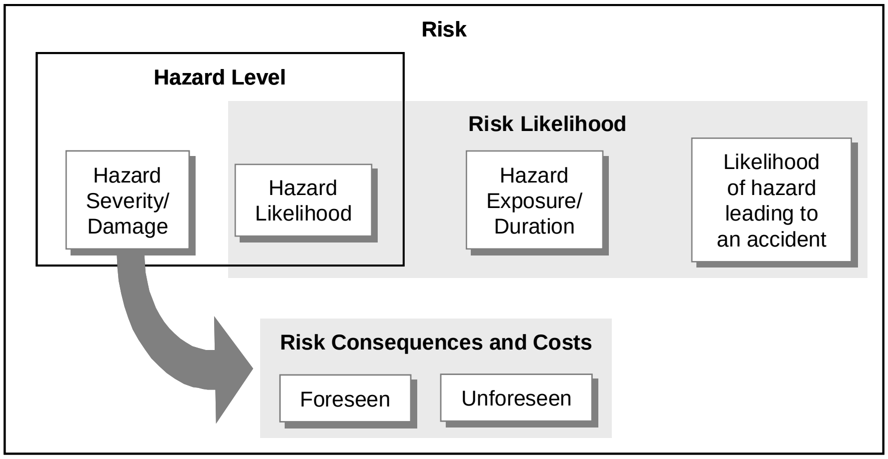

Risk
Hazard Level
Risk Likelihood
Hazard Severity/ Damage
Hazard Likelihood
Hazard Exposure/ Duration
Likelihood of hazard leading to an accident
Risk Consequences and Costs
Foreseen
Unforeseen

Figure 1.1 Components of Risk[a]

[a] Based on [Leveson, 1995]

The likelihood of an accident has three components: the likelihood of the hazard occurring, the hazard duration or exposure, and the likelihood of the hazard leading to an accident. Where these probabilities are quantified and independent, they can be multiplied to yield a single probability. Unfortunately, they are rarely independent, as discussed in Chapter 2. Looking at each probability individually may be useful as it provides better understanding of the way that an accident could occur. In the STAMP model, the likelihood of a particular risk is the likelihood of the system moving outside the constraint(s) associated with that risk.

1.3.2 Risk, Probability, and Consequence

Risk is usually measured in two ‘physical’ dimensions: probability and consequence. The expected value of risk is the combination of the probability and consequence.

Probability

There are two dominant philosophical concepts of probability. The *frequentist* approach views uncertainty as a true property of nature and considers probabilities to refer to the relative frequencies of objective physical events in repeated trials.

This approach is best illustrated by means of an example. Begin with a pair of fair dice. There are thirty-six possible combinations, but only eleven possible totals can be obtained from a single throw of the pair. There is only one way of throwing a double-six, so the probability of obtaining a total of twelve is $1/36$. There are four ways of obtaining a total of five, so the probability is $4/36 = 1/9$. This reasoning is familiar, and the understanding is that if a large number of throws are made, $1/9$ of them will result in a total of five.

Now look more closely at how these probabilities are obtained. The frequentist view of probability begins by saying that if a large number of trials is conducted, and the number of times that an event of interest occurs is counted, then the estimated probability of the event occurring in a future trial is the number of prior occurrences divided by the number of trials:

$$P(event) \to \frac{N_{events}}{N_{trials}} \tag{1.1}$$

Conducting large scale trials is often impossible, so the probability of an event is estimated based on a model of the system. In the dice example, it is assumed that each of the six faces of the die is equally likely to appear:

$$P(any\ given\ face) = \frac{1}{6} \tag{1.2}$$

This assumption is a model of a single die. By combining two such models the probabilities of the compound events corresponding to throwing two dice can be calculated. The rules of probability allow for the combination of multiple simple events to create probabilities for compound events. Probabilistic Risk Assessment (PRA), discussed in Chapter 2, is based on combining the probabilities of simple events to obtain the probabilities of system failures.

The *epistemic*, or *Bayesian*, approach views probability as a subjective measure of uncertainty used to express rational agents' degrees of belief in specific hypotheses. In this view probabilities can be assigned not only to random events, but to any other kind of statement. So, for example, a frequentist and a Bayesian might both assign a probability of $1/2$ to the event of getting a head on a coin toss, the Bayesian could also assign a probability to a personal belief, such as there being life on Mars, without making any assertion about relative frequency. In this sense, the Bayesian view of probability is wider. Bayesian inference can be used to update probabilities in a consistent and logical way to reflect a change in the state of knowledge. Under a Bayesian view of probability, all uncertainty results from a lack of knowledge. Thus there is no place for concepts like stochastic or random error.

Probability can also be expressed in different ways. For example, the reliability of a component could be defined as the probability of failing on demand or in terms of the mean time between failure (MTBF), which is actually an inverse probability. When probabilities are combined, care should be taken that they are all expressed in the same way.

Consequences and Costs

An accident may have many undesirable consequences, sometimes referred to as costs. Consequences can be financial (loss of income), human (loss of life or injury), loss of property etc. For example, the loss of the Space Shuttle Columbia resulted not only in the tragic deaths of astronauts, but also in the (temporary, it is hoped) grounding of the remaining shuttles.

When considering risks, all the associated consequences should be considered. The types of cost considered will affect the decisions that are made on the basis of the risk assessment. Should estimates be based on the worst possible consequence, or the most likely consequence, or on something else? How costs are defined determines which options are the most cost-effective or the safest. If the selected cost measures exclude an important cost, informed decision making is not possible. For example, launching a space mission without considering the cost of a loss of public support should the mission fail can result in an underestimation of the risk and in contingency measures to ensured continued public support not being put in place. Unfortunately, it may often be impossible to foresee all the possible consequences.

Costs range from direct and easily quantifiable to indirect and unquantifiable. For a project manager, a direct, quantifiable, cost of late system delivery is loss of revenue as a result of late-delivery penalties. An indirect, quantifiable, cost is unavailability of manpower for other projects. An indirect, unquantifiable, cost is loss of prestige by the project manager. For the space launch company direct, quantifiable costs associated with the launch failure are replacement costs for the rocket and possibly satellite (although these are usually partially covered by insurance so that the cost is partially transformed into increased insurance premiums). An indirect, less easily unquantifiable cost is satellite industry mistrust resulting from the failed launch.

Defining and estimating the costs associated with an accident is difficult because accident scenarios are usually not clearly defined or ‘waiting to be measured’. What is seen as undesirable and how undesirable it is relative to other options depends on individual value systems (cf. social complexity). For unprecedented accidents, it can be difficult to predict all the types of costs and how big they will be. The process of defining and estimating costs is subjective. For example, should human life be given a monetary value, and if so, how? Should only immediate loss of life be considered, and if not, how long into the future should the time horizon be extended? Different cost measures emphasise different values. When cost is measured in terms of loss of life expectancy, younger people are implicitly valued more than older people. There is no purely scientific way to answer this type of question.

Costs can be converted into a common form to allow comparison or summation of risks. The most common measurement of cost is monetary value. This conversion from one measurement to another assumes a common value system, is necessarily subjective, and may raise ethical issues.

Even when common cost measures can be found, accurately projecting costs is difficult. Usually the costs fall within a range of values. Both the endpoints of the range for a particular cost as well as the distribution of the cost within that range may not be known with certainty. Consider the simple case where a risk only has one cost associated with it. Assuming that the probability of the risk is accurately known, the cost could then be represented as spanning a range of values, and provide an indication of the uncertainty about the range limits and the distribution within the range. Generally there are multiple costs. Assuming that all the costs are expressed in commensurable form, how should they be combined? One approach is to sum over all the lowest and highest projected costs, generating worst and best case risks respectively. If the distribution of each cost within its range, and the uncertainty of the range limits, are known, they can be combined to yield the risk distribution and uncertainty. The distributions are often not known, however. The resulting risk limits then provide only a rough estimate of the risk, and determining the uncertainty associated with this number is difficult. Although combining all the costs into a single number is tempting, the resultant loss of information may be so severe that this number has little meaning. Simple metrics for complex systems are bound to fail in this way.

The Expected Value of Risk

Risk is often defined risk as some combination of probability and consequence, usually the expected value of risk:

$$E[Risk] = Probability \times Consequence \tag{1.3}$$

Conceptually, this definition should enable easy ranking of risks and thus simplify decision-making. In practice, it has limited use. It is not applicable to rare events or to events with widely differing costs. For example, consider two different lotteries where you pay $10 for a ticket. In the first lottery, you have a 1 in 10 chance of winning $100. The expected value, or risk, of this lottery is $10. In the second lottery you have a 1 in 10 000 chance of winning $100 000. Like the first lottery, this lottery has an expected value of $10. However, you are unlikely to value both lotteries equally.

The expected value definition further assumes that the consequences of different events can be estimated and are commensurate. When comparing two risks with different types of consequences, e.g. financial loss and loss of life, formula (1.3) is not readily applied.

Risk is not a simple concept and the expected value of risk is not useful except in the simplest cases. It permits easy ranking of options, but because the calculation allows so many inaccuracies, such a ranking is suspect. By keeping the factors separate, the risk associated with a particular choice can be better understood.

1.3.3 The Dual Nature of Risk: Objectivism and Constructivism

A major debate in risk management centers on the philosophical question of objectivism versus constructivism [Klinke and Renn, 2002]. Objectivism, or realism, holds that technical estimates of risk are true representations of observable hazards regardless of the beliefs or convictions of analysts. This approach can also be described as the technocratic mode [Cotgrove, 1981], where values and beliefs are excluded and emphasis is placed instead on ‘scientific’ data. Figure 1.2 illustrates the technocratic thinking process, and its application to risk assessment and management.

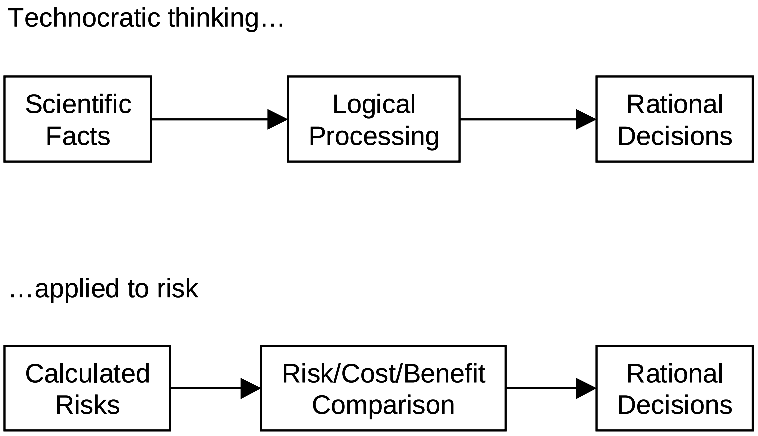


Figure 1.2 Technocratic Thinking in Risk Assessment[a]

[a] Adapted from [Cotgrove, 1981]

Many scholars have questioned the possibility of conducting objective analyses of risk. Perhaps ‘technical’ risk assessments reflect only the convictions of the professional risk assessors and as such have no more validity or universality than the assessments of other stakeholders, such as the lay public. The results of a risk assessment are affected by the values and backgrounds of those assessing risks. Extra-scientific or ‘soft’ factors always affect estimates of risk, but their effect is only noticed or manifested when there are value-based disagreements [Hatfield and Hipel, 2002]. For example group of engineers assessing risk in a plant probably have the same backgrounds and training and hence similar views of risk. It may be necessary to introduce an outside investigator with a different background to reveal underlying assumptions that affect the risk assessment. Constructivism, then, holds that risk assessments constitute mental constructions that can at best be checked against standards of consistency, cohesion, and internal conventions of logical deduction. These assessments have no validity outside the assessor’s logical framework. If risk assessments are nothing but social constructions, then they have no more normative validity for guiding regulatory action than stakeholder estimates or public perceptions.

Taking a purely objectivist view of risk ignores the social processing of risk information. Our attitude to risk, with its element of uncertainty, depends both on beliefs about risk and on value systems. Briefly put, $beliefs \times values = attitudes$ [Cotgrove, 1981]. When considering a given course of action, individuals have certain beliefs about the possibility of consequences, which do not necessarily coincide with ‘scientific’ assessments. Individuals do not all necessarily have the same perceptions of the consequences of a particular course of action. Value systems determine whether we view the consequences are desirable or undesirable and to what degree. Thus people may have completely different perceptions of risks and their origins. For example, consider the thousands of people who die every year on US Highways [Clarke and Short, 1993]. This statistic is often used to question why people are afraid of a one in a million cancer risk from toxic chemicals, when there are obviously much more likely dangers to worry about. Part of the answer is that people feel they have little control over their exposure to toxic chemicals. But another possible question is why so many people are killed on highways. The standard response is human error, but one could just as easily question the road design, lack of public transportation, or automobile design. Or one could question how the statistic was created in the first place, or how the data are defined, collected, and analyzed.

By acknowledging the constructivist aspect of risk, a better understanding of the way that people perceive risk can be gained. Socially constructed dimensions of risk affect how people perceive risk. This, in turn, affects which risks they will consider acceptable in a given situation. Because these dimensions are socially constructed, they will tend to differ from one situation to another. There is no master set of dimensions. Nevertheless, some example dimensions are useful as a start. Table 1.1 suggests dimensions that reflect both the technical and socially constructed aspects of risk.

On the other hand, taking a purely constructivist viewpoint can lead to the technical, human, organizational and other factors underlying the risks being underemphasized and so reduce understanding of how best to deal with the risks. A purely constructivist or purely objectivist view of risk provides incomplete understanding. It is not obvious, however, how to reconcile these different viewpoints. Two things are clear: the consequences of realised risks are real; and the socially constructed aspects of risk strongly affect how

TABLE 1.1 Aspects of Risk[a]

Aspect	Description
Extent of damage	Adverse effects such as deaths, injuries, property damage.
Probability of occurrence	Estimate of the frequency of the loss event.
Incertitude	Overall indicator for the uncertainty components (e.g. uncertainties in probability distributions).
Ubiquity	Geographic dispersion of potential damage.
Persistency	Temporal extent of potential damage.
Reversibility	Possibility of restoring the situation to its state before the damage occurred (e.g. oil spill cleanup).
Delay effect	Latency between the initial event and the actual impact of damage.
Violation of equity	Discrepancy between those who enjoy the benefits and those who bear the risks.
Potential of mobilization	Potential for social action arising from inequity and injustice; psychological stress and discomfort; potential for social conflict and mobilization; spill-over effects (e.g. highly symbolic losses).

[a] Adapted from [Klinke and Renn, 2002]. See also [Slovic, 2000] for additional dimensions such as controllability—the extent to which people can control their risk exposure.

people feel about risks and their exposure to them. However these additional dimensions are not criteria for determining whether someone is at risk, they merely determine people‘s attitudes towards risks. One approach to reconciling the two viewpoints is to use social, political and other processes to set the thresholds of acceptable risks (criteria for evaluating risks), and technical assessments to determine the extent of risks [Klinke and Renn, 2002].

1.4 Terminology

This section defines the various processes that will be referred to in this thesis.

Hazard Analysis. The identification of hazards and the assessment of hazard levels [Leveson, 1995].

Risk Analysis. The identification and assessment of risks.

Risk Identification. The process of identifying hazards and the associated environmental conditions that will inevitably lead to an accident.

Risk Assessment. The process of determining the likelihood and/or consequences and costs associated with identified risks.

Risk Evaluation. The process by which societal institutions such as agencies, social groups, or individuals determine the acceptability of a given risk [Klinke and Renn, 2002].

Risk Communication. The communication of the results of a risk analysis to stakeholders such as company executives and potentially affected communities.

Risk Management. The process of maintaining risk at an acceptable level throughout the lifetime of a system.

1.5 Research Goals and Thesis Outline

Complex socio-technical systems pose a challenge to risk managers. The complexity of these systems makes it difficult to identify and assess risk. When risks are not identified or are underestimated disaster, can ensue. In particular, recent accidents have demonstrated that an understanding of the way organizational behaviour contributes to risk is critical to effectively managing risk.

This thesis develops an approach to risk analysis that is applicable to modern, complex socio-technical systems. The research follows a two-fold approach to developing an alternative approach to risk assessment: theoretical and empirical. The theoretical approach draws on various disciplines and concepts to draw up a conceptual framework for assessing risk. The empirical approach uses examples of systems to identify common properties or behaviours that have an effect on safety. Figure 1.3 illustrates how these fields and concepts contribute to the development.

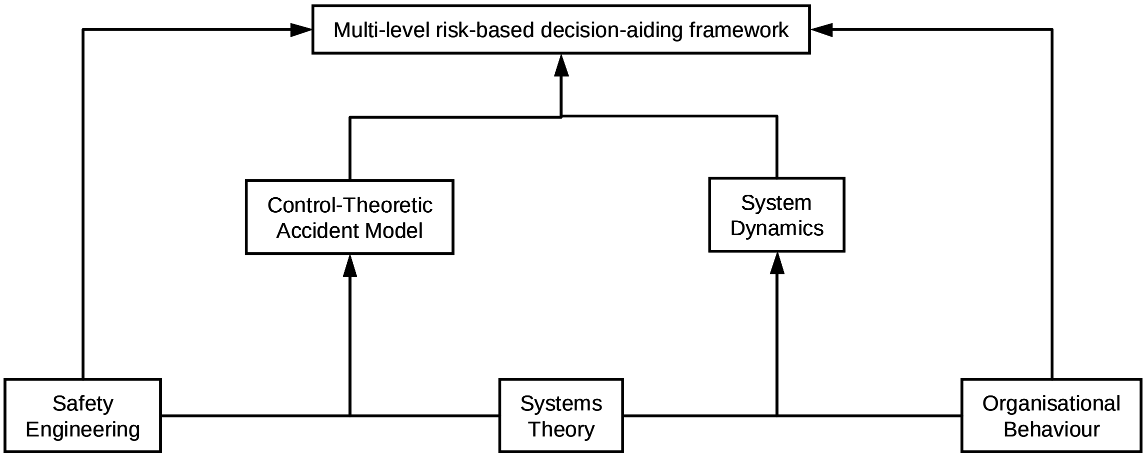

Multi-level risk-based decision-aiding framework
Control-Theoretic Accident Model
System Dynamics
Safety Engineering
Systems Theory
Organisational Behaviour

Figure 1.3 Research Foundations

Chapter 1 has introduced the concept of risk and the nature of accidents in modern complex socio-technical systems. Chapter 2 looks at risk management as a decision-making process and develops a set of criteria for what makes a good risk analysis. Based on these criteria, current risk analysis techniques are evaluated. These techniques do not appear to provide a good basis for decision making. An alternative approach is needed.

The importance of organizational factors in accidents has gained increasing acceptance. Numerous authors have written on the topic. In particular, sociologists have developed several theories to explain why accidents occur and how they can be prevented. Chapter 3 reviews three of the most popular sociological and organizational approaches to safety: normal accidents theory, high reliability organizations, and normalization of deviance.

Chapter 4 discusses the dynamic aspects of risk in organizations. First, a framework is developed to analyse the strategic trade-off between short and long term goals and understand why organizations tend to migrate to states of increasing risk. Next, a set of archetypes of organizational safety are developed. These archetypes describe specific mechanisms by which organizations unintentionally or unknowingly increase or fail to decrease risk, despite their best intentions. Each archetype is illustrated by means of examples from various industries.

Chapter 5 proposes an alternative approach to risk analysis and management based on systems theory. The proposed approach moves beyond event-based models to include risks that do not depend only on component or subsystem failures, and incorporate human, organizational, and societal factors. By taking an explicit lifecycle view of systems, the approach enables (1) the early identification of risks and risk mitigation strategies, (2) aids in the allocation of resources to best manage risk, and (3) provides for the continuous monitoring of risk throughout the system lifecycle. In addition, the approach emphasizes and enables the participation of members at all levels of the organization as well as other stakeholders in order to best identify, assess, and manage risks. The proposed approach addresses technical, human, organizational, and other factors.

Chapter 6 presents conclusions and suggestions for further work.

Chapter 2

ASSESSING RISK ASSESSMENT

> *Unless a person has calculated a risk, with its uncertainty, outlined his procedure, and highlighted the omissions, we cannot be sure he has made a reasoned, rational, decision... Risk assessment should not be an arcane discipline, carried out by narrowly focused experts in a back room.*
>
> Richard Wilson and Edmund Crouch[1]

Risk is a construct that we use to help us in choosing between different courses of action. We make decisions based on risk every day, whether explicitly or implicitly. For example, at a road crossing we may implicitly decide that it is too risky to cross when the traffic light is red. Risk assessment is the explicit identification and evaluation of the risks associated with different courses of action. Its purpose is to act as a decision-aiding tool. For example: Should a nuclear plant be built in a particular location? Therefore, in designing a risk assessment methodology, the aim should be to represent risks in a way that aids decision making. What, then, determines whether or not a risk assessment is good? This chapter develops a set of criteria for evaluating risk analysis methodologies. Then, current risk analysis techniques are reviewed and evaluated with respect to these criteria.

2.1 Criteria for Evaluating Risk Analysis Methodologies

Risk analyses are used to aid decision-making about uncertain situations. A risk analysis is good if it enables good decision-making. There is some debate about what makes a deci-

[1] [Wilson and Crouch, 2001]

sion ‘good‘. One way of defining decisions is by their outcomes: good decisions are those with good outcomes. But in some cases this approach equates good decision-making with good luck and bad decision-making with bad luck. An alternative definition of good decisions is decisions that reflect the best judgement given the information that was available at the time. Chapter 4 discusses this debate and the nature of good decision-making. Here, it is sufficient to note that it is difficult to make good decisions when the data on which the decisions are based is incomplete, inaccurate, or of questionable veracity.

This section develops a set of criteria for risk analysis based on the characteristics of socio-technical systems.

2.1.1 Criteria in the Literature

Although the literature on risk assessment, management and related fields is extensive, surprisingly little has been written about how one should evaluate a risk analysis methodology. In an early work, Fischhoff et al. discuss the generic complexities that decisions about acceptable risk present [Fischhoff et al., 1983]:

1. Uncertainty about how to define the decision problem;
2. Difficulties in assessing the facts of the matter;
3. Difficulties in assessing the relevant values;
4. Uncertainties about the human element in the decision-making process; and
5. Difficulties in assessing the quality of the decisions that are produced.

Using these complexities as a starting point, the authors develop a list of seven criteria for evaluating approaches to analysing risk. A risk assessment methodology must be

1. Comprehensive
2. Logically sound
3. Practical
4. Open to evaluation
5. Politically acceptable
6. Compatible with institutions

7. Conducive to learning

More recently, Haimes has built on these characteristics to provide ten characteristics of a good risk assessment, unfortunately without any explanation of how the additional criteria were derived [Haimes, 1998]:

1. Comprehensive
2. Adherent to evidence
3. Logically sound
4. Practical
5. Open to evaluation
6. Based on explicit assumptions and premises
7. Compatible with institutions
8. Conducive to learning
9. Attuned to risk communication
10. Innovative[1]

The next section adapts and augments these criteria to reflect the role that risk analysis should play in managing risk over the entire lifecycle of a system.

2.1.2 Criteria for Dynamic Participative Risk Analysis

The traditional view of risk assessment is that it is an activity that is carried out at the beginning of a system‘s lifecycle, thus providing a ‘snapshot’ of the risk associated with the system design. Based on this view, a set of static requirements can be developed, focussing on the accuracy of the risk assessment of the system design. By next expanding the view to consider risk over the lifecycle of the system (a constantly evolving picture), a set of dynamic requirements can be developed, focussing on providing the best information for managing risk over the lifetime of the system.

[1] This criterion is rather odd. It would seem that Haimes is pushing innovation for the sake of innovation!

2.1.3 Static Requirements

The first concern for a risk analysis methodology is that it can be used in practice by people other than academics. That is, the methodology must be practical (1), and compatible with the institutions (2) with which and in which it is intended to be used. Impractical and incompatible tools are rarely adopted outside academia.

From an engineering standpoint, the two ‘physical’ attributes of risk, probability and consequence, may appear to sufficiently characterise risk. But Chapter 1 argued that the perception of risk is also affected by social and psychological attributes, such as incertitude, ubiquity, persistency, irreversibility, latency, and inequity in distribution of risk [Klinke and Renn, 2002]. These attributes, though often ignored by the technical community, affect attitudes towards risk and whether or not a given risk is found to be acceptable, both inside and outside the organization. In developing a risk assessment methodology these attributes must therefore be taken into consideration, both in the evaluation and in the communication of risks (3).

Risks that are not known cannot be managed. The assessment must be comprehensive (4). It should consider more causal factors than component failure and deal with emergent behaviours such as system accidents. It should provide for the inclusion of non-technical factors such as human, organizational, political, and societal factors. If the assessment is not logically sound or does not adhere to the available evidence, risks be may under- or overestimated, and some risks may not be identified at all. In such cases the risk management program is hobbled and will be limited in its effectiveness. The assessment must be logically sound (5) and adherent to the evidence (6).

The purpose of risk assessment is to aid decision making about risks. Making good decisions requires understanding the results and limitations of the assessment. The results of the assessment must be represented in such a way that decision makers can understand the risks and where they come from (7). Any risk assessment must make assumptions and will therefore be subjective. The risk assessment technique itself has underlying assumptions about how accidents occur and about how risk should be modelled. For example, Probabilistic Risk Assessment (PRA) assumes that accidents happen only as a result of component failure, thus excluding most system accidents. When performing a risk assessment, the analyst must make assumptions about what is important and what is not. These assumptions must be made explicit to reveal the limitations of the assessment and so that they can be challenged and changed if necessary (8). Undocumented or implicit assumptions expose the entire risk evaluation process to second-guessing[1]. To build trust in the process inside and outside the organization, the assessment and its results must be open to evaluation (9). For example, varying levels of confidence in data should be made explicit. The methodology should not require such a high degree of specialised knowledge that other people in the organization cannot follow or understand the process or its results.

Static Requirements. The static requirements can be summarised as follows:

1. Be practical
2. Be compatible with institutions
3. Allow for inclusion of ‘non-physical’ risk attributes
4. Be comprehensive
5. Be logically sound
6. Be adherent to evidence
7. Represent results in a way that facilitates understanding
8. Explicitly document assumptions and premises
9. Open the process and results to evaluation

This list includes all the requirements given by [Fischhoff et al., 1983] and [Haimes, 1998], except for “politically acceptable”, which is here subsumed under (2); “innovate”,

[1] In the 1980s, questions were raised in Canada about the safety of the herbicide Alachlor. A controversy arose when three different stakeholder groups performed what they saw as objective, scientific assessments of the carcinogenic risk posed by Alachlor, only to come to three different answers. The different results arose not from bad science or incompetence, but from underlying value differences [Brunk et al., 1991]. Perhaps more explicit documentation and communication of the underlying values and assumptions would have revealed the underlying causes of the controversy and allowed an agreement to be reached [Hatfield and Hipel, 2002].

for which the reasoning is not clear; and “conducive to learning” and “attuned to risk communication”, which are included under dynamic requirements.

Now consider the requirements that arise from a view of risk as dynamic.

2.1.4 Dynamic Requirements

Risks are either accepted or rejected. They are accepted if they are judged to be sufficiently low, and they are rejected if they are judged to be too high. This judgement process may involve negotiation: risks that are initially deemed too high are sometimes accepted if no plausible alternatives can be found. If a risk is accepted, ‘coping’ strategies for dealing with the potential consequences of the risk must be devised and assessed. If a risk is rejected, system developers have two general options. They can abandon the system concept. Second, they amend the design to reduce risk to an acceptable level. This choice may entail a compromise on system performance or cost.

One approach to risk management is to perform the risk analysis once the system design has been completed. If necessary, the design is then adapted to bring risk to an acceptable level. Alternatively, if the risk is unacceptably high and cannot be lowered the project must be abandoned, and the time and money invested to that point lost.

This approach has obvious disadvantages. Once design decisions have been made, they are often not easily adaptable or reversible. Developing alternative designs late in the system development process is often not feasible. Manufacturing may already be in process, or there may not be sufficient time or resources to develop or implement alternative designs. Designers may in many cases be forced to rely on after-the-fact risk mitigation. Risk mitigation strategies developed after the system has been designed tend to have a ‘band-aid’ nature, since more elegant solutions often require fundamental changes in the system design. Such mitigation strategies are often less effective or reliable. For example, extensive and complicated operating instructions will in general not be followed without mistakes. The result is a system design that manages risk retroactively, which may increase cost, decrease performance, and leave the system with greater than necessary risk.

One of the reasons for using this apparently suboptimal approach to risk management is that existing risk assessment methods require a nearly complete system design as input. For example, as will be discussed in detail later, PRA works by assembling component data into system-level failure probabilities; it cannot be performed without detailed knowledge about components and their place in the system. A better approach is to start the risk analysis as soon as a system concept is considered, as discussed in Chapter 3.

Risk assessment should aid in the development of coping strategies (10), alternative designs (11), and mitigation strategies (12). Each time a decision is made, the impact of the decision on the system must be considered. Activities aimed at reducing one risk often give rise to a new risk elsewhere in the system. For example, the introduction of computers and software as a way of reducing the impact of human error has brought with it a whole new set of potential accident scenarios [Leveson, 1995]. The risk assessment should facilitate analysis of mitigation strategies and alternative designs and their impact on the system and its risk.

Chapter 4 will discuss the concept of adaptation in systems and how systems tend to migrate to the boundaries of safety. Safety tends to degrade in the absence of conscious effort [Rasmussen, 1997]. The risk assessment must allow for this dynamic behaviour in two ways. First, it must be structured in a way that allows convenient updating of the assessment as the system or its environment change (13). The results and method of the assessment must be open to examination and updating. Second, the assessment should indicate where and how adaptation may occur, so that risk managers can become aware when the boundaries of safe behaviour are being approached (14). The assessment should provide for the explicit representation of this migration, in order that it can be made obvious to decision makers before the boundaries are passed.

Employees can be a good source of information about an organization and the risks it faces. While expert risk assessors (should) have a good understanding of the risk assessment methodology and system safety, employees may have insights that are not formally documented. Maintaining or improving safety requires active participation from workers [Backström, 1997]. The results of the communication should be comprehensible and credible not only to management, but also to workers, at least as it pertains to their spheres of influence. Comprehensibility and credibility can be obtained by involving workers in the entire risk assessment process, from the structuring of objectives to the decision-making process. Alternatively, workers should be able to visualise the process and results of the assessment in an understandable way. By including employees in the risk assessment process (15), their insights can be used to obtain a more complete and accurate view of the risk. Furthermore, employees who are included in the risk assessment process will have a better understanding of how employees' behaviour affects risk. Including employees can encourage them to monitor risk in their part of the system and can empower them to speak up when problems arise.

Finally, the risk assessment must facilitate learning so that risk in the current and possible future systems can be better managed (16). The organization can avoid future mistakes by learning which risk management strategies work and which do not. For example, if a risk assessment reveals that a particular system design introduces unnecessary risks, such a design should be avoided in the future. Instead of relying on institutional memory, the risk assessment should provide mechanisms to retain this knowledge.

Dynamic Requirements. By viewing risk assessment as part of an iterative risk management process, the importance of taking a dynamic view of risk becomes apparent. This view gives rise to additional risk assessment criteria that are not traditionally considered. The dynamic requirements can be summarised as follows:

10. Aid in the development and assessment of coping strategies.
11. Aid in the development and analysis of mitigation strategies.
12. Be convenient to update to reflect changing risks.

13. Aid in identifying when the boundaries of safe behaviour are being approached.
14. Encourage and enable worker involvement.
15. Facilitate learning.

The next section reviews existing approaches to risk assessment of engineering systems and shows that, for modern socio-technical systems, these approaches do not satisfy the criteria for a good risk assessment methodology.

2.2 Qualitative Risk Representation

Risk analysis methods can be roughly classified as being quantitative or qualitative. In practice, specific instances fall somewhere on the continuum between being purely qualitative and being fully quantitative. Risks are usually classified according to their likelihood and/or consequences. High-consequence, high-likelihood risks are deemed the most critical, while low-consequence, low-likelihood risks are deemed less important. Although psychological and social dimensions of risk have been identified (see above), these are rarely used in practice, because they are difficult to assess, quantify or communicate.

Critical Item Lists (CILs) are used to maintain a record of those items whose failure could result in an accident. As such they can be useful in ensuring that attention is paid to these items. NASA classifies items whose failure could lead to a loss of the shuttle and crew as Criticality 1 (C1) and Criticality 1R (C1R), when there are redundant critical components. But this classification is easily changed and may even lead to problems. On the Space Shuttle Challenger the primary and secondary O-rings were mistakenly classified as C1R, under the erroneous belief that they provided redundant protection. The cold temperatures on the day of the launch affected both O-rings in the same way, an instance of a common-mode failure.

Risk matrices, or, likelihood-consequence grids, as shown in Figure 2.1, are commonly used to document the perceptions (inside or outside the organization) of the most critical risks in a system. They are usually used to prioritise hazards. Engineers may use their ‘gut feel' to estimate the risk of a system failure. Or the matrix may be used to represent perceptions of risks by different sectors of the public. Risk matrices may also provide a convenient summary of the results of another risk assessment.

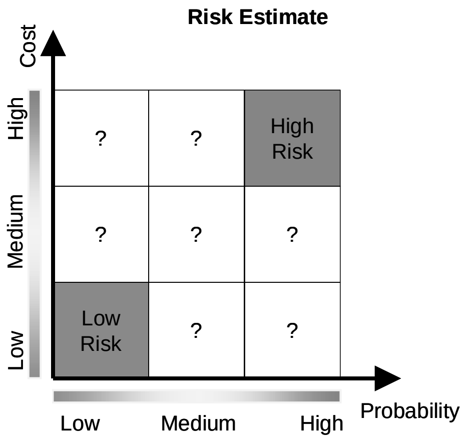


Figure 2.1 Risk Matrix

Strictly speaking risk matrices are not a risk analysis technique, but rather a way of eliciting or summarising perceptions of risks. They are included here however because they are often used instead of other risk assessment techniques.

Risk matrices have three major limitations. When they are constructed at the component or subsystem level, it is difficult to combine the information to estimate the risks for the entire system. When they are constructed at the system level they are of limited use in identifying leverage points for reducing risk (12, 13). Finally, they cannot be used for software design errors for which probability of failure is meaningless or unknowable.

2.3 Quantitative Risk Assessment

Quantitative Risk Assessment (QRA) also classifies risks according to their likelihood or consequence, but attempts to quantify these dimensions. Because the problem of determining the consequences of risks is difficult, fraught with ambiguity, and highly domain-specific (see Chapter 1), the focus is primarily on quantifying the likelihood of risks. Specialised techniques for estimating consequences are sometimes added to the standard QRA techniques to gain a more complete picture.

2.3.1 FMEAs and FMECAs

FMEAs and FMECAs attempt to identify a system’s weak points or components (critical items) whose failures could lead to accidents. By identifying weak points, FMEAs can be useful in identifying which parts of the system should be changed to reduce risks. While FMECAs are usually performed for parts with known failure rates, risk matrices can also be used to classify the estimated severity and frequency of each failure, as discussed above. The analysis considers each failure individually without regard to other failures in the system. Therefore FMECAs are not suitable for systems with a fair degree of redundancy or for those where common mode failures are a significant problem [Rausand and Hoyland, 2004]. More important, most accidents are caused by multiple interacting failures, a possibility that is ignored by these techniques. FMEAs and FMECAs examine and document all component failures, including those that do not have significant consequences. The resulting documentation burden is often unnecessarily large [Rausand and Hoyland, 2004]. Like other component failure based models, they are of no use in predicting accidents that are not caused by component failure (4). FMEAs and FMECAs are reliability engineering tools and have their uses in that area. But like other bottom-up approaches, they are of limited value in examining and ensuring the safety of complex socio-technical systems.

FMEAs and FMECAs have their place in system development and operation. They should not be mistaken for system level risk assessment tools however.

2.3.2 Actuarial Approaches

The two most common approaches to QRA are the actuarial approach and Probabilistic Risk Assessment (PRA) [Rasmussen, 1991]. The actuarial approach uses past accident data to estimate the future probabilities of accidents. This method is useful when there is a substantial record of past experience on which to base the estimate of accident frequency. When applying this estimate to future accidents, care must be taken to account for changes in the system or its environment that may affect the accident rate. For example, changes in technology may affect the accident rate. When Airbus introduced the A320, the first aircraft with substantial automation, the accident rate increased above the average for commercial jet airliners until unforeseen problems with the automation were resolved. Because it relies on past accident data, the actuarial approach is not applicable to new systems that do not yet have an operating history, systems that have changed significantly, or to systems where the accident rate is very low. An alternative approach that uses the system‘s design and operating environment to estimate risk is needed.

2.3.3 Probabilistic Risk Assessment

PRA was developed in the nuclear industry in the 1960s to estimate the probability of accidents in both existing and new nuclear plants in the absence of historical accident rates. The probability may be either an accident frequency or the probability of failure on demand, for example, the probability that a car will fail to start when the ignition switch is turned [Rasmussen, 1991]. PRA works by breaking a system down into subsystems and components, until a level is reached where reliability data for the subsystem or component can be estimated. For example, even though very few actual core melt accidents have occurred, significant data on the reliability of the pumps and valves in a nuclear plant may be available. The data are then re-aggregated, using some form of system models, such as event and fault trees, to estimate the overall probability of accidents for the entire system. The technique is therefore dependent on the concept of component failure as the driver for accidents. Component reliability data are estimated on the basis of test and experimental data, extrapolations of historical data, surrogate or generic data, and/or expert judgment. Obviously the accuracy of the results depend firstly on the accuracy of the model (e.g. are crucial failure modes included?), and secondly on the accuracy of the component reliabilities. Apart from the problem of estimating component reliabilities, there are insurmountable problems resulting from the bottom-up modelling approach.

A PRA of a system begins by defining a set of undesirable consequences, or end states (accidents). PRA assumes that each end state can be traced back to an initiating event (IE) through a sequence of discrete events. These IEs are identified as disturbances to normal operation, such as hardware failures, human errors, or natural phenomena. It is assumed that the likelihood of IEs can be estimated using physical models (e.g. structural analysis of a damwall), models of human performance[1] (e.g. human error rates), historical data (e.g. component failure rates), expert opinion, and so forth. The postulated sequence of events between the IEs and undesirable end states is identified using logic diagrams such as event and fault trees.

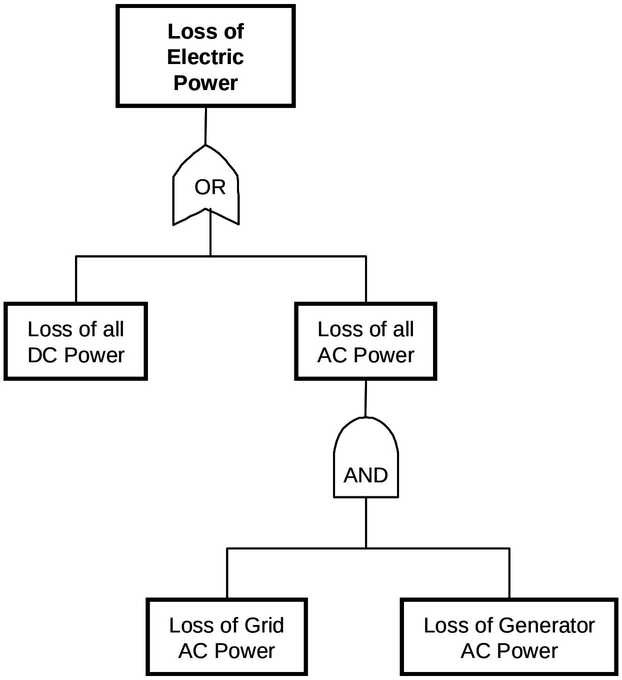


Figure 2.2 Fault Tree

[1] The only existing models of human performance are for extremely simple skills. Individual differences make it difficult to use even these physical skill estimates.

Fault Trees. A fault tree is a directed graph that uses Boolean logic to represent how binary-valued primary conditions combine to result in the top event. A fault tree analysis begins with an undesirable end state and then works backwards (deductively) to find which combinations of component failures can result in the end state. For example, for a nuclear plant the top event may be ‘loss of electric power’ [Rasmussen, 1991], as shown in Figure 2.2. In the nuclear plant both AC and DC power sources are required to power the safety systems. In other words, if the AC or DC power or both fail, the safety systems lose power, as indicated by the “OR” gate in the diagram. AC power is supplied from both an off-site grid and from an on-site generator. Either is sufficient, as indicated by the “AND” gate in the diagram. Using Boolean logic, an expression for the top event can be derived:

$$PowerLoss = AllDC \cup (ACCoffsite \cap ACgenerator) \tag{2.1}$$

The analysis continues downwards in this manner until it reaches components for which reliability data are known. A detailed expression for the top event can then be derived, and the probability of the top event can be obtained by filling in the probabilities for the individual components. In practice it is computationally difficult to propagate the probabilities, and a Monte Carlo analysis is often used to estimate the top event probability.

Event Trees. Event trees are almost the reverse of fault trees in that they work forward (inductively) from an initiating event and develop a time-sequence of events to determine which, if any, undesirable end states can be reached from the initiating event [Rasmussen, 1975]. The outcomes of each intermediate event are usually assumed to be binary (yes/no) but may also include multiple outcomes (yes, partly, no).

Fault and event trees can be combined in ‘bowties’ where the top event in each fault tree is propagated forward using an event tree, as shown in Figure 2.4. The fault tree is used to

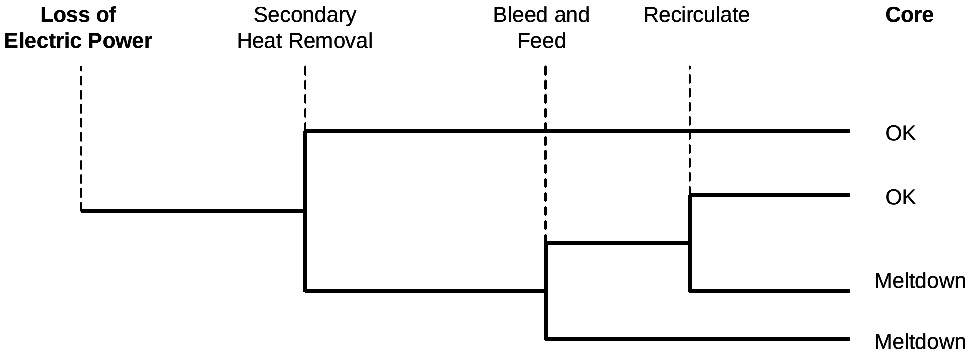


Figure 2.3 Event Tree

identify how the initiating event could occur and with what probability, while the event tree is used to model the possible end states and their severity.

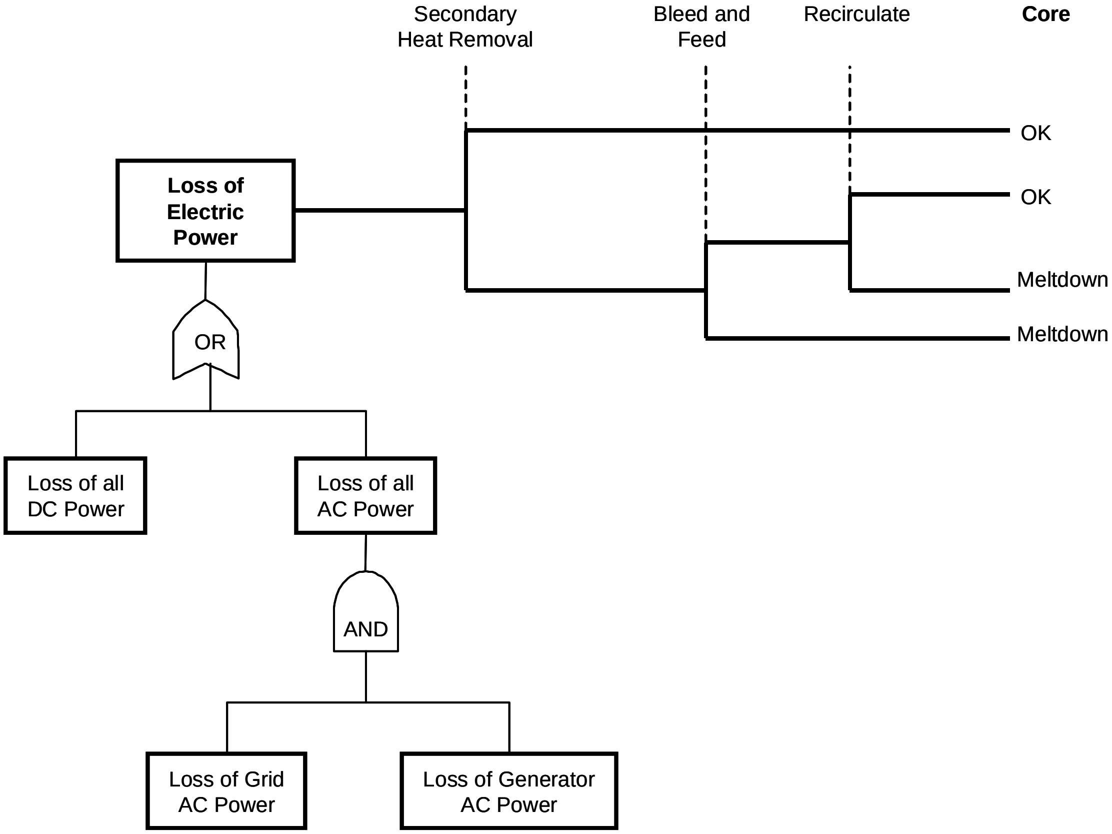


Figure 2.4 Fault Tree and Event Tree combined in a ‘Bowtie‘

Thus, using fault and event trees, sets of accident scenarios are developed for each undesirable end state. When used in a quantitative analysis, probabilities for the individual events are combined to obtain probabilities for the end states. Theoretically, common-cause and other dependent failures can be accounted for using probability techniques. In practice coverage is limited both by the difficulty of identifying dependencies and by the computational costs of accounting for these dependencies.

Fault and event trees for complex systems rapidly become very complex themselves. For example, a fault tree for the International Space Station (ISS) was constructed with 28 undesirable end states. It has 65 event trees, 450 fault trees, and 1500 basic events (component failures) [Futron, 2000]. This fault tree omits all human and software errors.

PRA is widely used in the nuclear industry, where it is seen as a cost-effective approach to risk reduction [Garrick, 1987]. In other industries it is less widely used (2). For example, its use in the U.S. space program remains contentious [cf. NASA, 1995; and Pate-Cornell and Dillon, 2001]. The first PRA on the shuttle as a whole, in 1988, gave a then shockingly high probability of $1/78$ for the loss of the shuttle [Buchbinder, 1989]. Subsequent studies have steadily improved the odds, from $1/90$ (1993) [SAIC, 1993], to between $1/76$ and $1/230$ (1995) [SAIC, 1995], to $1/245$ (1998)[1]. In each case the assumptions behind the assessment changed, as well as the definition of the failure space (e.g. which mission phases were included in the assessment). Whether this improvement in the estimated odds is a clearer reflection of the true probability of failure is not clear. Currently the historical frequency of failure for the Space Shuttle System is 2 failures out of 113 (or about $1/52$) total flights by the shuttle fleet. To illustrate the difficulty in determining the correct measure of reliability, note that one could instead look at the historical failure rate for each individual shuttle, as shown in Table 2.1

[1] Unpublished analysis. See, for example, [Vesely, 2003]

TABLE 2.1 Space Shuttle Historical Failure Rates

	First Flight	# Flights	Historical Failure Rate
Columbia	1981	28	$1/28$
Challenger	1982	10	$1/10$
Discovery	1983	30	$0/30$
Atlantis	1985	26	$0/26$
Endeavour	1991	19	$0/19$
Enterprise[a]			
Total		**113**	**2/113**

[a] Test vehicle not used for space flight

Another PRA study, started before the Columbia accident, is in process. Current results indicate a median failure rate of $1/165$ excluding the threat of micrometeorite orbital damage (MMOD), and $1/123$ including MMOD [Vesely, 2003].

2.3.4 Evaluating Probabilistic Risk Assessment

This section uses the risk analysis criteria to demonstrate the limitations of PRA when applied to complex socio-technical systems. For convenience, numbers in parentheses are used to indicate when a particular criterion is being referenced. It may also be useful to consult the summary table at the end of the chapter when reading these two sections.

Note that there are other approaches to risk assessment, some have been mentioned earlier in this chapter. This evaluation focusses on PRA because it is a popular technique in the nuclear and chemical industries and is gaining increasing acceptance in other areas, such as the space industry.

Problems with Component Reliability Estimates. The system-level failure probability estimates are dependent on the component reliability estimates. It is difficult for PRAs to deal with the absence of data. Estimates of system-level failure probabilities require that all the component failure probabilities be ‘filled in’. The quality of the reliability data for components or subsystems is often questionable, even though numbers may be stated to several decimal places [e.g. Crawford, 2001]. Tests and experiments used to obtain component reliability data do not always reflect the operating conditions of the component. For example, a component may be operating in a different temperature range to that in which it was tested. Similarly, extrapolations of historical data may not take changes in the operating conditions or component manufacturing processes, which may affect the component reliability, into account. Surrogate or generic data, where a component‘s reliability is estimated on the basis of a similar component’s reliability, can easily be wrong. Expert judgement can be prone to irrational biases such as organizational pressures and professional ideologies [Dietz and Rycroft, 1987]. Expert judgments do not always agree. If more than one expert judgment is obtained, whose opinion is ‘best’ and how should the data be combined? Software presents an even greater (and perhaps insurmountable) challenge. Software ‘reliability’ is not well-defined and most incorrect outputs have no relationship to particular hazards. How does one determine the probability of just one or two particular erroneous outputs or behaviours?

The varying levels of confidence in data are not always explicitly made visible[1] (8). Specifically, when Bayesian updating is used, the subjectivity associated with determining prior distributions is often not made apparent [Bier, 1999]. Worse, where data is unknown, the tendency is therefore to use guesstimates or expert judgements. Poor data at the component level will compromise the system-level results. PRAs thus do not necessarily provide a true reflection of the available evidence (6). Even apparently small changes in component probabilities can have dramatic impacts on the top level probability. For example, probability assignments that are assumed to be negligible sometimes have a significant effect on the final results [Bier, 1999].

Limitations of Fault and Event Trees. Fault and event tree analysis was developed by reliability theorists for relatively simple and primarily electro-mechanical systems, which

[1] Simply putting a probability distribution on component failure rates does not adequately address this problem. Probability distributions are usually selected based on their convenience, or some feeling about how the data ‘ought’ to be distributed.

could often be represented completely and directly by trees. For example, assuming that the component level failure rates were accurate, a fault tree of a system could be used to determine the top level failure probability. As a result fault and event trees were and often still are seen as being objective. But when these techniques are used on more complex systems the construction of the trees becomes subjective. Because these systems involve so much uncertainty, judgement is required to decide what to include in the trees and crucial elements may be omitted because of ignorance, error, or lack of imagination.

Fault trees are limiting because a complete set of system failures cannot be defined for even moderately complex systems [NASA, 1995]. For all but the most simple systems, there will always be another way that something could go wrong, whether it is an additional branch on a fault tree, or an additional failure that had not been thought of before. It is often difficult to incorporate additional scenarios or components into a completed fault tree, and top level probabilities must be recalculated each time. The British nuclear submarine ‘Tireless’ was forced to abandon exercises in the Mediterranean in 2000 when a cooling-system pipe cracked. The possibility of this pipe cracking had not been considered in the probabilistic risk analysis and (perhaps consequently) the pipe was never checked during maintenance [Redmill, 2002]. Precision is not the same thing as accuracy and does not imply it.

Similarly, for event trees it is difficult to define a complete set of initiating events that produce all the important accident sequences. In complex systems, there is no way of guaranteeing that all the event sequences that can lead to a particular accident have been identified. It is also difficult or impossible to determine the order of functions across the top of the event tree. When these functions interact with each other, as is the case in complex or tightly coupled systems, there is no correct order of events, because any particular ordering will exclude some interactions. Both event and fault trees can only model linear, direct relationships between events, specifically, they cannot represent feedback or adaptation [Rasmussen, 1997].

Fault and event trees do not handle common-cause or dependent failures well, and fall apart if multiple common-cause failures must be considered (5, 6). PRAs tend to consider only a small subset of common-cause failures. One unfortunate side-effect is that the impact of redundancy on increasing system reliability can therefore be overestimated. High levels of redundancy increase system complexity and can therefore even lead to accidents [Perrow, 1999a]. Risk mitigation strategies based on increasing redundancy may perversely decrease system safety (12).

Due to its bottom-up reassembly nature, PRA rapidly becomes intractable (1) for large systems, both in the sense of performing the PRA and in the sense of communicating and understanding the results. PRAs have been performed for large systems such as the International Space Station, but the fault and event tree models are so large that it is difficult to conceive of them being free from error. Results of the PRA tend to be communicated at a very high level, or focus on components or subsystems that the analysis indicates contribute significantly to the system-level failure probability. It is difficult for the decision makers to understand where and what the risks really are (7), and hence to make decisions about dealing with these risks. Similarly it is difficult to include workers (e.g. operators, supervisors) in the process (15). Developing coping and mitigation strategies and conceiving of alternative designs is therefore difficult (10, 11, 12). The complexity of the models makes evaluation of the risk assessment by outsiders (to the risk assessment) difficult (9). It is not practical to review exhaustively each individual component‘s reliability data, nor can the fault and event tree models easily be evaluated without redoing the assessment.

Questionable Rigour. PRA breaks a problem down into smaller problems until a point is reached where the problems can be solved. The approach is familiar to engineers and scientists, who are trained to solve problems in this manner. As a result, PRA is seen as being objective and rigorous when in fact it relies on subjective judgement. For complex systems, the PRA approach offers no guarantee of objectivity or rigour (6). In multiple assessments of the same system, different assessors come up with different evaluations [Pate-Cornell and Dillon, 2001]. Two reasons immediately come to mind. First, the quality of the reliability data for components or subsystems is often questionable. Second, like any risk assessment technique, PRA does not evaluate the risks of the system itself, but the risks of a model of the system. A system model is never more than an approximation of the system and is therefore subjective. System modelers decide, for example, where to draw the boundaries of the system and to what level of detail different system aspects are modelled. PRA models are not excluded from this limitation. Modelers must decide what types of failures (what goes wrong) to look at, and which failure scenarios (how it goes wrong) to consider. The construction of fault and event trees is part art, especially in the case of complex systems. It is thus highly likely that different modelers will not consider the same set of failures or failure scenarios, and will construct different fault and event trees, resulting in different assessments of risk. Any other risk assessment methodology will run into the same problem, but in the case of PRA it can be very difficult to determine how and why one model differs from another because the models are so large and involved (8, 9).

Hidden Assumptions. PRA analyses generally do not make their assumptions explicit, often because the analysts may not even be aware they are making these assumptions (8). Of course any risk assessment is based on assumptions, but when these assumptions are not made explicit the limitations of the analysis are not apparent. If the system or its environment change and these assumptions become inappropriate, it may not be obvious that the risk assessment must be adjusted. The focus on random component failures means that there is an implicit assumption that the system design is correct, that the realised system matches the design, and that the system is operated as intended. In reality, the design may be flawed, the realised system almost never matches the design, and systems are never operated in quite the way the designers intended. The PRA for nuclear power plants like Chernobyl certainly never included the assumption that the operators would disable the safety protections because they were under pressure to get a test done that would otherwise have to wait six months. Under those circumstances, the probability was actually quite high, certainly higher than in the calculations, which assumed the protections would be operational (or would only fail due to random events). But if one considers the production and other pressures on the operators, their behaviour appears less bizarre and even ‘reasonable’ (or at least understandable). Such pressures can be predicted, but they are usually ignored because considering them makes PRA impractical or infeasible.

Omission of System Accidents. When it comes to modern, complex, socio-technical systems, PRA is becoming even less applicable. Because it is based on the concept of component failures and event chains, it does not cover system accidents where no components fail (4). It has an inherent bias toward random, mechanical-type errors. In fact, it is difficult to find examples of accidents that are caused by genuinely random component failures [Crawford, 2001]. Because it depends on component failure, PRA cannot assess the contribution of software to system-accidents. Software (4) is often simply ignored, for example by modelling it as a component with zero probability of failing (5). While it is true that software does not fail in the sense of physically breaking like hardware does, software has been blamed for many accidents. For example, the Ariane 5 rocket exploded shortly after launch due to a bug in its software. Attempts have been made to include software in PRA, but these methods all revert to the component failure concept at some level and do not adequately account for the origin of software-related failures. See, for example, [Garret and Apostolakis, 1999]. PRA therefore excludes an entire group of accident scenarios.

Poor Treatment of Human Factors. Incorporating the effects of cognitively complex human interaction and organizational and societal influences in PRA is difficult (4). Attempts to include these factors have focussed on extending the “component failure” concept. When human interaction is modelled in a PRA, it is usually limited to models of human error on the level of the individual. That is, the human is viewed as just another system component that can fail. The best known technique for estimating human reliability is the Technique for Human Error Rate Prediction (THERP) developed by Swain *et al.*, see for example [Swain, 1990]. This technique includes only physical mistakes and not mental errors. It handles the things that are most easily automated and therefore usually are automated. Human error rate estimation techniques tend to have significant shortcomings, especially when combined with each other and with machine component reliabilities [Wickens and Hollands, 2000, pp.500-502]:

- **Lack of Database**: Empirical data for performance on simple acts and non-stressed conditions exists. But data on the frequency of cognitive errors related to diagnosis and problem-solving, and the effects of stress are much more limited and tend to rely heavily on expert opinion, which may be wrong.

  Although most accidents continue to be blamed on human error, there is growing recognition that human error is in general not random and often occurs because the design of the system makes it difficult or impossible to avoid erring [Leveson, 1995].
- **Error Monitoring**: Humans monitor their own performance and often correct errors before they can affect system performance. The probability of human error cannot be directly associated with the probability that the error will result in a system failure.
- **Non-independence of Human Errors**: The assumption that human errors are independent is untenable. The dependence may work in two directions. Committing one error may result in an increase in frustration and stress and so increase the probability of committing another error. Alternatively, committing one error may result in an increase in care and vigilance and so decrease the probability of committing another error. In either case, one cannot make the assumption that the probability of committing an error is independent of whether an error was made at an earlier time. Similarly, the error probabilities of two or more people cannot be assumed to be independent.
- **Integrating Human and Machine Reliabilities**: Human and machine reliabilities cannot be assumed to be independent. When a machine component fails, it is likely to affect the probability of human failure. For example, the operator may become more cautious and vigilant if a component has failed.

One result of the human error viewpoint is that safety improvement programs have tended to direct their efforts at reducing human error. But it is now recognised that human error is often the result of the interaction between the human operators and the rest of the system and its environment. In particular, people are unlikely to undertake actions that do not seem reasonable to them at the time, so that the majority of errors of commission (doing the wrong thing) result from short-cuts, trying to satisfy competing goals and incentives (e.g. production pressure), or misdiagnoses [Julius et al., 1995]. Similarly, errors of omission (neglecting to do the right thing) may also be driven at least in part by characteristics of the system. Efforts to reduce human error by increasing punitive measures, for example, may therefore only have limited effect (11, 12). For example, in an aircraft, both the pilot and co-pilot may make the same mistakes if the aircraft interface is poorly designed. It has also been argued that some level of error must be tolerated in order to reduce the risk of catastrophic accidents [Amalberti, 1996].

Poor Treatment of Organizational and Societal Factors. Likewise, organizational and societal factors have a significant impact on human performance and safety. Research into incorporating these factors into PRA has again attempted to extend the component failure and event chain concepts by quantifying the impact of these factors on component or subsystem failure probabilities, see, for example SHERPA [Embrey, 1992], SAM [Murphy and Paté-Cornell, 1996], and WPAM [Davoudian et al., 1994a]. In these approaches[1], organizations are seen as influencing the likelihood of human errors, for example using influence diagrams [Oliver and Smith, 1990]. While these approaches do provide insights into organizational behaviours that can influence safety, the insights are so general that they are not useful for improving the safety of a specific system (4, 11, 12) [Abramovici, 1998]. More important, they still rely on the component failure concept and so retain all the associated limitations discussed above [cf. Hollnagel, 1998]. For example, they do not take feedback into account (15).

Negative Impact on Risk Management. As illustrated by the examples of component redundancy and human error reduction, the type of risk assessment influences the risk management strategies that an organization uses. The result of the PRA focus on component failure (and human error) is that most risk reduction or safety improvement programs tend to focus on reducing the effects of component failures and reducing human error. Because PRAs tend to model the physical system in great detail while omitting qualitative effects such as organizational factors, they can lead to a false sense of accuracy [Murphy, 1994] and to ineffective mitigation measures (12). The focus on managing component

[1] See also [Abramovici, 1998] for an overview of these and other techniques.

failures means that many safety management programs are in fact reliability improvement programs: quality control and various forms of redundancy are used to reduce the likelihood of individual component failures and their impacts, respectively. While redundancy can improve system reliability, it is only useful to a point. It does not provide any protection against common-mode failures. It increases system complexity, making it even harder to understand the system. As long as the risk assessment techniques focus on component failure and human error, risk mitigation strategies too will focus on component failure and human error (12). PRA does not provide a way for understanding the complexities of modern systems, and cannot aid us in learning how to deal with these systems (16).

PRA does not convincingly meet any of the risk assessment criteria. Because it provides only a limited view of risk, risk management efforts based on the results of a PRA can have only limited effectiveness, and may even decrease safety. An alternative approach that complies with these criteria is needed.

The results of this section are summarised in Table 2.2 at the end of this chapter together with requirements for an alternative risk analysis methodology that would better satisfy the criteria.

2.4 Summary

This chapter has described the role of risk assessment in the risk management process and developed a set of risk assessment criteria based on this view. Current risk analysis techniques and their extensions do not address the requirements for risk assessment of complex socio-technical systems. An alternative approach to risk analysis that addresses some of these limitations is presented in Chapter 5.

TABLE 2.2 Summary of PRA with respect to Risk Assessment Criteria

	Risk Criterion	PRA	Requirements for a New Approach
1	Practical	PRAs usually have to be performed by experts. The models cannot be understood by those unfamiliar with PRA techniques such as fault and event trees and Bayesian updating.	Throughout the development of an alternative approach, emphasis must be laid on making the approach user-friendly and practical.
2	Compatible with institutions		
3	Take psychological and social attributes of risk into account.	Psychological and social attributes of risk are not considered.	Incorporating risk attributes other than likelihood and consequence is difficult. Psychological and social attributes of risk contribute to our perception of risk, and contribute to how people define what is acceptable and what is not. A new approach must find ways of acknowledging and dealing with these factors. This problem is easier stated than solved, however.
4	Comprehensive	Does not include system accidents, where components/subsystems work as they are supposed to. The contribution of software to system accidents is often simply ignored. Adequately accounting for the effects of cognitively complex human interaction and organizational and societal influences is difficult.	Stop using component failures as the building blocks for the risk assessment. Component failure is not a good model for software or human behaviour. Techniques based on component failure cannot incorporate system accidents. An alternative might be to use “constraints” as the building blocks, as discussed in Chapter 1.
5	Logically sound	Does not include system accidents, where components/subsystems work as they are supposed to.	An approach that is not based on component failure may be able to incorporate system accidents.
6	Adherent to evidence	Tendency to use guesstimates where data is unavailable.	Move away from component failures and the associated models (fault trees, event trees, etc.) to open the way up for models that can accommodate gaps in information.
7	Represent results in a way that facilitates understanding	Single number descriptions of risk and a focus on components does not adequately describe complex risks.	Clearly state assumptions and reasoning on which results are based.

TABLE 2.2 Summary of PRA with respect to Risk Assessment Criteria

8	Explicitly document assumptions and premises	The focus on component failure carries with it an underlying assumption that accidents where all the components work as they are supposed to, do not happen, or at least are vanishingly rare. This assumption is rarely made visible. Varying levels of confidence in underlying data are not always made visible.	The approach must help in identifying and laying bare underlying assumptions. It is never possible to operate without assumptions, what is important is that decision makers are aware of them as far as possible.
9	Open process and results to evaluation	The complexity of the models makes evaluation of the risk assessment by outsiders (to the risk assessment) difficult.	Assumptions and reasoning must be clearly documented.
10	Aid in the development and assessment of “coping” strategies	Inadequate identification and understanding of risks makes it difficult to devise coping strategies.	An approach that contributes to a system-level understanding of behaviour and risk will also aid in the development of ways of dealing with risk. A system-level understanding will help in escaping the problem of unwittingly replacing one risk with another. What is needed are fundamental solutions to risk-based problems, not symptomatic solutions like redundancy.
11	Aid in the development and analysis of mitigation strategies	Risk mitigation strategies and alternative designs that are proposed in response to PRA results tend to focus on eliminating unreliable components and increasing redundancy (i.e. reliability engineering techniques). These strategies do not necessarily improve safety and may perversely even decrease it.	
12	Be convenient to update to reflect changing risks	Updating component failure rates is simple, but adding additional components necessitates re-analyzing fault and event trees.	Assume that change will occur and provide mechanisms to incorporate change over the system lifetime.
13	Aid in identifying when the boundaries of safe behaviour are being approached	Because current PRA techniques do not account well for organizational and psychological factors, they cannot identify when the boundaries of safe behaviour are being approached	Incorporate both technical and non-technical factors. Do not rely solely on component failure. Provide ways of easily incorporating updated information and identifying new or changed risks.
14	Encourage and enable worker involvement	It is difficult to involve workers and to communicate the results.	The approach must be open to and understandable by people other than the risk assessors. It should encourage a system-level understanding of behaviour and risk.

TABLE 2.2 Summary of PRA with respect to Risk Assessment Criteria

15	Facilitate learning	PRA does not provide a way for understanding the complexities of modern systems, and will not aid us in learning how to deal with these systems.	The approach must be open to and understandable by people other than the risk assessors. It should encourage a system-level understanding of behaviour and risk. To avoid overwhelming users with detail, the technique should provide for some form of “selective detail”. Users should be able to select the detail with which they view different aspects of the system. The aim is to promote understanding, not spread confusion.

Chapter 3

ORGANIZATIONAL RISK THEORIES

"When the facts change, I change my mind. What do you do, sir?"

John Maynard Keynes[1]

The importance of organizational factors in accidents has gained increasing acceptance. Numerous authors have written on the topic. In particular, sociologists have developed several theories to explain why accidents occur and how they can be prevented. This chapter reviews three of the most popular sociological and organizational approaches to safety: normal accidents theory, high reliability organizations, and normalization of deviance.

3.1 Normal Accidents Theory

Charles Perrow's initial formulation of Normal Accident Theory[2] (NAT) was developed in the aftermath of the accident at the Three Mile Island nuclear power plant in 1979 [Perrow, 1982]. He introduced the idea that in some technological systems, accidents are inevitable or 'normal', and defined two related dimensions—interactive complexity and loose/ tight coupling—which he claimed together determine a system's susceptibility to accidents [Perrow, 1999a].

[1] See http://www.economist.com/research/Economics/alphabetic.cfm?TERM=KEYNES%2C%20JOHN%20MAYNARD

[2] The sections on Normal Accident Theory and High Reliability Organizations are based on [Marais et al., 2004].

Interactive complexity refers to the presence of unfamiliar or unplanned and unexpected sequences of events in a system that are either not visible or not immediately comprehensible. A *tightly coupled* system is one that is highly interdependent: Each part of the system is tightly linked to many other parts and therefore a change in one part can rapidly affect the status of other parts. Tightly coupled systems respond quickly to perturbations, but this response may be disastrous. Loosely coupled or decoupled systems have fewer or less tight links between parts and therefore are able to absorb failures or unplanned behavior without destabilization. Figure 3.1 provides examples of industries as classified by Perrow according to coupling and interactive complexity[1].

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

Perrow made an important contribution in identifying these two risk-increasing system characteristics. His conclusion that accidents are inevitable in these systems is however overly pessimistic. The argument advanced is essentially that efforts to improve safety in interactively complex, tightly coupled systems all involve increasing complexity and therefore only make accidents more likely.

The flaw in this argument is that the only engineering solution he considers is redundancy. He correctly argues that redundancy introduces additional complexity and encourages risk

[1] Note that Perrow’s diagram classifies industries, not particular implementations. This level of classification ignores the possibility of alternative design approaches to particular functional requirements. For example, pebble bed nuclear reactors are designed to be less complex and less coupled than traditional boiling water nuclear reactors.

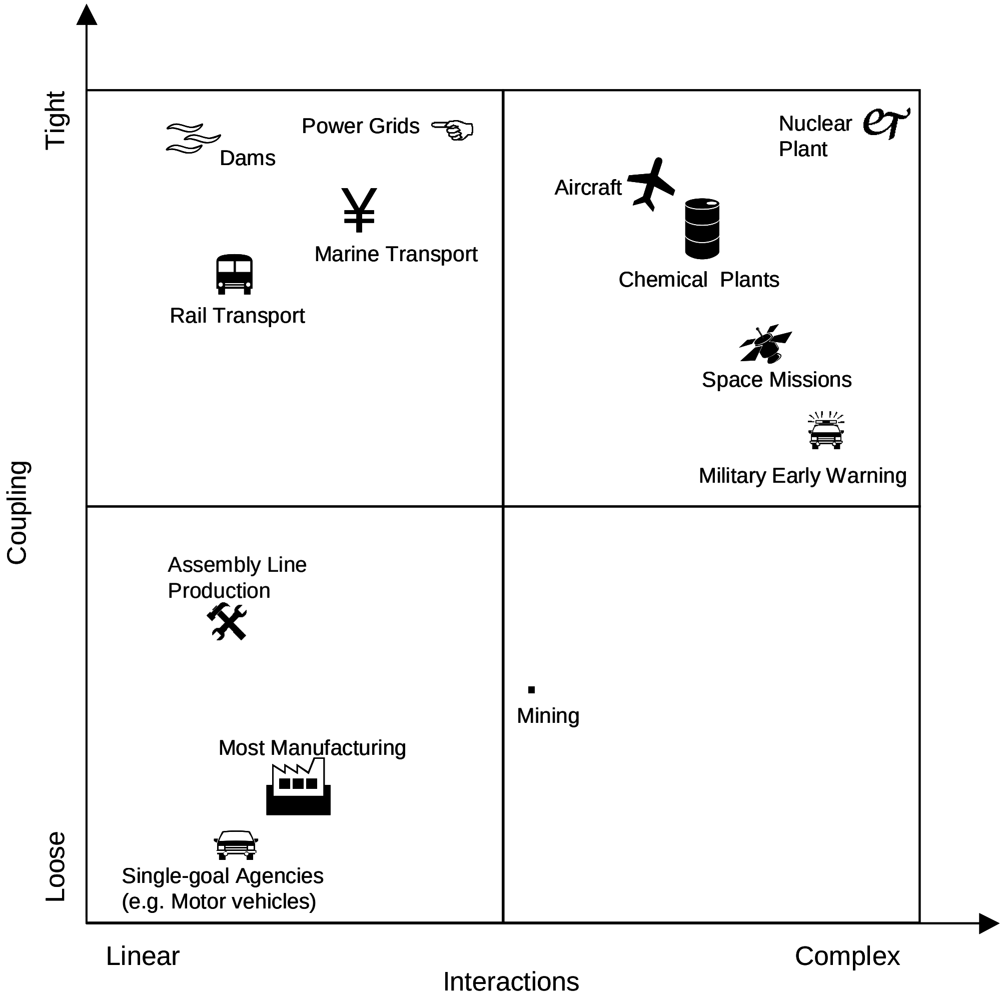


Figure 3.1 Interaction/Coupling Chart[a]

[a] Adapted from [Perrow, 1999a, Figure 3.1]

taking. He provides many examples of how redundant safety devices or human procedures can 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 (see, e.g., [Nuclearfiles, 2004]). The core was partially melted and the reactor was permanently disabled.

While Perrow's basic argument about redundancy is compelling, the use of redundancy is not the only way to increase safety and many of the alternatives do not involve increasing complexity. Redundancy and the use of protection systems are among the least effective and the most costly approaches to designing for safety [Leveson, 1995]. The most effective approaches involve eliminating hazards or 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, and interlocks of various kinds. Operations can also be made safer by eliminating and reducing the potential for human error. A simple example is the use of colour coding and male/female adapters to reduce wiring errors.

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. The consequences of accidents can be effectively mitigated without completely identifying all potential causes. For example, life boats can be used to evacuate passengers and crews from a sinking ship, regardless of the cause of the sinking.

Safer systems can be designed by limiting the interactive complexity and tight coupling in designs. Interactively complex and tightly coupled designs are created because they often allow greater functionality and efficiency to be achieved, but in some cases simpler, decoupled designs may be able to achieve similar goals. Where systems cannot be decoupled or made simpler, an awareness of the risk posed by such system can aid in developing positive safety cultures. Interactive complexity is by definition subjective and can therefore be reduced through training and experience. The better people know and understand a system the less likely they are to encounter interactions that they do not expect or understand.

3.2 High Reliability Organizations

High Reliability Organizations (HROs) are defined as the subset of hazardous organizations that enjoy a record of high safety over long periods of time[1]:

> 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’” [Roberts, 1990a, p.160].

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

Definitional problems aside, 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 [La Porte and Consolini, 1991].

The critical flaw in the HRO argument is that the systems they studied are not interactively complex nor tightly coupled according to Perrow‘s definitions. Air traffic control (ATC), for example, is as safe as it is precisely because the system design is deliberately decoupled in order to increase safety. The ATC system is carefully divided into non-interacting sectors and flight phases (en route, arrival, and takeoff and landing) with the interfaces between the sectors and phases (for example, hand-off 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).

[1] The researchers most often associated with HROs are Todd La Porte, Gene Rochlin, Karlene Roberts, Karl Weick, and Paula Consolini.

Additional warning devices, such as the traffic collision avoidance system (TCAS), are used to further reduce the likelihood of accidents.

The functions of aircraft carrier landing and take-off 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. These systems are relatively simple and loosely coupled *and* safe, thus seeming to support Perrow‘s arguments rather than contradict them (see also the afterword to [Perrow, 1999a]).

The HRO researchers emphasize the low level of complexity in the 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 [La Porte and Consolini, 1991, pp.29-30] (emphasis added).

Systems that allow perfect knowledge are by definition not interactively complex. If technical knowledge is complete, as required for HROs, it is relatively easy to lower risk through standard system safety and industrial safety approaches. Unfortunately, most complex systems, particularly high-tech systems, do not fall into this category.

The important factor here is uncertainty: technical, organizational, and social. It is uncertainty that makes engineering difficult and challenging and occasionally unsuccessful. The Space Shuttle is an example of a system at the other extreme from those studied by the HRO researchers. Technical uncertainty is inherent in any system as complex as the Shuttle. For example, although foam shedding from the external tank during liftoff has been identified as a problem for two decades, it is still not fully understood. All attempts to fix the problem have been unsuccessful. In complex systems, particularly those operating at the edge of technical innovation, there are always many technical uncertainties that cannot be resolved as required for HROs. At any time, the Shuttle has over 3000 waivers (allowing flight even though potential problems have not been completely resolved) out of a total of over 4000 Criticality 1 and 1R components[1]. Deciding which outstanding problems should be given priority is a difficult problem in itself. In addition, because many high-tech systems use new technology, understanding of the physical phenomena that may cause problems is often limited. NASA (and other organizations) cannot wait for complete understanding before launching technically complex systems. While the Shuttle is admittedly an extreme example, most high-tech systems have unresolved technical uncertainty. If it were necessary to resolve all uncertainty before use or operation, as required for HROs, most high-tech systems would never come into operation.

Organizational and social uncertainties further exacerbate the problem. The Shuttle, for example, exists in a highly uncertain political and budgetary environment. Budgets can change dramatically from year to year and 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 primarily 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 accidents do occur, despite the claims of some HRO proponents that there are no or few accidents. For the Shuttle, the task itself is not only high risk, but the technical, organizational, and social structures used to perform the task also involve significant uncertainty. The risk on aircraft carrier stems from uncertainty about individual operator behavior. The equivalent for the Shuttle is uncertainty about the behaviour of those employees manufacturing the components, preparing the Shuttles for launch, and operating in Mission Control. 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

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

decision-making environment, which does not exist in the systems that the HRO researchers studied.

Some HRO researchers claim that their example systems operate in an environment that is constantly changing [La Porte and Consolini, 1991; Roberts, 1990b], but there is little evidence of this and the quote above contradicts this assumption of constant change. Air traffic control has essentially remained the same for the past thirty 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 seventy-five years of aircraft carrier existence, only a few major changes have occurred; the greatest changes resulted from the invention of jet aircraft. The introduction of improvements in carrier aviation, such as the angled flight deck, the steam catapult, and the introduction of mirror landing systems, have occurred slowly and over long time periods.

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

The HRO researchers have identified four primary organizational characteristics that they claim substantially limit accidents and ‘failures’ and simultaneously result in high levels of performance: (1) prioritization of both safety and performance and consensus about the goals across the organization [La Porte and Consolini, 1991]; (2) promotion of a “culture of reliability” in simultaneously decentralized and centralized operations [Weick, 1987]; (3) use of organizational learning that maximizes learning from accidents, incidents, and near misses [La Porte and Consolini, 1991]; and (4) extensive use of redundancy [Rochlin et al., 1987]. The next four sections consider the practicality and usefulness of each of these for organizations trying to operate high-risk systems.

3.2.1 Goal Prioritization and Consensus

In HROs, leaders prioritize both performance and safety as organizational goals, and consensus about these goals is unequivocal [La Porte and Consolini, 1991]. While this state of affairs is clearly desirable, it is almost never possible. Safety goals usually do not coincide with performance goals (unless the sole or primary purpose of the system is to maintain safety). 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, safety is either the only or primary goal of the existence of the organization so prioritizing it is easy. For example, in peacetime aircraft carrier operations, military exercises are performed to provide training and ensure readiness. There are no goal conflicts with safety: The primary goal is to get aircraft landed and launched safely or, if that goal is not successful, to safely eject and recover the pilots. If conditions are risky, for example, during bad weather, flight operations can be delayed or canceled without major consequences.

In wartime, the carrier‘s goals are subordinated to the larger goals of the military operation. The peacetime primary goal of safely getting aircraft on and off the carrier must now be combined with additional goals from strategic planners, including speed of operations. Human safety, aircraft safety, and even carrier safety may no longer be the highest priority. Further complicating the situation is the stress of being in combat. 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 [Rochlin, 1991].

Safety is not the primary goal of most organizations. Usually, the mission of the organization is something other than safety, such as producing and selling products or the pursuit of scientific knowledge. In addition, it is often the case that the non-safety goals are best achieved in ways that are not consistent with designing or operating for lowest risk. In that case, each employee reliably performing their job is not enough to ensure safety—the employees may behave in a highly reliable manner to satisfy the non-safety goals while compromising safety in the process. Management statements that safety is the primary goal are often belied by pressures on employees to bend safety rules in order to increase production or to meet tight deadlines. An example was the issuance of “countdown to Node 2 launch” screensavers to all NASA Shuttle employees before the Columbia accident [Gehman, 2003, p. 133]. 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, 2004].

Consensus on prioritization of conflicting goals may waver when a company is forced to choose between operational safety goals (i.e., performing 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—but making difficult trade-offs and decisions about how much risk is acceptable and even how to measure the risk. Suggesting that NASA or any large organization should prioritize safety goals and operate reliably overly simplifies the difficulty in accomplishing these goals and is not very helpful in resolving the necessary trade-offs 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 most other organizations are subject to financial, political and social pressures from within and without 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 (see Section 4.3.3).

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 because the pressure to maintain supplier relationships is substantial. NASA 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, 2004].

Organizations can improve conflict resolution and implement safeguards to counteract pressures to undervalue safety, but some of the HRO arguments do not take into account the extreme difficulty of achieving this goal in practice.

3.2.2 Simultaneously Decentralized and Centralized Operations

The second characteristic of HROs is that organization members are socialized and trained to provide uniform and appropriate responses to crisis situations [Weick, 1987]. 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, while the operation of aircraft carriers is subject to the Navy‘s chain of command, even the lowest-level seaman can abort landings [La Porte and Consolini, 1991]. Clearly, this local authority is necessary in the case of aborted landings because decisions must be made too quickly to go up a chain of command. Overtraining of emergency responses is a standard practice in the training of operational personnel working in potentially dangerous, time-critical conditions. Note also that low-level personnel on aircraft carriers may only make decisions in one direction, that is, they may only abort landings. The actions governed by these decisions and the conditions for making them are relatively simple.

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

3.2.3 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 and supplements to trial-and-error learning [Weick, 1987]. 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".

HROs also try to maximize learning from accidents, incidents, and near misses [La Porte and Consolini, 1991]. While it is difficult to argue against learning from mistakes, the costs of implementing effective organizational learning are high and the problem of competition for resources arises again. In addition, the difficulty of implementing effective organizational learning should not be underestimated.

The organizations studied by HRO researchers are characterized by unchanging or very slowly changing designs and technology, which increases the effectiveness of learning from accidents and incidents. Organizations like NASA and the military that operate at the edges of technological innovation do not always have past experience from which to learn. Lessons learned on old technologies are also often inapplicable to newer ones. For example, digital systems are changing the nature of accidents and the types of errors made by operators [Sarter and Woods, 1995]. Experience with older, electro-mechanical systems does not apply to these new system designs and technology.

Organizational learning is important, but the difficulty of achieving effective learning should not be underestimated. Improved methods of risk and hazard analysis, system safety engineering, and understanding of the organizational and social factors that allow accidents to happen can be used to reduce the need to learn from accidents and ad-hoc scenario generation.

3.2.4 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," [La Porte, 1996] where redundancy is defined as "the ability to provide for the execution of a task if the primary unit fails or falters” [La Porte and Consolini, 1991]. According to Roberts, 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) [Roberts, 1990b]. 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. 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. Under these circumstances redundancy can actually increase the risk of an accident, as Perrow so persuasively argued.

The systems studied by HRO researchers are interactively *not* complex, *not* tightly coupled, and, according to their own accounts, are characterized by low levels of uncertainty. In these relatively simple, decoupled systems, redundancy can be effective in preventing a single component failure (or sometimes multiple component failures) from leading to an accident. Even in these cases, however, there are limitations. For example, common-mode failures, where supposedly independent redundant components fail due to the same cause, limit the effectiveness of redundancy in protecting against component failure. An Eastern Airlines Lockheed L-1011 descending into Nassau in 1983 lost oil pressure in all three engines simultaneously because both mechanics did not put O-rings on three newly installed engine oil plugs [NTSB, 1983]. Inadequate preventive maintenance is one type of common error that affects all components, including backups components. Redundancy depends on an assumption of random and independent component failure to be effective. But many causes of accidents in interactively complex and tightly coupled systems do not involve random or independent component failure.

The use of redundancy can lead to dangerous decision making when it encourages complacency and the need for additional safety measures is discounted. As Sagan notes, “when redundancy makes the system appear more safe, operators often take advantage of such improvements to move to higher and more dangerous production levels” [Sagan, 1993, p. 40]. The decision to launch the Challenger Space Shuttle on its fatal flight was partly based on overreliance on redundant O-rings. The failure of the primary O-ring led to the failure of the secondary O-ring [Rogers, 1986]. 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.

Redundancy is not useful in protecting against software commands that can lead to accidents. Most software-related accidents can be traced back to errors in the software requirements, that is, 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’. Software redundancy management systems are so complex that they often introduce errors and can lead to system failures themselves [Leveson, 1995].

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

***

Both NAT and HRO research oversimplify the cause of accidents. HRO underestimates the problems of uncertainty, while NAT recognizes the difficulty of dealing with uncertainty but underestimates and oversimplifies the potential ways to cope with uncertainty. The contribution of Perrow to understanding accidents in complex systems by identifying interactive complexity and tight coupling as critical factors should not be discounted. But the theory is incomplete and leads to more pessimism than required with respect to designing and operating complex high-risk systems. While the HRO researchers do offer more suggestions, most of them are inapplicable to complex systems or oversimplify the problems involved.

3.3 Normalization of Deviance

Diane Vaughan developed the theory of normalization of deviance to explain the Challenger accident, and subsequently applied it to the Columbia accident as well [Vaughan, 1996; Gehman, 2003, Chapter 8]. The theory claims that risk is "normalized" over time, so that an organization ends up accepting higher levels of risk than initially intended. While the theory has received great attention from other sociologists, it does not quite reflect the way that real engineers think about risk when designing high-risk systems. The Challenger accident occurred not because an escalated level of risk was knowingly accepted, but because some aspects of system behaviour had come to be seen as acceptable and, further, that this conclusion had been reached without adequate supporting data [Leveson, 2004b].

This section develops a critique of the theory, which is illustrated by a review of one of the occasions of 'normalization of deviance'. According to the theory, normalization of deviance occurred when members of NASA and its contractors repeatedly accepted incremental increases in risk as the following sequence of events repeated:

1. Signals of potential danger
2. Official act acknowledging escalated risk
3. Review of the evidence
4. Official act indicating the normalization of deviance: accepting risk
5. Shuttle launch

Consider the proposed process of normalization by reviewing some of the events surrounding the performance of the solid rocket booster (SRB) joints on the Space Shuttle.

The Space Shuttle Challenger was lost in 1986 when O-rings on one of the joints failed to seal, allowing an explosive mixture of hydrogen and oxygen propellants to escape. The explosion destroyed the External Tank and exposed the Shuttle to severe aerodynamic loads that caused complete structural breakup [Rogers, 1986, Chapter 3].

The O-rings in the SRB joints were not expected to show any erosion, but starting with the second Space Shuttle flight in 1981, erosion of $0.053''$ was observed on the primary O-ring of the right SRB's aft field joint[1]. According to Vaughan, this was a "signal of potential danger" (Step One). Despite the erosion, the primary O-ring sealed the gap, indicating that under similar conditions it could provide sealing in the presence of erosion of at least $0.053''$.

Systems are designed with nominal performance in mind. In complex systems it is impossible to predict performance precisely based solely on design parameters. It is therefore expected that the actual performance will deviate from the nominal performance. That is why performance requirements are usually not phrased in absolute terms, but rather in terms of upper and lower bounds of acceptable performance. Remedial action is necessary only when the actual performance falls outside the specified performance limits. System developers then have three options: they can redesign the relevant parts of the system to bring performance closer to the desired performance; they can renegotiate the performance requirements; or they can do a combination of redesign and requirements adjustment. Acknowledging a deviation from nominal performance is not the same thing as acknowledging an "escalation of risk".

Engineers at NASA and Thiokol did not expect to see erosion because none had ever been observed in tests on Titan rockets, on which the SRBs were based. But erosion was not an unknown phenomenon in the aerospace industry and it was therefore not seen as conclusive evidence that the sealing ability of the O-rings was compromised. Subsequent investi-

[1] The erosion on STS-2 proved to be the worst observed on a primary O-ring in a field joint in any recovered solid rocket booster [Rogers, 1986, Chapter. 6].

gation (Step 3) of the erosion indicated that it was caused by a localized deficiency in the zinc chromate putty that lined the space between the booster segments. Engineers at Morton-Thiokol believed that the reason for the erosion had been determined and they began testing the method of putty lay-up and the effect of the assembly of the rocket stages on the integrity of the putty [Rogers, 1986, Chapter. 6]. Contrary to Vaughan’s assertion, the erosion was not seen as a “signal of potential danger”. Rather, it was an unexpected but not unprecedented deviation in performance that nevertheless still fell within the performance requirements of the O-rings. It is easy to identify events such as deviations from expected performance as “signals of potential danger” in hindsight by tracing backwards from an accident. But complex systems will experience countless such deviations. Determining which ones, if any, could lead to accidents is difficult and often impossible. When actual performance falls within the limits of acceptable performance, it is even more difficult to identify this deviation as a potential contributor to an accident.

Continuing the case of erosion on STS-2, note that the risk that the O-rings would not seal did not change because erosion was observed. There was no “escalation of risk” (Step Two). The O-rings did not suddenly become less likely to seal, and the consequences of not sealing did not suddenly increase. What changed was the evidence relating to this risk. Vaughan’s use of the terms “construction of risk” and “negotiation of risk” is unfortunate. Risk is not constructed or negotiated, it is an inherent feature of a system. How well the assessed risk compares to the actual risk depends on the type of system, how well the technology is understood, the risk analysis methods, and so forth. An improved assessment of risk does not mean that the level of risk has changed, it simply means that the assessed risk is now closer to the actual risk. How well the perceived risk compares to the actual risk depends on who is doing the perceiving, what their relationship is to the risk analysts and to the system, how the risk is presented to them, and so forth[1]. What is negotiated is the level of acceptable risk, although the huge uncertainty associated with systems like the shuttle make it impossible to set a firm boundary or determine with certainty

[1] There is a vast literature on risk perception. See [Slovic, 2000] for an extensive discussion.

whether a given risk is above or below this boundary. In the case of the O-ring erosion on STS-2, the perceived risk did not increase, because engineers believed that the worst-case erosion would not compromise sealing of the joint. There was no perception of “escalated risk”. While the deviation from nominal performance was accepted (as is standard), there was no acceptance of increased risk (“Step 4”) because neither the actual nor the perceived risk increased.

There was one official acknowledgement that risk was higher than previously believed. In December 1982, the O-rings were reclassified from Criticality 1R to Criticality 1. The earlier classification indicated that failure of the O-rings could lead to loss of the Shuttle, but that there was a redundant O-ring. The reclassification was made on the basis of analyses of joint rotation, which indicated that under certain conditions the primary O-ring would not be able to provide sealing [Rogers, 1986, Chapter 6]. Despite the reclassification, many engineers continued to believe that the O-rings did provide redundant sealing. This belief was partially responsible for the decision to launch Challenger. But the engineers’ continuing belief was not a case of “normalization of deviance” either, it was simply a case of overreliance on redundancy.

Normalization of deviance argues that by accepting incremental increases in risk, organizations eventually end up accepting higher levels of risk than they would have at the beginning of system development. The Challenger story does not support this theory however. In most cases, deviations in performance were examined and found to be within the limits of acceptable performance. The problem was that these analyses were incomplete and were improperly communicated to a management audience that had preconceived notions that the O-rings did not pose a flight risk. The engineers tasked to investigate the problem encountered administrative resistance and had difficulty convincing management of the seriousness of the problem [Winsor, 1988]. By the eve of the Challenger launch they were still unable to prove unambiguously that the O-rings would not seal. But these problems had no relation to “escalation of risk” or “normalization of deviance”.

The proposed process of normalization of deviance is an oversimplified and inaccurate representation of system development. Real systems are developed in an atmosphere of uncertainty. Properly understanding the risks posed by deviations from expected performance often requires extensive technical analyses. Such analyses require management support and may require extensive resources and time. Engineers and technicians may therefore find themselves in a difficult situation where they are unable to obtain the necessary managerial support because they can only obtain the necessary supporting data indicating the seriousness of the problem by doing the analysis. Neither the Challenger nor the Columbia accidents resulted from a knowing acceptance of increased risk. In both cases, the underlying technical phenomena (O-ring erosion and foam impacts) were poorly understood and decision makers unknowingly accepted high risks.

3.4 Summary

This chapter has reviewed three popular sociological theories of organizational risk, namely normal accidents theory, high reliability organizations, and normalization of deviance. HRO and normalization of deviance underestimate the challenges posed by uncertainty, while NAT recognizes the difficulty of dealing with uncertainty but underestimates and oversimplifies the potential ways to cope with uncertainty. NAT and normalization of deviance do not offer practical suggestions for dealing with uncertainty and developing safer systems. While the HRO theorists do offer more suggestions, most of them are inapplicable to complex systems or oversimplify the problems involved. The next chapter examines the problem of risk in complex socio-technical systems from both an engineering and an organizational perspective. Chapter 5 proposes a new approach to analyzing and managing risk in these systems.

Chapter 4

ORGANIZATIONAL RISK DYNAMICS

> *The machine does not isolate man from the great problems of nature but plunges him more deeply into them.*
>
> Antoine de Saint-Exupéry[1]

> *People know what they do; they frequently know why they do what they do; but what they don't know is what what they do does.*
>
> Michel Foucault[2]

One of the worst industrial accidents in history occurred in December 1984 at the Union Carbide chemical plant in Bhopal, India. The accidental release of methyl isocyanate (MIC) resulted in at least 2000 fatalities, 10 000 permanent disabilities (including blindness), and 200 000 injuries [Shrivastava, 1992]. The Indian government blamed the accident on human error in the form of improperly performed maintenance activities. Numerous additional factors involved in the accident can be identified. But further analysis shows that the plant had been drifting over a period of many years toward a state of high-risk where almost any change in usual behaviour could lead to an accident [Leveson, 1995]. A better understanding of risk therefore requires understanding how systems migrate towards states of increasing risk.

This chapter examines the dynamics of risk in organizations. First, a framework is developed to analyse the strategic trade-off between short and long term goals and understand

[1] [Saint-Exupéry, 1992]

[2] [Foucault, 1970]

why organizations tend to migrate to states of increasing risk. Next, a set of archetypes of organizational safety are developed. These archetypes describe specific mechanisms by which organizations unintentionally or unknowingly increase or fail to decrease risk, despite their best intentions.

4.1 Organizational Dynamics and Complex Goal Environments

Organizations that operate complex systems have to make trade-offs between multiple, interacting, sometimes conflicting, and often changing goals at both the individual and organizational levels. Resolving the conflict between goals is difficult because the potential outcomes (e.g., revenues, costs, and risks) of different courses of action are often poorly understood, uncertain, or ambiguous. Resource pressures (e.g., time, money) limit the ability to clarify the situation or reduce uncertainty, further complicating the trade-off process. In some cases, goals may directly oppose one another. For example, operators may be required to work faster to increase throughput. However, they may also be required to perform delicate tasks that require high precision, which is enhanced by working more slowly. If the organization is tightly coupled and interactively complex [Perrow, 1999a], the situation is further complicated because it is difficult to predict the consequences of actions and it is difficult to determine which information is relevant to the situation. When an organization is in a crisis mode, appropriately resolving conflicting goals becomes even more difficult because organizational and individual resources are stretched ever more tightly. This makes it more likely that inappropriate decisions will be made thus further escalating the crisis situation [Woods and Cook, 1999]. Which goals were most important and what was the most appropriate way to meet these goals is often visible only in hindsight.

Maintaining an acceptable level of risk in complex goal environments is difficult for a number of reasons, one of which is that safety goals are often poorly articulated (i.e., what is an acceptable level of risk and how should it be achieved) and the long-term effects of performance-related decisions on safety are often not obvious. Understanding how goal conflicts arise and how they can be resolved is the first step towards formulating a robust strategy to successfully resolve the apparent conflict between performance and safety.

This section discusses some of the inherent tensions between short and long term organizational goals, how this tension can result in inappropriate resolution of performance and safety goals, and how organizations can develop a strategy that maintains performance and safety over the long term.

4.1.1 Organizational Efficiency and Thoroughness

The individual approach to coping with complex goal environments can be seen as a trade-off between efficiency and thoroughness: “On the one hand people genuinely try to do what they are supposed to do—or at least what they intend to do—and to be as thorough as they believe is necessary. On the other hand they try to do this as efficiently as possible, which means that they try to do it without spending unnecessary efforts or wasting time” [Hollnagel, 1993]. For example, by omitting apparently unnecessary steps in a procedure, throughput and thus efficiency can be increased. Because short-term performance pressures dominate, people and organizations tend to be more efficient and less thorough. Rasmussen noted that organizations move toward the boundaries of safety under pressures to maintain economic performance and reduce workload [Rasmussen, 1997]. This adaptation results in migration to a system state where any small deviation in behaviour can lead to an accident. While Hollnagel's discussion refers primarily to individual activities, it can be applied at the organizational level by replacing individual efficiency and thoroughness with organizational analogues.

Organizational efficiency refers to those aspects of organizational behaviour that promote meeting performance goals (at least in the short term) such as productivity, defect-elimination, on-time delivery, quality, cost and rapid product development. Organizations can improve their efficiency by changing aspects of their organizational design such as organizational structure (e.g., how functions are allocated to departments), processes (e.g., manufacturing and accounting procedures), and composition (e.g., types of skills). Similarly, they can change the design of the systems they operate to obtain better performance (e.g., faster, more fuel-efficient aircraft). Available resources and market size limit efficiency because some activities are more efficient when performed in larger volumes (economies of scale).

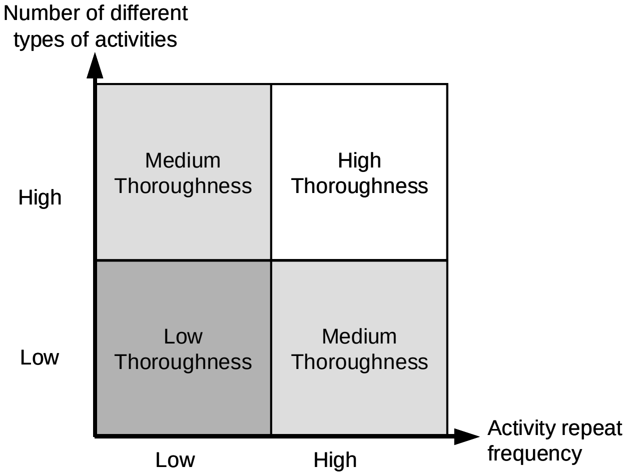


Figure 4.1 Levels of Thoroughness

Organizational thoroughness refers to those aspects of organizational behaviour that promote meeting long-term goals like safety or sustained growth. In the safety context, organizational thoroughness therefore refers to performing activities that promote the safety of the systems the organization manufactures or operates. These activities can take place at any stage of the system lifecycle. For example, hazard analysis identifies and classifies hazards during system development. During operation, root cause analysis can promote learning from accidents and help organizations improve their safety. Some activities, such as hazard analysis and root cause analysis, are appropriate for all systems. Appropriate additional safety activities depend on the stage of the lifecycle, the type of organization, and the type of systems that the organization operates. Additional activities should be defined depending on the characteristics of the system (e.g., code inspection in software companies).

An organization with a high degree of thoroughness is one that performs many relevant types of safety activities, and performs these activities frequently, as shown in the top right hand corner of Figure 4.1. Conversely, an organization with a low degree of thoroughness is one that performs few types of safety activities, and does not perform these activities frequently, as shown in the bottom left hand corner of the figure. Organizations that perform few types of activities frequently, or many types of activities sparingly, are characterised as having a medium degree of thoroughness.

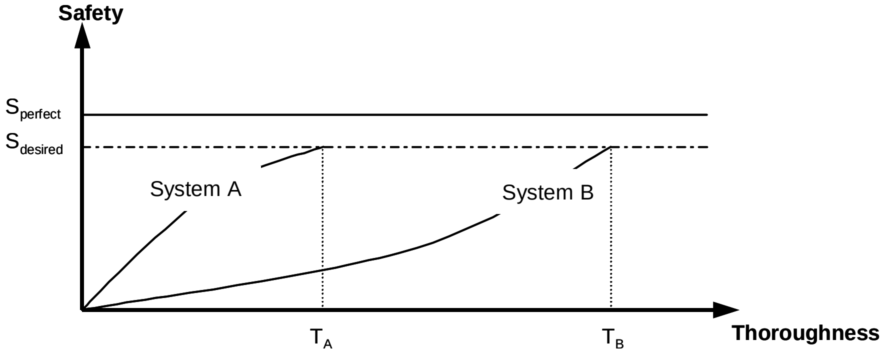


Figure 4.2 The relationship between thoroughness and safety is system dependent

Not all organizations need to exhibit a high degree of thoroughness to achieve a given level of safety. The degree of thoroughness required in order to achieve the required level of safety is system dependent, as shown in Figure 4.2. For example, fuel lines on a motor vehicle need not be checked every day—if the vehicle loses power the occupants will generally not be harmed. On the other hand, if the engines of an aircraft do not receive fuel, disaster will likely result. Aircraft fuel lines must be checked frequently. Note though that continuing to apply one type of safety activity more intensively does not necessarily result in increased safety, and may even paradoxically result in decreased safety, as discussed in Section [Amalberti, 1996]. Once a given safety activity has been applied to its maximum extent, additional different activities may be necessary to further improve safety. For example, preventative maintenance on its own cannot guarantee safety, no matter how well it is performed. Additional activities such as personnel training and root cause analysis are also required.

The Thoroughness-Efficiency Space for Organizational Design

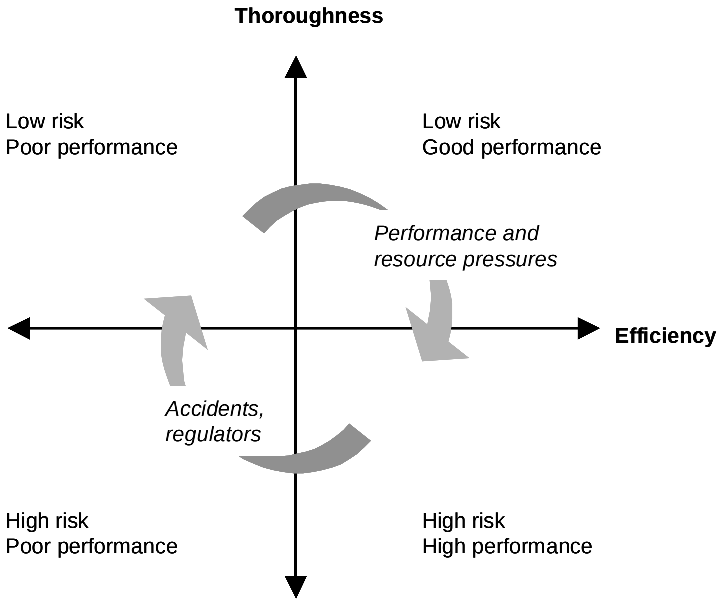


Figure 4.3 The Thoroughness-Efficiency Space

In the short term efficiency and thoroughness do not complement each other: they are orthogonal, as shown in Figure 4.3. In the short term, activities that promote performance and activities that promote safety tend to work against each other. The diagram illustrates two aspects of the tension between efficiency and thoroughness. First, it classifies organizations according to the degrees of efficiency and thoroughness that they exhibit. The sustainable organization balances performance and safety goals, as shown in the upper right hand quadrant of the graph, labelled “low risk, good performance”. Such an organization is realistic about the performance it can achieve given its resources, design, and the necessary level of safety[1]. The worst type of organization is neither efficient nor thorough and is therefore exposed to high risk and performs poorly, as shown in the lower left quadrant, labelled “high risk, poor performance”. Organizations that focus on short-term performance while skimping on thoroughness fall in the lower right hand quadrant, labelled “high risk, high performance”. These organizations perform well in the short-term, but because they are insufficiently thorough they are exposed to high risk and may eventually experience a disastrous accident. Conversely, organizations that exhibit a high degree of thoroughness and low efficiency fall in the upper left quadrant, labelled “low risk, poor performance”. These organizations have a low risk exposure to the detriment of poor performance. Organizations may temporarily operate in this quadrant while they determine how to increase performance without compromising safety. For example, following the Challenger and Columbia accidents, shuttle flights were suspended while changes were made to the organizational and system designs to improve safety.

Organizational Dynamics in the Thoroughness-Efficiency Space

Figure 4.3 also illustrates the pressures that cause organizations to move from one quadrant to another. Resource and performance pressures push organizations away from thoroughness (e.g., reducing training programs) and towards efficiency (e.g., increased productivity demands). Resource pressures occur at the organizational level when resources are limited and the organization does not or cannot scale its goals down commensurately. Organizations are almost always subject to resource pressures. Government agencies such as NASA have their budgets set by Congress, usually below the amount requested. Performance pressures include pressure to improve productivity, decrease development times, and develop better products. Faced with intense public criticism of the International Space Station, NASA pressured employees to meet ISS deadlines with a “Countdown to Node 2 Launch” screensaver [Gehman, 2003, p. 133]. The screensaver reinforced the message that meeting the schedule requirements was paramount. Perfor-

[1] Acceptable performance and risk levels must be set by the organization. In the case of regulated industries, the regulatory agencies determine acceptable risk levels. Note that while performance is usually unambiguous and easily measured and quantified, risk levels are not so easily measured or quantifiable.

mance pressures may come from within the organization or from outside. For example, public companies must maintain a healthy financial scorecard in order to keep satisfying their shareholders and Wall Street. Organizations must carefully choose how to translate, propagate, or dampen, external pressures into internal pressures. For example, they may push back on the external source by requesting additional resources or by limiting what they had committed to delivering.

Consider now how resource and performance pressure push organizations away from thoroughness. Begin with a sustainable organization in the upper right hand quadrant. This organization is performing well and is sufficiently thorough to maintain risk at an acceptable level. It is difficult for organizations to move into or remain in this quadrant. Performance and resource pressures tend to push organizations away from or out of this quadrant, towards short-term profitability and high risk, in the lower right hand quadrant. Next consider an organization in the lower right hand quadrant. This organization is performing well but is exposed to high risk. Organizations operate in this quadrant because short-term performance goals such as on-time delivery tend to dominate, thus driving the organization to emphasise performance at the detriment of safety. Resource pressures (e.g., financial, personnel, time) are experienced most sharply in the present. Because the value of safety practices or measures is not always clearly visible and is not easily measured, safety usually has a lower priority compared to other goals such as performance or efficiency. Benefits from investments in safety tend to emerge only in the long run, and may only be indirectly observable, as non-accidents or the avoidance of modifications or retrofits to improve safety [Leveson, 1998]. However, the costs (time, financial, performance, etc.) of safety practices or measures can usually be measured. Faced with intangible long-term benefits but visible short-term costs, it is understandably tempting for project managers to knowingly or tacitly compromise on safety. Emphasising performance goals works in the short term because it is usually possible to simplify procedures and omit safety activities without immediately increasing risk to the point where an accident is imminent. If the probability of accidents is low then organizations can be both efficient and safe in the short term. Over time, however, the drive towards greater efficiency increases risk, pushing the system closer to the border of safe behaviour [Rasmussen, 1997].

Pressures resulting from accidents and from industry regulators push organizations towards thoroughness and away from efficiency. Not all pressures result in improved safety however. Accident investigation recommendations may be inadequate, incomplete, or even inappropriate. For example, a common finding is that procedures were not followed, leading to a recommendation to enforce procedures more strictly. But this type of finding ignores the systemic factors that led to procedures being violated in the first place. Recommendations and requirements from regulators may also be inappropriate. Even when pressures, recommendations and requirements are appropriate they may be short-lived. In the aftermath of accidents, organizations do try to improve their safety. But this goal is usually replaced by other goals, such as productivity, when the memory of the accident fades. Pressures from regulators may also fade away between inspection cycles. If inspections occur infrequently, there may be substantial periods when the organization is not subject to any regulatory pressure.

Because short-term pressures dominate, organizations generally move into the top left hand corner of Figure 4.3 (low risk, poor performance) only under duress. For example, following a serious accident they may make an effort to improve their safety. Or regulators, as is the case with nuclear energy, may impose safety requirements on the organization as a condition of operation. In the long term, investment in safety is economically justifiable. But it is difficult for organizations to operate in a long-term manner, because of the immediacy of performance goals and resource pressures.

4.1.2 The Performance–Safety Barrier

An organization’s design, the design of the systems it operates, and its available resources determine how much emphasis it can place on performance while maintaining the necessary level of safety. A conceptual performance-safety barrier is defined, as shown in Figure 4.4. The shape of the curve is defined by noting that increasing emphasis on safety detracts from short-term performance. Note that there is a maximum possible performance given a level of available resources. Decreasing emphasis on safety further does not increase performance, because risk is increased to the point where accidents occur often enough to detract from short-term performance. The performance–safety barrier can be moved outward towards greater efficiency by increasing the level of resources or by changing the system design (e.g., more efficient and safe motor vehicles).

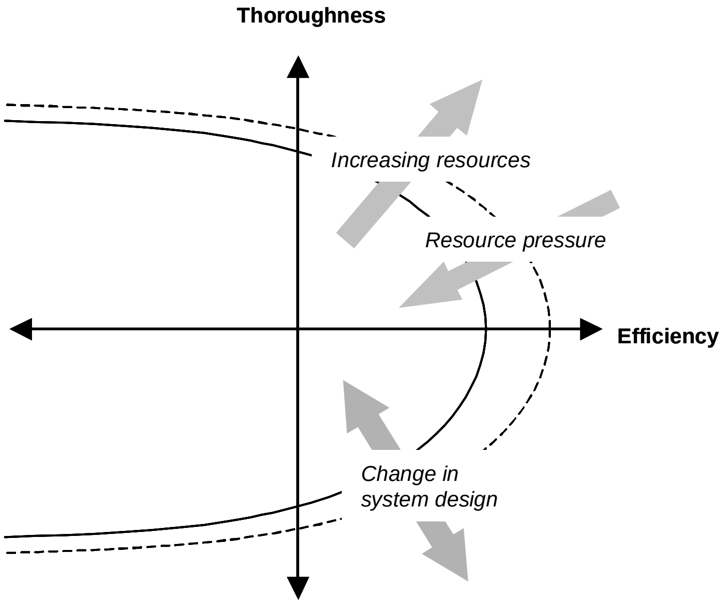


Figure 4.4 The Performance-Safety Barrier

4.1.3 Balancing Safety and Performance

Emphasising performance goals to the detriment of safety goals may work in the short term. But in the long term, continuing this emphasis can result in potentially disastrous accidents, as illustrated by the Bhopal accident. Organizations that wish to survive in the long term should operate in the upper right hand quadrant of the thoroughness–efficiency space where both performance and safety goals are met. The performance–safety barrier limits the maximum efficiency obtainable for the required level of safety and therefore the necessary degree of thoroughness, thus creating a sustainable operating space, as shown in Figure 4.5.

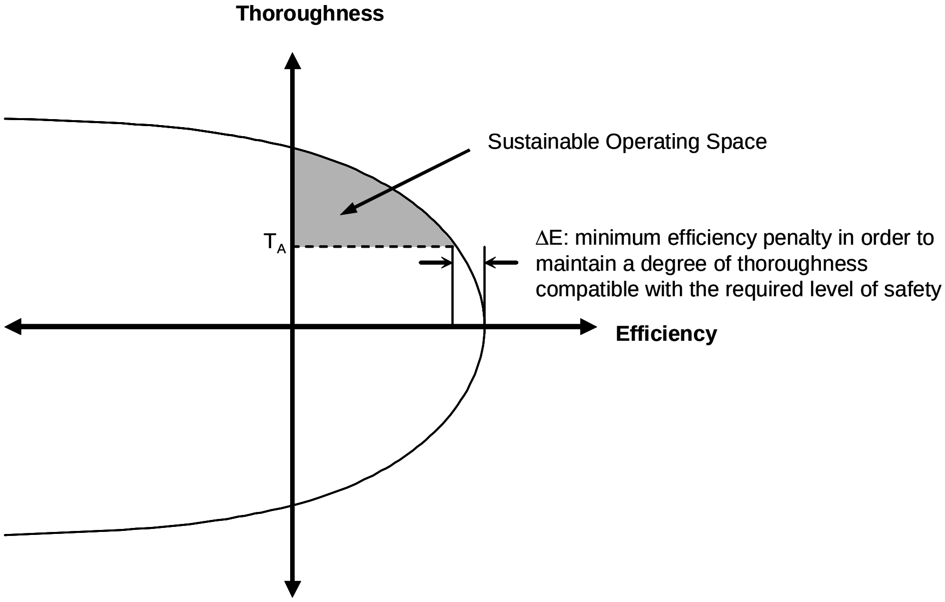

Thoroughness
Sustainable Operating Space
T_A
ΔE: minimum efficiency penalty in order to maintain a degree of thoroughness compatible with the required level of safety
Efficiency

Figure 4.5 The sustainable operating space for an organization operating System A as shown in Figure 4.2.

The desired, or ideal, operating point, as shown in Figure 4.6, is the point where an organization maximises performance while maintaining the required level of safety[1]. If an organization’s performance and safety goals place it outside the performance-safety barrier, either the barrier must be moved outwards (by increasing the level of resources or changing the system design), or the performance and/or safety goals must be revised. Conversely, if the organization is operating inside the performance-safety barrier, higher performance is possible without decreasing thoroughness. Note that while the desired operating point optimises performance and safety, an organization should operate slightly inside the barrier because of uncertainty in determining the desired and actual operating

[1] Note that it may be difficult to measure the level of safety. The operating point is a theoretical construct that illustrates the trade-off between efficiency and thoroughness.

points and to provide resilience in the face of short-term variations in the operating point and performance-safety barrier.

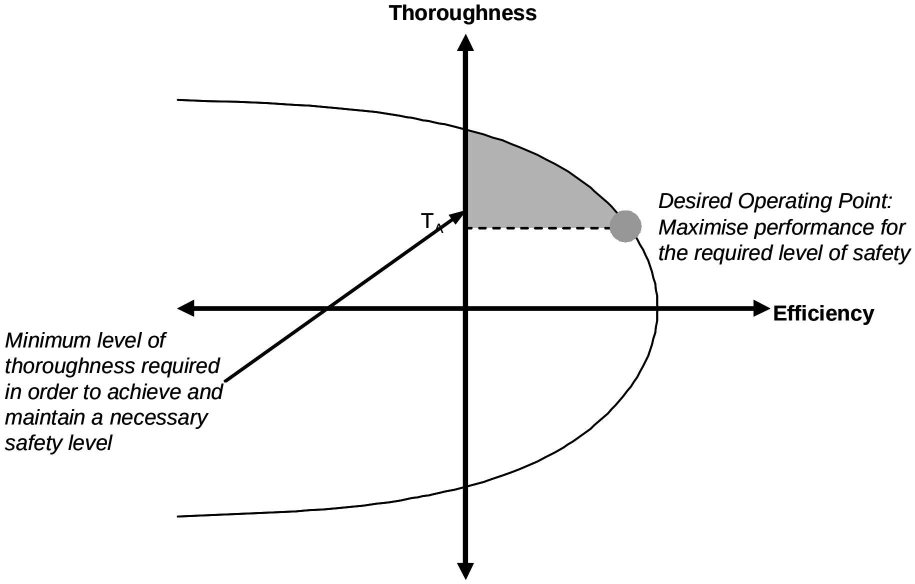

Thoroughness
T_A
Desired Operating Point: Maximise performance for the required level of safety
Efficiency
Minimum level of thoroughness required in order to achieve and maintain a necessary safety level

Figure 4.6 Balancing Safety and Performance

In practice, the trade-off between performance and safety goals, resources, and system design is usually not made explicitly. Organizations are often not conscious of performing trade-offs [Woods and Cook, 1999], or of the criteria upon which these trade-offs are made. When the desired performance is not attained, safety goals are often tacitly or explicitly traded for higher performance. The possibility of obtaining the desired performance by increasing the level of resources, or changing the system design, may be ignored, or discounted because of cost and other considerations. The criteria by which trade-offs are made are not always visible either. They may be explicit, implicit, or emergent properties at various levels of the organization. They may be susceptible to influence and change over time, or they may be firmly fixed and inflexible in the face of changing conditions. By becoming aware that they are making trade-offs, organizations can be pro-active about identifying trade-offs, formulating strategies to make these trade-offs, and determining where they want to lie in the efficiency–thoroughness space.

4.1.4 Migration to Boundaries

Migration toward states of increased risk, as discussed earlier, occurs when performance and resource pressures cause the organization to emphasise efficiency at the detriment of thoroughness, thus moving it to the edge of safety. Continuous effort is needed to ensure that performance and resource pressures do not result in a decrease in thoroughness that moves the organization away from safety. Changes that occur over time in the system move the performance-safety barrier and may result in the organization operating below the minimum thoroughness threshold, as shown in Figure 4.7. The organization is therefore faced with two tasks: monitoring where it lies with respect to the barrier, and ensuring that it remains inside (for safety), and preferably on (for maximum performance), the barrier.

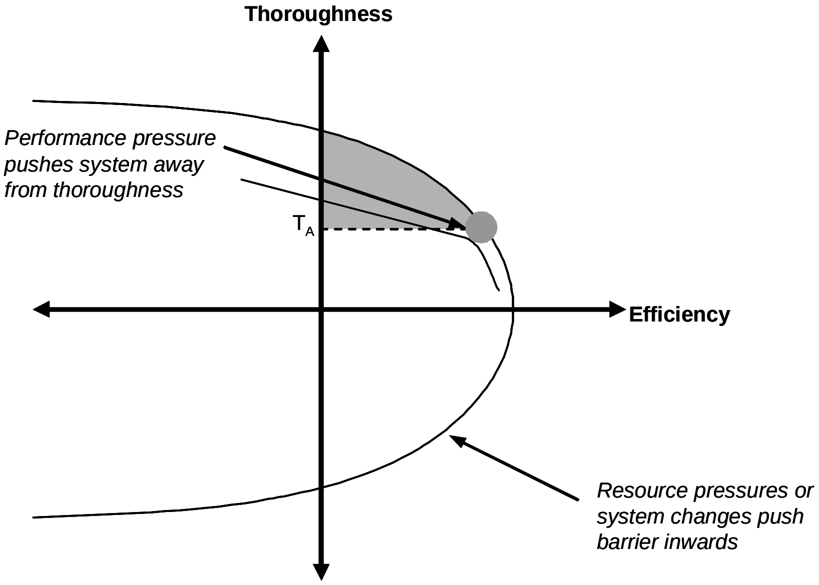

Thoroughness
Performance pressure pushes system away from thoroughness
T_A
Efficiency
Resource pressures or system changes push barrier inwards

Figure 4.7 Migration from the Desired Operating Point

Determining where an organization lies with respect to the barrier and the thoroughness threshold requires a combination of risk management and performance monitoring. A description of performance monitoring is beyond the scope of this thesis, but the literature provides numerous references. Chapter 5 describes one approach to assessing and monitoring the level of risk.

Ensuring that an organization remains inside the barrier requires a positive safety culture, as described in Appendix B. A positive safety culture can be seen as "pulling the slippery slope" up, as shown in Figure 4.8, making it easier for the organization to withstand performance and resource pressures and to resist changes in the system that make it less safe.

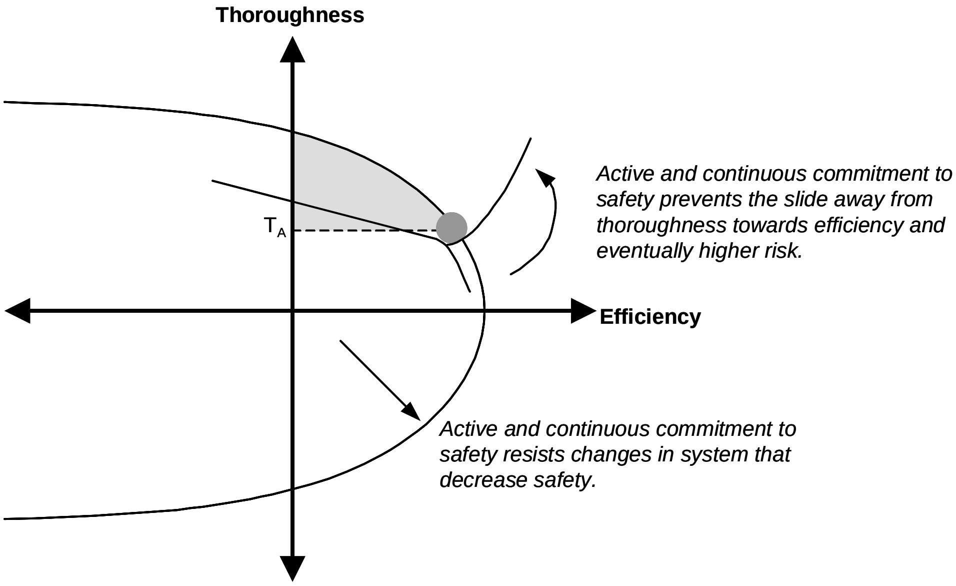

Thoroughness
T_A
Efficiency
Active and continuous commitment to safety prevents the slide away from thoroughness towards efficiency and eventually higher risk.
Active and continuous commitment to safety resists changes in system that decrease safety.

Figure 4.8 A positive safety culture prevents sliding away from safety

This discussion has shown how tension arises between safety and performance because performance is measured in the short term, while safety, or the lack thereof, is only observed over the long term. When viewed in the short term, safety and performance goals tend to promote opposing actions. By taking a long-term view the tension can be resolved. An awareness of the often implicit trade-offs between safety and performance can empower organizations to avoid decisions that gradually push the system towards the boundary of safe behaviour.

4.2 Organizational Safety Archetypes

While individual accidents usually have unique features at the surface, further probing often reveals common underlying systemic patterns. By identifying these patterns, or archetypes, organizations can better understand past accidents, monitor risk, and decrease the likelihood of future accidents.

This section introduces the concept of safety archetypes. General system behavioural archetypes have been described by various authors in fields such as system dynamics [Braun, 2002; Wolstenholme, 2003] and organizational behaviour [Masuch, 1985; Miller and Friesen, 1980]. While the general archetypes apply to all behaviour, the safety archetypes developed here address specific behaviour related to flaws in an organization's safety processes and culture.

In risk analysis the archetypes can be used to understand how and why the level of risk changes over time, as discussed in Chapter 5. They explain how undesired side-effects arise from apparently good decisions, why organizations become complacent, and why it is difficult for organizations to successfully implement safety improvement programs. An awareness of these pitfalls can help organizations avoid them or at least decrease their negative impact.

In accident analysis, the archetypes can be used to develop dynamic models that describe the systemic and organizational factors contributing to the accident. The archetypes help clarify why safety-related decisions do not always result in the desired behaviour, and how independent decisions in different parts of the organization can combine to impact safety.

The archetypes are explained using elements of the system dynamics modelling language. A brief review is provided below.

4.2.1 Brief Overview of System Dynamics

System dynamics is an approach to identifying, explaining, and eliminating problem behaviours in socio-economic systems, primarily by identifying feedback loops in the system. It provides a framework for dealing with dynamic complexity. Whereas the controllers used in engineered feedback control systems typically employ negative feedback, socio-economic and natural systems may exhibit both negative and positive feedback loops. System dynamics is grounded in the theory of nonlinear dynamics and feedback control, but also draws on cognitive and social psychology, organization theory, economics, and other social sciences [Sterman, 2002a].

With its explicit recognition of the time dimension, system dynamics is designed to address the problem of dynamic complexity. This type of complexity refers to our inability to predict the often counterintuitive behaviour of complex systems over time.

The archetypes are constructed from three basic building blocks: the reinforcing loop, the balancing loop, and the delay.

Reinforcing Loop.

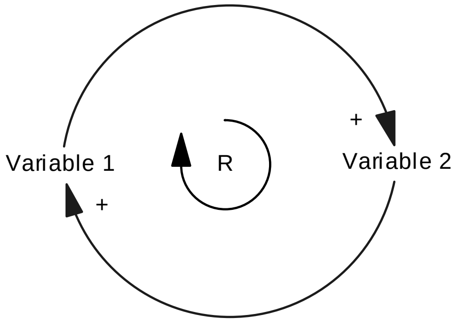


Figure 4.9 Reinforcing Loop

A *Reinforcing Loop* is a structure that feeds on itself to produce growth or decline. It corresponds to a positive feedback loop in control theory. An increase in State 1 causes an increase in State 2, as indicated by the '+' sign, which in turn causes an increase in State 1, and so on. In the absence of external influences, both State 1 and State 2 will grow or decline increasingly rapidly. Because initial growth/decline is often slow, it may be unnoticed until it becomes rapid, at which point it may be too late to control the growth/decline. Reinforcing loops "generate growth, amplify deviations, and reinforce change" [Sterman, 2000].

Balancing Loop

A *Balancing Loop* is a structure that attempts to move a current state to a desired or reference state through some action. It corresponds to a negative feedback loop in control theory. The difference between the current state and the desired state is perceived as an error. An action proportional to the error is taken to decrease the error, so that, over time, the current state approaches the desired state. While the reinforcing loop tends to display growth or decline, the balancing loop tends to settle down to the desired state. Because the size of the remedial action is proportional to the size of the error, the current state initially rapidly approaches the desired state. As the error decreases, the rate with which the current state approaches the desired state decreases.

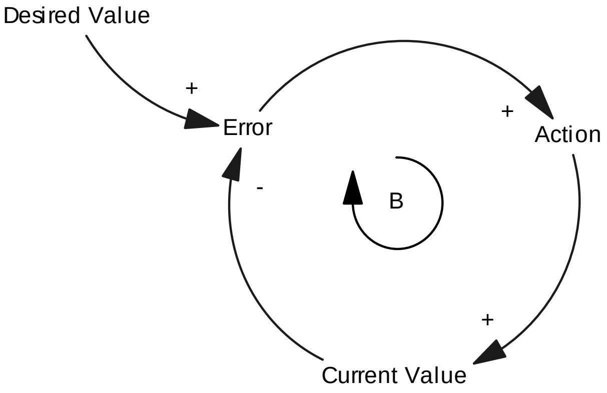


Figure 4.10 Balancing Loop

Delay

Delays are used to model the time that elapses between cause and effect, and are indicated by a double line, as shown on a balancing loop in Figure 4.11.

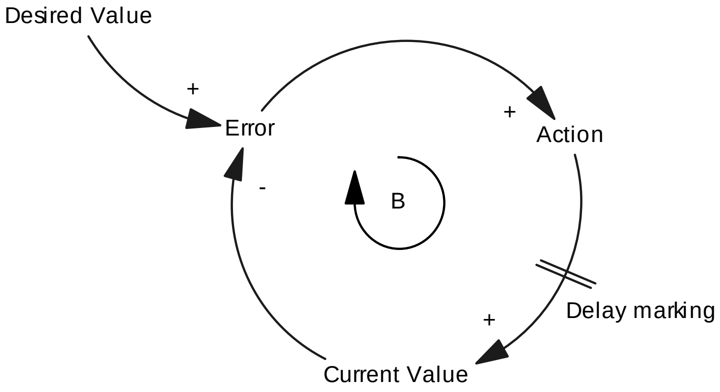


Figure 4.11 Balancing Loop with Delay

Delays make it difficult to link cause and effect (dynamic complexity) and may result in unstable system behaviour. Consider, for example, the problem of navigating a ship down a narrow channel. Suppose that the ship is veering to one side of the channel, and the helmsman wishes to correct the course. Due to the ship's inertia, adjusting the rudder will not result in an immediate course change. There is a delay between a change in the rudder position and the resulting course change. In stressful situations, even experienced helmsmen may interpret a delayed response as a complete lack of response, and accordingly make a larger change in the rudder position. When the ship's inertia is eventually overcome, the helmsman finds himself sailing towards the opposite side of the channel. If the helmsman continues to over-correct in this way, the ship may veer wildly from one side of the channel to the other, and may run aground.

***

The next sections introduce the following sets of archetypes:

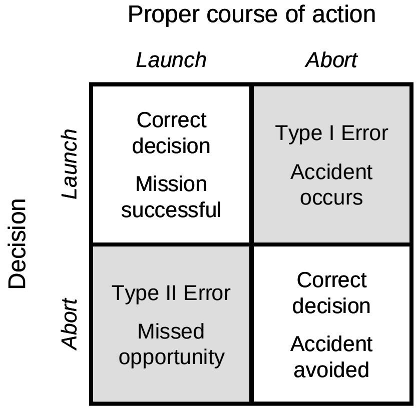
1. Cycles of Error
2. Challenges of Maintaining Safety:
- Stagnant Safety Practices in the Face of Technological Advances
- Decreasing Safety Consciousness
- Eroding Safety Goals
- Complacency
3. Side-Effects and Symptomatic Responses:
- Unintended Side Effects of Safety Fixes
- Fixing Symptoms Rather Than Root Causes
- The Vicious Cycle of Bureaucracy
4. Challenges of Successfully Addressing Root Causes:
- The Short-Term Performance Trap
- Employee Commitment

4.2.2 Cycles of Error

This archetype demonstrates why organizations may oscillate between periods of few accidents and incidents and periods of many or serious accidents.

Decisions about whether to take an action or not can be wrong in two ways: action is taken when it should not be taken, or action is not taken when it should be taken. By analogy to hypothesis testing, taking an inappropriate action is a Type I error, while not taking an appropriate action is a Type II error [Bendor, 1985]. Figure 4.12 illustrates the concept in the case of deciding whether to launch the Space Shuttle.

Avoiding the possibility of Type I errors necessarily requires accepting some Type II errors and vice versa. A Type I error such as an unsafe launch decision can have obvious and tragic consequences. But Type II errors are wasted opportunities and can be costly in terms of lost revenue, wasted resources, and the impact on other projects. In the framework of the discussion in Section 4.1, Type I errors occur when efficiency is pursued to



Decision	Proper course of action: *Launch*	Proper course of action: *Abort*
Launch	Correct decision Mission successful	Type I Error Accident occurs
Abort	Type II Error Missed opportunity	Correct decision Accident avoided

Figure 4.12 Type I and II errors[a]

[a] Adapted from [Heimann, 1993]

the detriment of thoroughness, while Type II errors occur when thoroughness is pursued to the detriment of efficiency. In the case of the Shuttle, repeated launch scrubs delay construction of the International Space Station as well as numerous scientific experiments. Some scientific opportunities may even be lost. When NASA missed the launch date for the ASTRO mission, the opportunity to study Halley’s comet was lost for at least seventy-six years. The inability to avoid both Type I and II errors simultaneously is one manifestation of the tension between performance and safety discussed in Section 4.1.

In complex systems, it is not possible to eliminate the possibility of either Type I or II errors because there is never complete certainty that an accident will or will not occur. Organizations tend to cycle between Type I and II errors [Heimann, 1997], as demonstrated by the *Cycles of Error* archetype (Figure 4.13).

Consider an organization that begins with a concern to *avoid Type I errors*. This focus on avoiding accidents results in an increasing number of *missed opportunities* over time. As more and more opportunities are missed, performance and cost-effectiveness pressures eventually cause the focus to change from avoiding Type I errors to *avoiding Type II errors*, thereby increasing the probability of *accidents*. When an accident does occur, the focus shifts back to avoiding Type I errors, and the cycle begins again ($R_{\text{cycles of failure}}$).

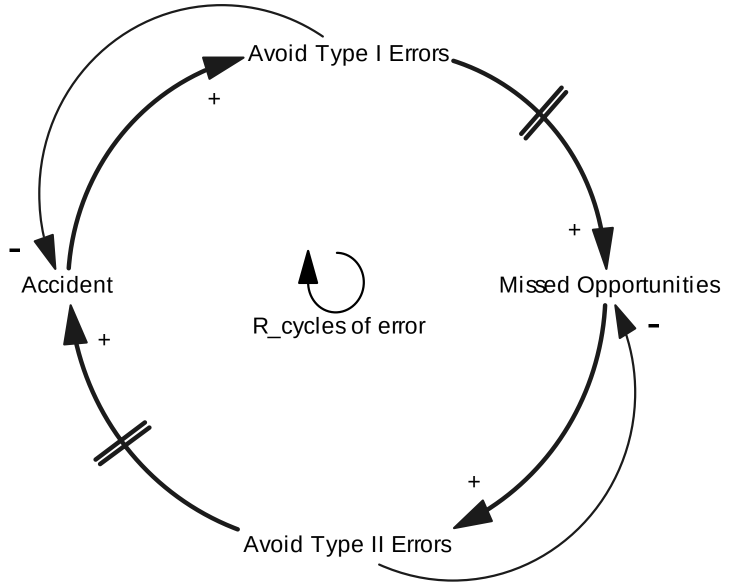


Figure 4.13 Cycles of Error

At NASA this behaviour is exemplified by the change from the culture of “prove it’s safe” during the Apollo program and in the early days of the Shuttle program to the culture of “prove it’s not safe” that prevailed before both the Challenger and Columbia accidents. Both accidents were preceded by an atmosphere in which maintaining launch schedules had become increasingly important. For example, before the Columbia accidents, employees were issued with a “Countdown to Node 2” screensaver that reinforced the message that meeting the ISS construction milestones was more important than other goals, despite management claims to the contrary. NASA reacted to both accidents by cancelling all launches while the accidents were investigated, recommendations made, and efforts made to decrease the likelihood of future accidents. But the increased emphasis on safety following the Challenger accident rapidly faded away and set the stage for the Columbia accident seventeen years later. *Cycles of Error* indicates that history is likely to repeat itself again.

Type I errors can be tragic and have far-reaching effects. *Cycles of Error* can be used to increase awareness of risk and fight against the complacency (see *Complacency*) that arises from repeated success.

4.2.3 Challenges of Maintaining Safety

The next four archetypes illustrate the challenges of maintaining safety over long periods of time.

Stagnant Safety Practices in the Face of Technological Advances

When technological advances are not accompanied by concomitant understanding of the associated risks, safety may be compromised, as shown in Figure 4.14. This structure consists of a reinforcing loop ($R_{\text{growing performance}}$) and two balancing loops ($B_{\text{decreasing safety}}$ and $B_{\text{lagging understanding}}$).

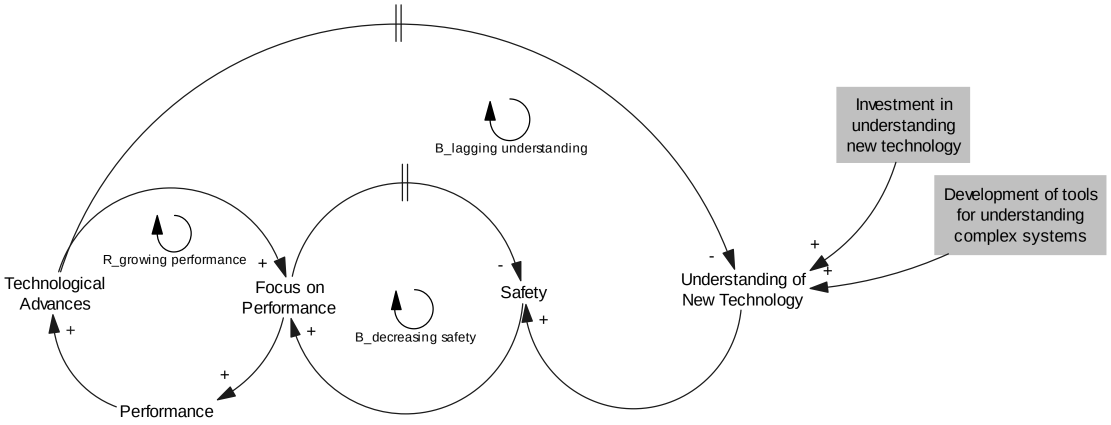


Figure 4.14 Stagnant Safety Practices in the Face of Technological Advances

The constraint on safety is the understanding of the risk associated with the new technology and the resulting risk associated with the systems in which it is embedded. Technological advances result in a focus on performance and a corresponding increase in performance, which in turn motivates more advances ($R_{\text{growing performance}}$). At the same time, the focus on performance detracts attention from safety ($B_{\text{decreasing safety}}$). As the speed of change accelerates, understanding of the safety implications lags further behind ($B_{\text{lagging understanding}}$).

One particular area of concern is software. Software is becoming an increasingly significant part of most systems. Unfortunately, the field of software engineering has not kept pace with the uses to which software is put. Numerous aerospace accidents have involved software which behaves in a way that could have been foreseen but was not because the software, system, and safety engineering functions did not detect the problems [Leveson, 2004b]. For example, the Mars Polar Lander is believed to have crashed when the software on the Mars Polar Lander erroneously interpreted a spurious signal from the spacecraft legs as indicating that the spacecraft had landed and shut the engines down prematurely, causing the spacecraft to crash into the Martian surface [Albee, 2000].

The problem of stagnant safety practices can be ameliorated by applying new technologies only when their risks are understood, investing more resources in the understanding of new technologies, and by developing tools for understanding complex systems.

Decreasing Safety Consciousness

The success of a safety program may be limited by the characteristics of the system to which the program is applied, or by the nature of the program itself. This archetype illustrates how a strategy, policy, or process that initially promotes improved safety may eventually reach a point where its continued application cause a decline in safety (Figure 4.15). Incident reduction measures may initially improve system safety ($R_{\text{reduce incidents}}$). But the absence of incidents (near-misses, unscheduled downtimes, etc.) renders the system mute, and situational awareness of the system is decreased. The result is a decrease in system safety ($B_{\text{awareness limits safety}}$).

Consider the case of ultra-safe systems such as commercial air travel. Common sense tells us that in order to increase safety, errors, incidents and breakdowns must be reduced or eliminated. This is true for systems where the rate of incidents and accidents is high. In the case of ultra-safe systems, continued elimination of errors, incidents, and breakdowns may

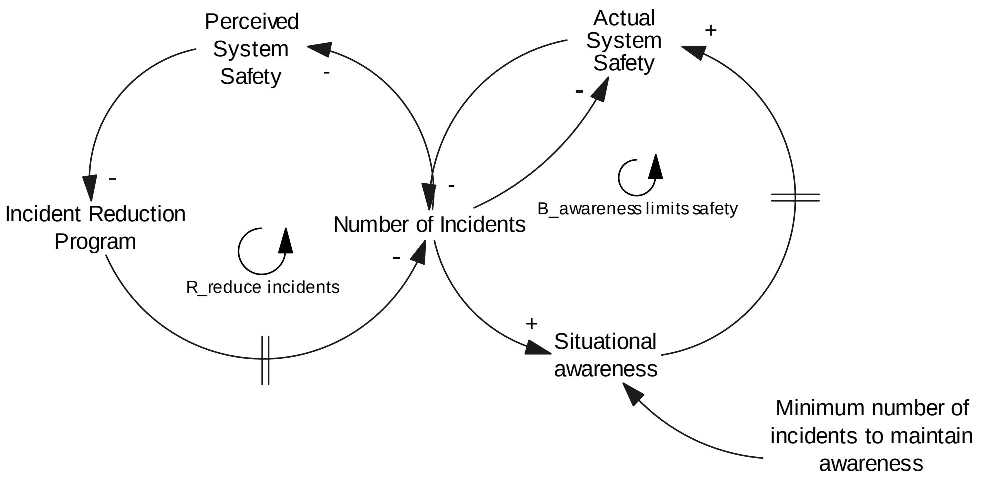


Figure 4.15 Decreasing Safety Consciousness

paradoxically decrease safety [Amalberti, 1996]. Amalberti argues that the combination of a system with a given set of safety measures bears within itself a maximum safety potential, which cannot be exceeded by continued optimization of those safety measures. Continued optimization of a particular safety measure mutes some system aspects, thereby decreasing system awareness and adversely affecting safety. To obtain further increases in safety beyond this limit, additional, new safety measures are necessary. Therefore, to maintain safety, safety measures must be aggregated, but no single safety measure should be overly optimised.

Over-optimization numbs the adaptive capabilities of human and technical systems, while covering up minor system failures. In the case of error reduction, for example, it has been found that error plays an ecological role in the control of performance, and that detected errors are necessary to maintain situational awareness. Similarly, programs to reduce the number of incidents and breakdowns may also perversely decrease safety. As the perceived level of safety increases, the temptation is strong to redirect investments away from safety measures and towards improving system performance. Over-stretched system performance leads to new risks, which may materialise in the form of disastrous accidents [Rasmussen, 1997]. Beyond a certain incident reduction quota, the absence of incidents, as opposed to the presence of a minimum number of incidents, does not prevent accidents from occurring. Information that can only be gained from incidents is lost when all incidents are eliminated. Responding to incidents provides organizations with the motivation to adapt and increases organizational resilience [Sitkin, 1992]. It may therefore be necessary to tolerate a certain level of errors, incidents, breakdowns, and even accidents to protect the system against disastrous accidents and prepare the organization for responding to accidents if they do occur.

Another response to the problem is to gradually lower the incident detection threshold as the number of incidents is reduced, thus maintaining a certain minimum number of incidents. Investigations into incidents have several uses and benefits [cf. Carroll, 1998b]. First, they can uncover the causes of the particular incident. Second, they can uncover root causes that may lead to other incidents and even accidents. Third, they encourage members of the organization to think about how their role in the organization affects safety. Fourth, they maintain the organization's ability to investigate incidents. By setting the incident detection threshold sufficiently low so that incident investigations do not cease entirely, organizations can continue reaping the benefits of incident investigations.

Consider the strong emphasis on redundancy as a safety and reliability measure in many systems. Some degree of redundancy is useful in increasing reliability, and possibly safety. But more redundancy is not necessarily better, and may be worse [e.g., Sagan, 2004]. While redundancy may increase reliability, it does not necessarily increase, and may even decrease, safety. First, a reliance on redundancy may lead to decreased emphasis on other safety engineering techniques. If system designers believe that redundancy will limit the effect of design errors they may be less motivated to find and eliminate these errors. In practice, redundancy may 'cover up', or mute, design errors and prevent them from becoming visible until something catastrophic occurs. Second, increasing redundancy increases system complexity. More complex systems are less amenable to testing and maintenance, and their properties and behaviour are difficult to predict accurately [Graham, 1971].

The use of redundancy can lead to dangerous decision making when it encourages complacency and the need for additional safety measures is discounted [Marais et al., 2004]. The decision to launch the Challenger Space Shuttle on its fatal flight was partly based on overreliance on redundant O-rings. The failure of the primary O-ring led to the failure of the secondary O-ring [Rogers, 1986]. The overconfidence inspired by the redundancy convinced decision makers that the Shuttle would survive a cold-weather launch even if the primary O-ring failed. Redundancy led to decreased safety consciousness at the exact time when concern for safety was most needed because redundancy on its own was not enough to improve system safety.

Eroding Safety Goals

This archetype illustrates how safety goals may erode or become subverted over time. *Eroding Safety Goals* behaviour often precedes accidents, but is generally only observed in hindsight. *Eroding Safety Goals* is difficult to observe while it is occurring because change tends to happen gradually. At short time scales, changes may be imperceptible. It is only after an accident has occurred that the extent of change is noticed, if at all.

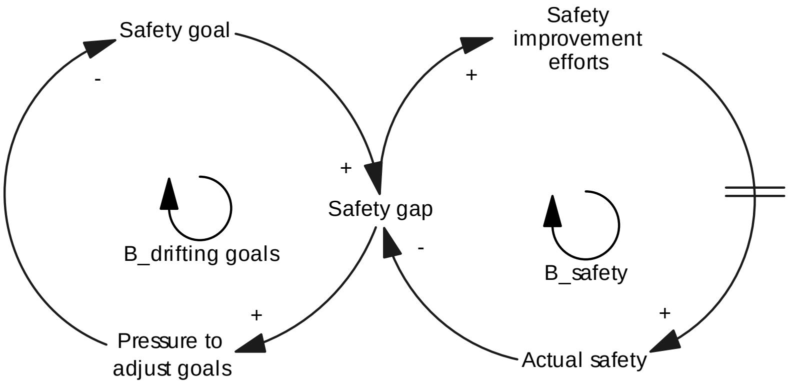


Figure 4.16 Eroding Safety

Figure 4.16 illustrates the basic structure. A *safety gap* between the *safety goal* and *actual safety* inspires *safety improvement efforts*, which improve *actual safety*, but usually not immediately (B_{safety}). The *safety gap* can also be decreased by adjusting the *safety goal* downwards. The greater the gap, the greater the *pressure to adjust goals* ($B_{\text{drifting goals}}$). Because $B_{\text{drifting goals}}$ makes the safety gap smaller, B_{safety} becomes less effective at maintaining safety at the required level.

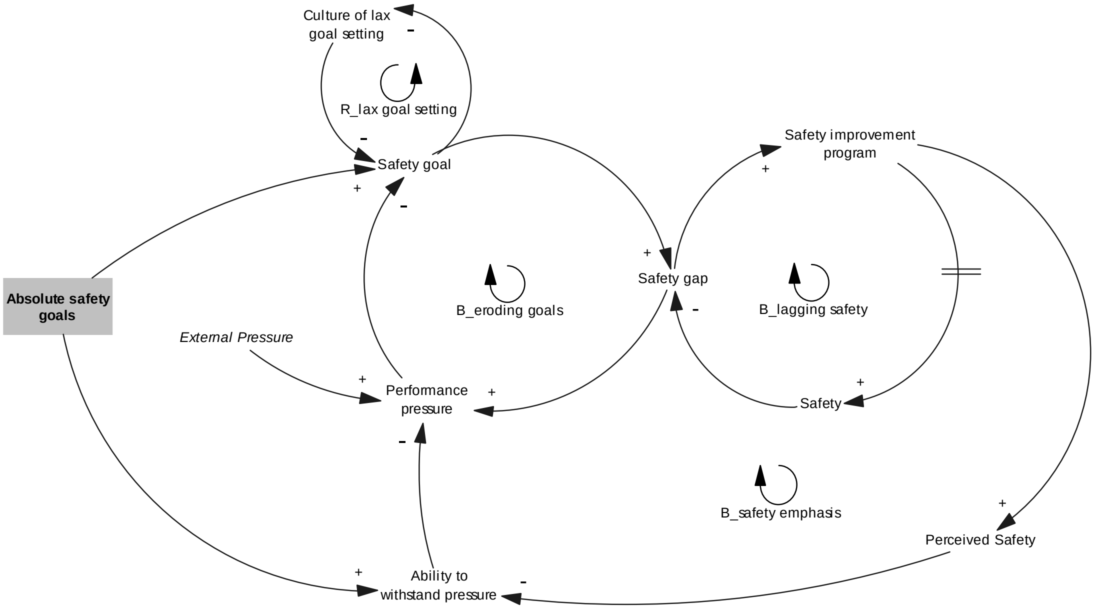


Figure 4.17 Disappointing Safety Programs

Disappointing Safety Programs. *Eroding Safety Goals* illustrates one reason why safety programs do not always live up to expectations (Figure 4.17). Safety improvement programs can be expensive and often do not show immediate results ($B_{\text{lagging safety}}$). While the eventual costs of not improving safety can be high, the immediate cost of a safety program is subject to external pressures (e.g., budget and performance pressure). The combination of seeming ineffectiveness and external pressures makes it tempting to place less emphasis on safety and adjust the goals of the safety program ($B_{\text{safety emphasis}}$ and $B_{\text{eroding goals}}$). This adjustment is not necessarily seen as a failure, and may even be viewed as an improvement. These balancing loops interact to repeatedly lower the safety goal. Repeated lowering of safety goals results in a reinforcing dynamic ($R_{\text{lax goal setting}}$) that encourages lax goal setting in the future. The problem can be addressed by setting absolute safety goals, perhaps based on some external standard. Such external safety goals will only be effective however if organizations adhere to these standards, either of their own accord, or because regular inspections, audits, and/or punitive measures force them to do so.

For example, a common response to failed programs is to restructure parts of, or the entire organization in question. After the Challenger accident NASA responded by reorganising the safety and quality programs at NASA Headquarters and the field centers. A new office of Safety, Reliability, Maintainability and Quality Assurance (SRM&QA) was established and overall management of the safety function was elevated to the level of associate administrator, in an attempt to increase awareness of significant safety and quality issues at the highest levels of NASA management. This reorganization was presented as one of the most significant improvements following the Challenger accident [NASA, 1988]. This reorganization failed to achieve its goal over the long term, and many of the same “silent safety program” characteristics were made evident by the Columbia accident. While restructuring and reorganization is sometimes necessary, it does not always address the underlying problem.

Another, more subtle form of downward goal adjustment is the eternally receding deadline. In this case, the goals remain the same, but the deadline for meeting the goals is continually shifted back, effectively lowering the goals.

Pressure for increased performance (e.g., shorter delivery times, increased profit) can make it difficult to remain focussed on safety goals. *Eroding Safety* illustrates how these pressures can contribute to safety improvement goals not being met. The challenge is to resist external pressures that work against safety improvement programs, whether overtly or in a less obvious manner. Anchoring the safety goals to externally generated and enforced standards or deadlines can make adjustments in goals more visible or more difficult to make. For example, government regulators impose certain minimum safety standards on some industries, such as the nuclear power industry.

To be successful a safety program must provide a clear plan and a realistic time frame for improving safety. It must provide concrete steps towards achieving the safety goal, as well as interim measures of progress. If a safety program is seen as working against performance (e.g. preventing on-time delivery of goods) there will be a reciprocal tendency to work against the program, thereby decreasing its effectiveness. Managers who pay lip service to safety programs but simultaneously demand increased performance encourage a lax attitude to safety at lower organizational levels. Only when there is buy-in at all levels of the organization can a safety program succeed.

Complacency

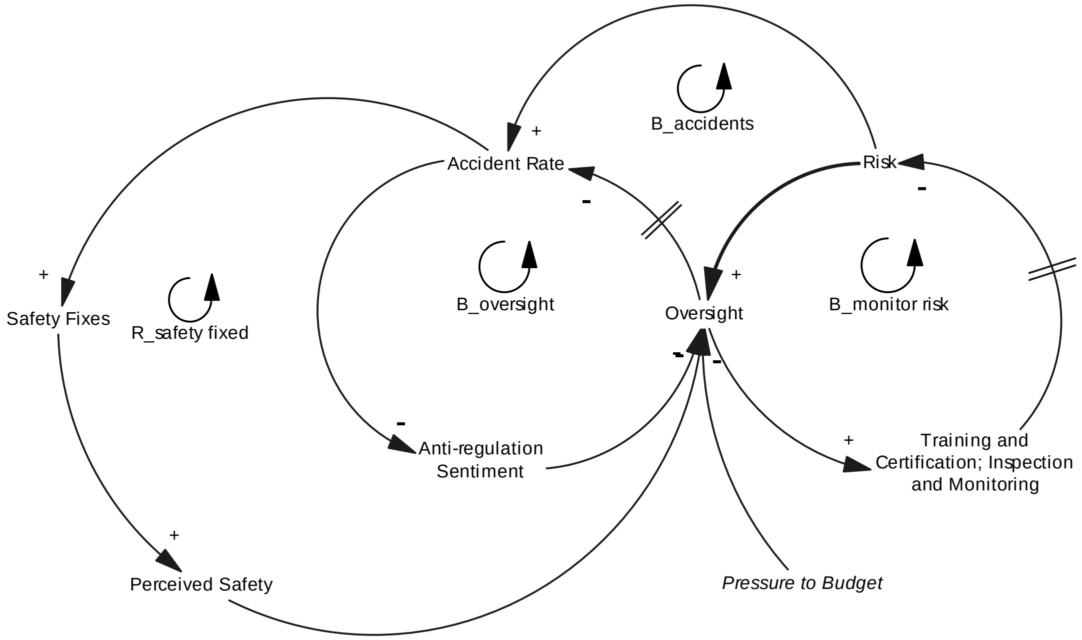


Figure 4.18 Complacency

A history of operations without incidents often leads to growing complacency as shown in Figure 4.18. Consider a system that initially operates with a high accident rate. In order to bring the accident rate down, the system is closely monitored, possibly both internally (company rules and procedures) and externally (government regulation). Close *oversight* may eventually contribute to a decreased *accident rate*, and may bring it to the point where people do not believe that accidents can or will occur. In the apparent absence of a threat to safety, extensive oversight may seem draconian and unnecessarily costly. Coupled with budgetary pressures, this anti-regulation sentiment creates pressure to decrease oversight ($B_{oversight}$). Decreased oversight is manifested on the one hand by less *training* and fewer or less strict *certification requirements*, and on the other hand by decreased *inspection and monitoring*. A decrease in these activities eventually leads to an increase in the risk of accidents, and so the accident rate increases ($B_{accidents}$). *Safety fixes* implemented in response to accidents increase *perceived safety* and thus encourage the relaxation of *oversight* ($R_{safety\ fixed}$). One way to avoid the complacency trap is to continuously monitor risk (thus maintaining awareness that risk is not negligible) and set the level of oversight accordingly ($B_{monitor\ risk}$).

Following the Apollo launch pad fire in 1967, NASA established one of the best system safety programs of the time [CETS, 1993]. But nearly two decades later the Rogers Commission report on the Challenger accident referred to a "Silent Safety Program" that had lost some of its effectiveness since Apollo. In particular, the report cited growing complacency at the agency, as the perception grew that Shuttle operations were routine [Rogers, 1986]:

> Following successful completion of the orbital flight test phase of the Shuttle program, the system was declared to be operational. Subsequently, several safety, reliability and quality assurance organizations found themselves with reduced and/or reorganized functional capability… The apparent reason for such actions was a perception that less safety, reliability and quality assurance activity would be required during 'routine'' Shuttle operations. This reasoning was faulty. The machinery is highly complex, and the requirements are exacting… As the system matures and the experience changes, careful tracking will be required to prevent premature failures… Complacency and failures in supervision and reporting seriously aggravate these risks.

Improved safety consciousness at NASA in response to the Challenger accident rapidly deteriorated as a series of successful missions helped the memory of the accident fade. Many engineers were highly conscious of the risks and even tried warning NASA and the government. But neither NASA management nor the government appear to have taken these concerns seriously. When Columbia was launched in January 2003, managers were much more concerned about maintaining the launch schedule than about safety. Although a team was assembled to investigate the foam strike that occurred during the launch, they received little support from management, who appeared to have already decided that the foam strike did not pose a flight risk.

The problem with complacency is twofold. First, it is difficult not to become complacent when success follows upon success. Second, it is difficult for an organization to realise that it is becoming complacent, and often a serious accident is required to shake the complacency.

Complacency arises because the accident rate usually does not immediately increase following a decrease in oversight. Inertia in the system temporarily keeps the accident risk at a low level, creating the impression that oversight is set at the appropriate level. All the while, the system is migrating toward the boundary of safe behaviour [Rasmussen, 1997]. When accidents start occurring, the link to decreased oversight is not immediately obvious. When making the connection between risk and the level of oversight, the long-term trend in the risk level must be considered, rather than short-term fluctuations.

4.2.4 Side-Effects and Symptomatic Responses

The next set of archetypes addresses the problem of poorly designed (safety) improvement efforts. The first archetype illustrates how the unforeseen side-effects of safety improvement efforts may worsen the problems these efforts are intended to address. The second archetypes shows how addressing symptoms and not root causes not only does not solve the underlying problem, but may actually make it worse. The third archetype illustrates how inappropriate responses to organizational challenges can lead to stifling rules and apathy.

Unintended Side Effects of Safety Fixes

The unintended consequences of poorly designed responses to safety problems, whether they are symptomatic treatments or supposed fundamental solutions, can worsen the problem.

This structure consists of a balancing loop (B_{fix}) and a reinforcing loop ($R_{side\text{-}effects}$). The loops interact so that the desired result initially produced by the safety fix in the balancing loop is, after some delay, offset by the undesired side effects in the reinforcing loop. Initially, the *safety fix* ameliorates the *safety problem* (B_{fix}). After a delay, the *unintended system reaction* becomes visible, worsening the problem and accordingly the safety fix is applied more strongly ($R_{side\text{-}effects}$). The safety fix ironically contributes to the worsening of the problem.

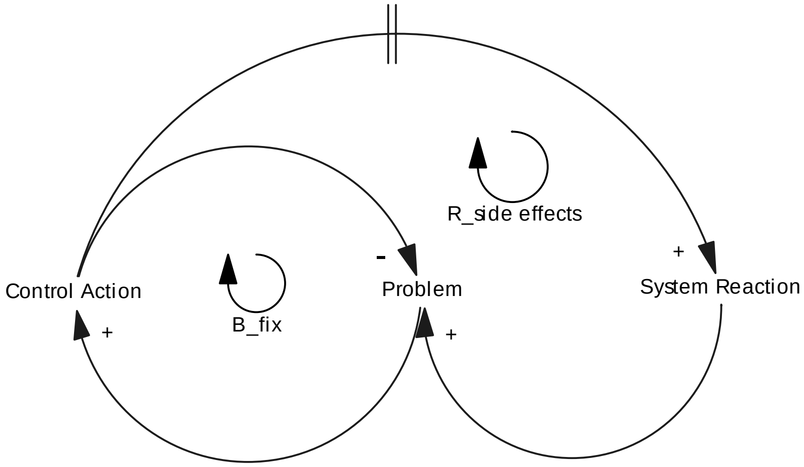


Figure 4.19 Unintended side-effects of safety fixes

Well-intentioned, commonplace solutions to safety problems often fail to help, have unintended side effects, or exacerbate problems. The example below illustrates how disciplining workers and writing more detailed procedures may fail to reduce the number of equipment breakdowns.

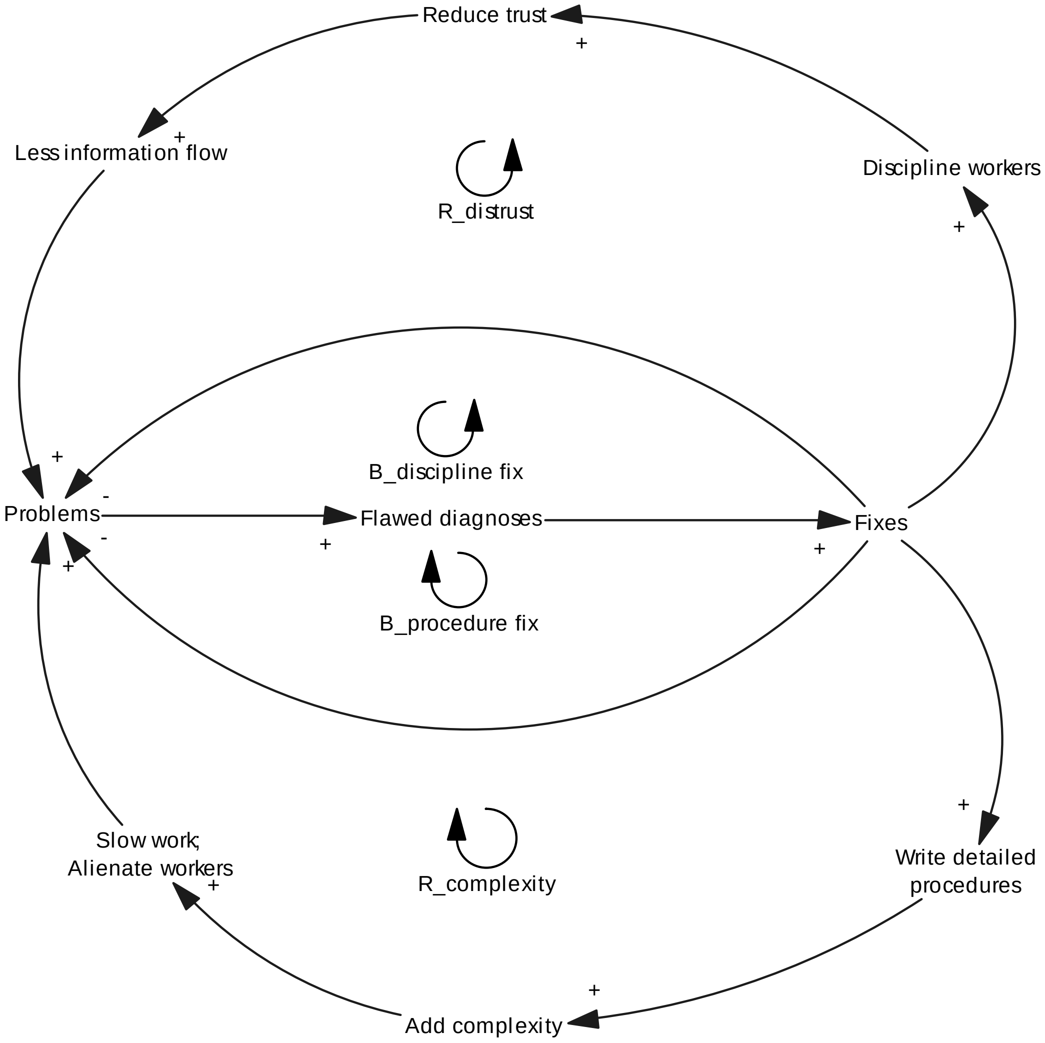


Figure 4.20 Unsuccessful Problem Resolution[a]

[a] Adapted from [Carroll, 1998b]

Consider a plant that is experiencing increasing equipment breakdowns, which are attributed to poor maintenance (Figure 4.20) [cf. Carroll et al., 1998]. A typical 'fix' for maintenance-related problems is to write more detailed maintenance procedures and to monitor compliance with these procedures more closely ($B_{\text{discipline fix}}$ and $B_{\text{procedure fix)}}$. But these fixes often result in reinforcing loops (R_{distrust} and $R_{\text{complexity}}$) that eventually make the problems worse. More detailed procedures can translate to fewer errors in a particular task. Workers tend to view more detailed procedures and closer supervision as mistrust and regimentation, causing them to lose motivation, or comply blindly or maliciously with procedures that may be incomplete or incorrect. Skilled workers may find the new regime intrusive and look for more interesting work elsewhere. Excessive restrictions on behaviour discourages problem-solving and encourages blind adherence to procedures, even when such compliance is not optimal in terms of safety or productivity. Blaming or disciplining individual workers, designed to create an atmosphere of accountability, encourages all workers to hide problems. For example, when the Federal Aviation Administration provided immunity from prosecution to pilots who reported near-collisions, the number of reports tripled; when immunity was later retracted, the number of reports decreased six-fold [Tamuz, 1994]. When incidents are deliberately concealed, the underlying problems do not become visible, often worsen, and may lead to more problems.

Unintended Side Effects behaviour occurs when the fundamental problem is not understood, or when the solutions to the fundamental problem are not appropriate or are improperly implemented. This behaviour can be avoided by correctly identifying the fundamental problem and designing appropriate solution strategies. Identifying the fundamental problem is often difficult, and designing and implementing solution strategies can be challenging. An awareness of the long-term negative implications that fixes often have can provide the impetus to search for fundamental solutions instead.

Fixing Symptoms Rather Than Root Causes

This archetype is a special case of *Unintended Side Effects* that shows how symptomatic solutions, or fixes, can undermine the ability to determine and address root causes (Figure 4.21). A *fix* is implemented in response to a *problem* (B_{fix}), temporarily decreasing the problem symptoms. The fix may also make it more difficult to identify the fundamental problem and/or implement a fundamental solution: If the fundamental solution is known, *side effects* of the fix may either decrease the desire to implement the fundamental solution, or act to decrease the effectiveness of the fundamental solution ($R_{\text{side effects}}$). If the fundamental problem or solution is not known, the symptomatic solutions may decrease the ability to find the fundamental solution, for example by masking the problem symptoms.

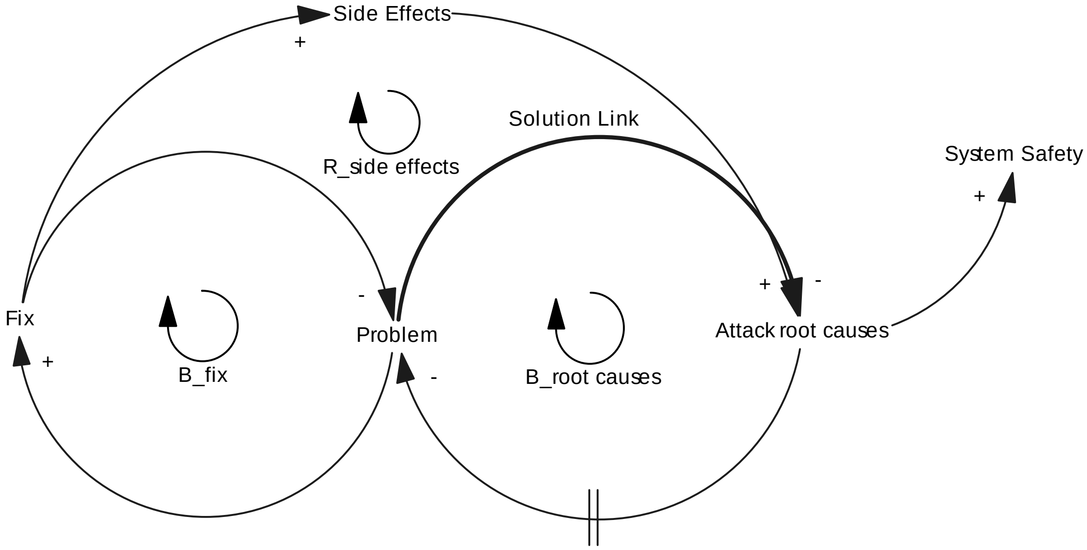


Figure 4.21 Fixing symptoms rather than root causes

Instead of applying reactive fixes when problems arise, organizations should perform root cause analysis and use the resulting insights to formulate fundamental solutions that address the underlying systemic causal factors ($B_{\text{root causes}}$).

Fixing Symptoms illustrates the tension between the appeal of short-term, symptomatic solutions, and the long-term impact of fundamental solutions. Symptomatic solutions are usually easier, faster, and cheaper to implement than long-term fundamental solutions. Initially, positive results to symptomatic solutions are seen immediately, as the visible symptoms are eliminated. Once a symptomatic solution has been successfully applied, the pressure to find and implement a fundamental solution tends to decrease. Over time, the symptomatic solutions may become less effective, or different symptoms of the underlying problem may arise—in response new symptomatic solutions are devised. The underlying problem remains. If the fundamental problem is not dealt with, symptoms can be expected to continue surfacing in various forms. Long-term, fundamental solutions, on the other hand, may be more difficult to devise, more difficult to implement, take longer to show results, and are often initially more costly. At the same time, external pressures often demand a ‘quick-fix’ to the problem.

The reactive focus of many safety programs results in placing primary emphasis on investigating previous incidents and accidents in an attempt to prevent future accidents. These efforts are not always fruitful. Excessive focus is placed on preventing recurrence of exactly the same accident, without taking sufficient account of the underlying systemic factors that allowed safety to deteriorate [Leveson, 2004a]. Attempts to identify the deeper factors or conditions that allowed the accident to occur (i.e., root cause analysis) are often insufficient.

For example, Carroll has identified instances of inadequate root cause analysis at nuclear plants [Carroll, 1998b]. In the nuclear and chemical industries, problem investigation teams are assigned to examine serious incidents and troubling trends. These investigations are part of corrective action programs to improve safety and performance. Although considerable resources are devoted to these programs, the investigations do not always result in effective learning. The authors found that the investigations tended to focus on only a few proximal causes. These causes were typically technical or involved human error, and their solutions were obvious, easily implemented, and acceptable to powerful stakeholders. Little effort was made to uncover root causes or devise fundamental solutions.

Symptomatic solutions to accidents often only decrease the likelihood of that particular accident recurring. They do not eliminate the deeper structural deficiencies that led to the accident in the first place and may lead to other accidents in the future. Once a symptomatic solution has been successfully applied, the perceived need to solve the underlying structural problem may disappear, reducing the pressure to find a fundamental solution. To improve safety in the long term the fundamental problem or structural deficiency that is causing the symptoms must be identified.

For example, if an aircraft rudder failure is shown to be the result of insufficient or poor maintenance, the recommended action may be to improve the rudder maintenance procedures. But deeper problems, such as subtle management pressure to increase maintenance throughput, may have caused the maintenance to be poorly performed in the first place.

Identifying the root causes of incidents and accidents is not always easy to do. Symptomatic solutions may be suppressing the symptoms, creating the illusion that no problem exists. These solutions may be consciously or unconsciously formulated and applied. Unconsciously applied solutions (e.g. unconsciously correcting for misaligned steering on a motor vehicle) may so successfully mask the underlying problem that operators are not aware of the problem symptoms, let alone the fundamental problem. Because any individual only has a limited view of the system, obtaining different viewpoints of the symptoms, the problem, and the system can help in identifying the fundamental problem.

Eliminating root causes is likely to be more difficult, time-consuming, and costly to implement than implementing symptomatic solutions. It is essential to obtain commitment from all parties involved with the implementation of the proposed solution. Without such commitment, the solution is unlikely to be successfully applied. Side effects of the solution must be identified as far as possible. Of course it may not be possible to foresee all the side effects. Awareness of the potential for side effects makes it easier to identify and deal with them if they do occur. Where side effects of symptomatic solutions may undermine the fundamental solution, it is necessary to stop applying these solutions before applying the fundamental solution.

Consider now two examples illustrating symptomatic responses. The first example illustrates the application of the Fixing Symptoms structure in the healthcare industry. The second example illustrates how incident reporting schemes may decrease safety by encouraging workers to hide incidents.

Not Learning from Failure. In a study of hospital nursing staff, Tucker *et al.* investigated why hospitals often do not learn from failures [Tucker et al., 2002; Tucker and Edmondson, 2003]. They found that the organizational structure encouraged ‘first-order’ problem-solving (fixing symptoms), which addresses only the particular problem, to the detriment of ‘second-order’ problem-solving, which addresses the underlying (or root) causes of the problem. As a result the hospitals repeatedly had to deal with the same types of problems, despite, and to some extent because of, their best intentions. They developed a causal loop model to explain this behaviour, as shown in Figure 4.22.

This failure to learn and its systemic causes are not unique to the health care industry. Graphical manipulation of the model reveals the underlying *Fixing Symptoms* structure, as shown in Figure 4.23.

Systemic (or fundamental) problems become visible as *problem symptoms* (e.g. accidents, incidents, and other barriers to task completion such as unavailability of equipment), which trigger *first-order (or symptomatic) problem-solving efforts*. These efforts may be quite effective at reducing problem symptoms. But first-order problem-solving decreases the chances that the underlying systemic problems will be addressed. First, effective first-order problem-solving results in immediate gratification and reduces the need to do second-order problem-solving. Second, because first-order problem-solving does not resolve the underlying causes, problems tend to recur. The result is that employees have to keep on ‘fighting fires’, which leads to employee burnout. Frustrated, stressed, and tired employees are both less willing and less able to identify and solve underlying problems and more prone to make decisions leading to accidents.

Organizations that engage only in first-order problem-solving do not address the underlying systemic issues. At best, such organizations can maintain the status quo of safety, but they are more likely to experience deteriorating safety as problems continue to arise and employees continue to be overworked. Clearly no organization wishes to contribute to its

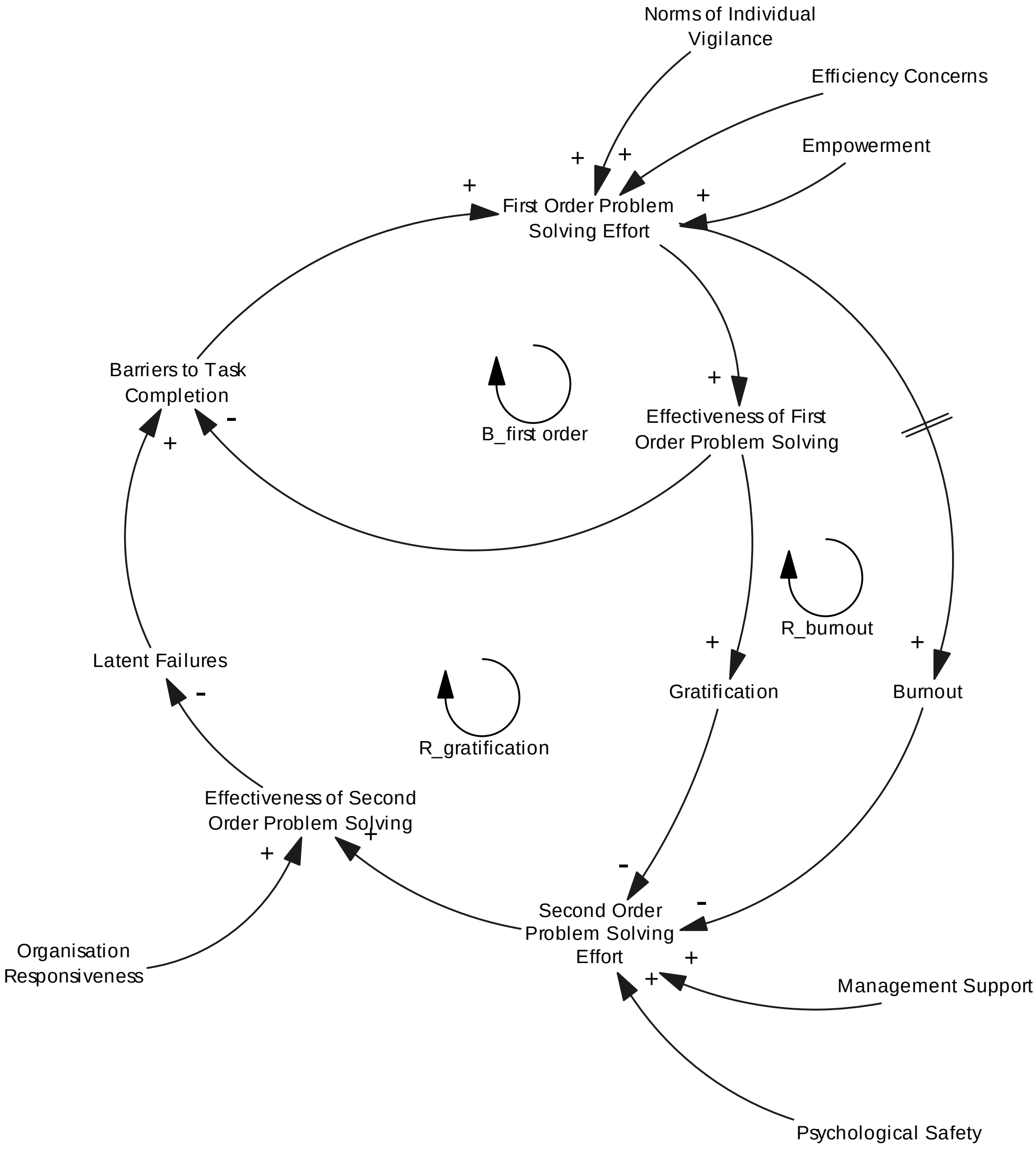

Norms of Individual Vigilance
Efficiency Concerns
Empowerment
First Order Problem Solving Effort
Barriers to Task Completion
B_first order
Effectiveness of First Order Problem Solving
R_burnout
Latent Failures
R_gratification
Gratification
Burnout
Effectiveness of Second Order Problem Solving
Second Order Problem Solving Effort
Organisation Responsiveness
Management Support
Psychological Safety
+
-

Figure 4.22 First- and Second-Order problem-solving in Hospitals[a]

[a] Adapted from [Tucker and Edmondson, 2003]

own deterioration. Why, then, would an organization engage in this behaviour? The dynamics described above explain why, once begun, this behaviour tends to persist.

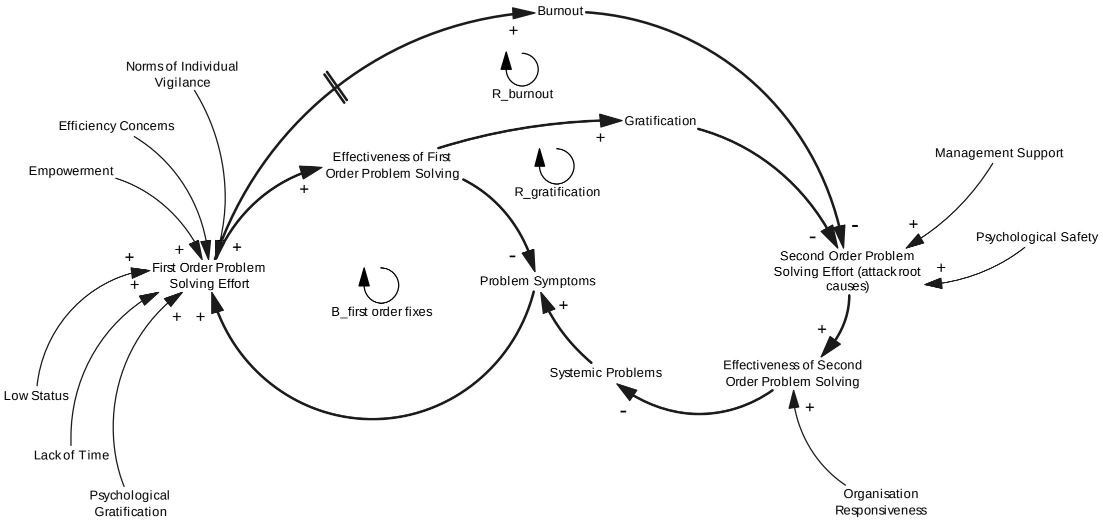

Burnout
+
R_burnout
Norms of Individual Vigilance
Efficiency Concerns
Empowerment
Gratification
+
Effectiveness of First Order Problem Solving
+
R_gratification
Management Support
Psychological Safety
+
+
+
+
First Order Problem Solving Effort
+
+
+
B_first order fixes
-
Problem Symptoms
+
-
-
+
Second Order Problem Solving Effort (attack root causes)
+
+
Effectiveness of Second Order Problem Solving
Systemic Problems
-
+
Low Status
Lack of Time
Psychological Gratification
Organisation Responsiveness

Figure 4.23 Not Learning from Failure

Now consider the factors that encourage first-order problem-solving in the first place. The hospital study identified several psychological factors and organizational features, some of which are positive, that encourage nurses to address problems on their own and as quickly as possible. Nurses' low status relative to doctors and constant time pressure encouraged them to address problems on their own and as quickly as possible. Nurses were often aware that a problem was recurring and serious and might require more senior intervention. But they generally preferred not to inform more senior personnel or suggest remedial action, both because of problems associated with the status gap between doctors and nurses and because doing so would require time that they did not have. The psychological gratification that nurses derived from dealing with problems on their own encouraged future first-order problem solving. Ironically, some positive human resource attributes encourage first-order problem-solving and so discourage learning [Tucker and Edmondson, 2003]: First, norms of individual vigilance encourage individuals to be independent and solve problems independently without considering the impact on the system. The *Safety Fixes* archetype shows how such behaviour not only does not solve the underlying problem, but can even give rise to new problems. Second, concerns about unit or subdivision efficiency mean that workers do not have time or incentives to consider factors beyond their subdivision. Third, individual empowerment is intended to help organizations address problems such as quality and productivity. But a focus on empowerment often leads to the removal of managers and other non-direct support from day-to-day activities. Workers have to deal on their own with problems that may stem from other parts of the organization. Managers, who have the status and ability to implement second-order solutions, are left unaware of problem symptoms and so cannot gain the broad perspective necessary to properly identify and solve systemic problems.

Organizations can avoid the nearly inevitable deterioration associated with first-order problem solving by working to actively encourage second-order problem-solving. Chapter 4 discusses how organizations can provide management support, create a climate of psychological safety, and become responsive to second-order problem identification and solving.

Reporting Schemes. *Reporting Schemes* is a special case of *Fixing Symptoms* that illustrates how improperly designed incident reporting schemes and other regulatory requirements can work against safety. Consider what often happens when incident reporting schemes are implemented (Figure 4.24). The primary purpose of these schemes is to encourage workers to be more careful on a day-to-day basis, thus reducing the number of incidents. As an incentive to reduce the number of incidents, workers with the best safety records (as measured by fewest reported incidents) are rewarded ($B_{\text{inappropriate incentives}}$). Rewarding workers who report the fewest number of incidents is an incentive to withhold information about small accidents and near misses. Underreporting of incidents creates the illusion that the system is becoming safer, when, in fact, it has merely been muted. Management becomes less aware of the behaviour of the system, and safety may therefore decrease (B_{safety}). At the worker level, the original goal of increasing safety is subverted into one of reporting the fewest incidents. Ironically, the introduction of an incident reporting scheme can decrease safety, as found in a study of the California construction industry [Levitt and Parker, 1976].

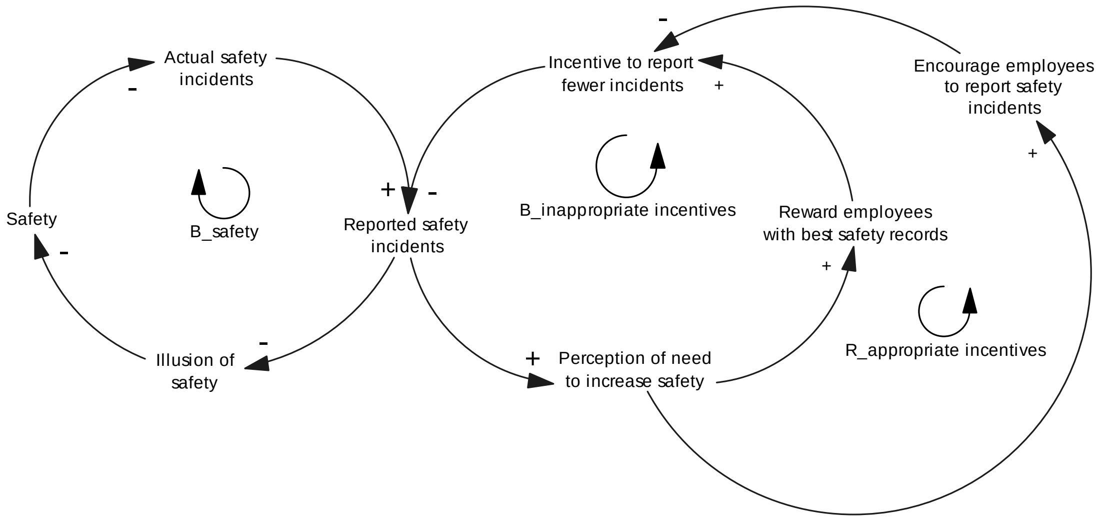

Actual safety incidents
-
Safety
B_safety
-
Illusion of safety
-
+
-
Reported safety incidents
Incentive to report fewer incidents
+
-
B_inappropriate incentives
Encourage employees to report safety incidents
+
Reward employees with best safety records
+
R_appropriate incentives
+
Perception of need to increase safety

Figure 4.24 Incident Reporting Schemes

The *Reporting Schemes* archetype illustrates how the goal of improving safety can be displaced by goals that subvert safety improvement. Regulatory requirements can have a similar negative effect on safety when they are more focussed on bureaucratic requirements than on improving safety. The investigation into the Three Mile Island accident criticised the Nuclear Regulatory Commission for displacing the objective of operational safety with a demand that the nuclear industry comply with ineffective rules:

> “The existence of a vast body of regulations by NRC tends to focus industry attention narrowly on the meeting of regulations rather than on a systematic concern for safety. Furthermore, the nature of some of the regulations, in combination with the way rate bases are established for utilities, may in some instances have served as a deterrent for utilities or their suppliers to take the initiative in proposing measures for improved safety” [Kemeny, 1979, pp. 52-53].

In implementing safety programs it is essential to consider carefully what incentives or rewards will be used to ensure compliance. One way of avoiding this type of behaviour is to encourage employees to report safety incidents, rather than rewarding employees with the best safety records ($R_{\text{appropriate incentives}}$). If symptomatic behaviour is rewarded (e.g. fewest reported incidents), it is likely that workers will find other ways to generate the same symptoms (e.g., underreporting incidents) (see also [Repenning and Sterman, 2001]). If incentives are inappropriately formulated, compliance with the intent of the program may be lower than if no incentives were offered. This behaviour can also be observed in organizations that operate according to process certification standards. In this case the purported rewards are often not visible and employees view the requirements as impeding their normal working processes. Employees therefore obey the letter of the process and documentation standards, but do not comply with the underlying intentions.

In order to maximize the likelihood that safety programs will be successful, the intent of these programs should be communicated at all levels of the organization and employees must be provided with the necessary resources to perform their part in the programs. They must be empowered to make safety-based choices in cases where such decisions might adversely affect productivity (e.g., see the discussion of worker initiated procedure changes at the North Anna nuclear plant in Section 5.3.2). If employees understand the intent of the program, and are therefore committed to it, they are more likely to comply with its intent than with the letter of the law.

Organizations can further increase the likelihood of success of improvement programs by including employees in the development of these programs. When employees are included they not only understand the intent and mechanisms of the program better, but are likely to be more committed to it.

The Vicious Cycle of Bureaucracy

This archetype demonstrates another aspect of inappropriate responses to problems. It explains the tendency of organizations to become caught up in bureaucratic rules and apathy[1] [e.g., Merton, 1968]. The vicious cycle arises as follows. Pressure on the system, due to poor performance (in this case, lack of safety), creates a desire for change. When management tries to bring this change about through increasing formalization and bureaucracy (e.g. rule making and closer supervision), employees may react dysfunctionally, becoming

[1] Bureaucracy is not necessarily bad. See []Adler and Borys, 1996] for a discussion of ‘good’, or enabling, bureaucracies, and ‘bad’, or coercive, bureaucracies.

apathetic, alienated or even subversive. The *quality of feedback about system behaviour* may therefore decrease. Despite management's best efforts, safety does not improve. Management, unaware of the real problem, institute further formalization, thus completing the vicious circle ($R_{\text{bureaucracy}}$). Eventually, management exhausts the supply of control devices ($R_{\text{control devices}}$), but by this point the organization is so caught up in bureaucracy and apathy that implementing even well-considered changes is difficult.

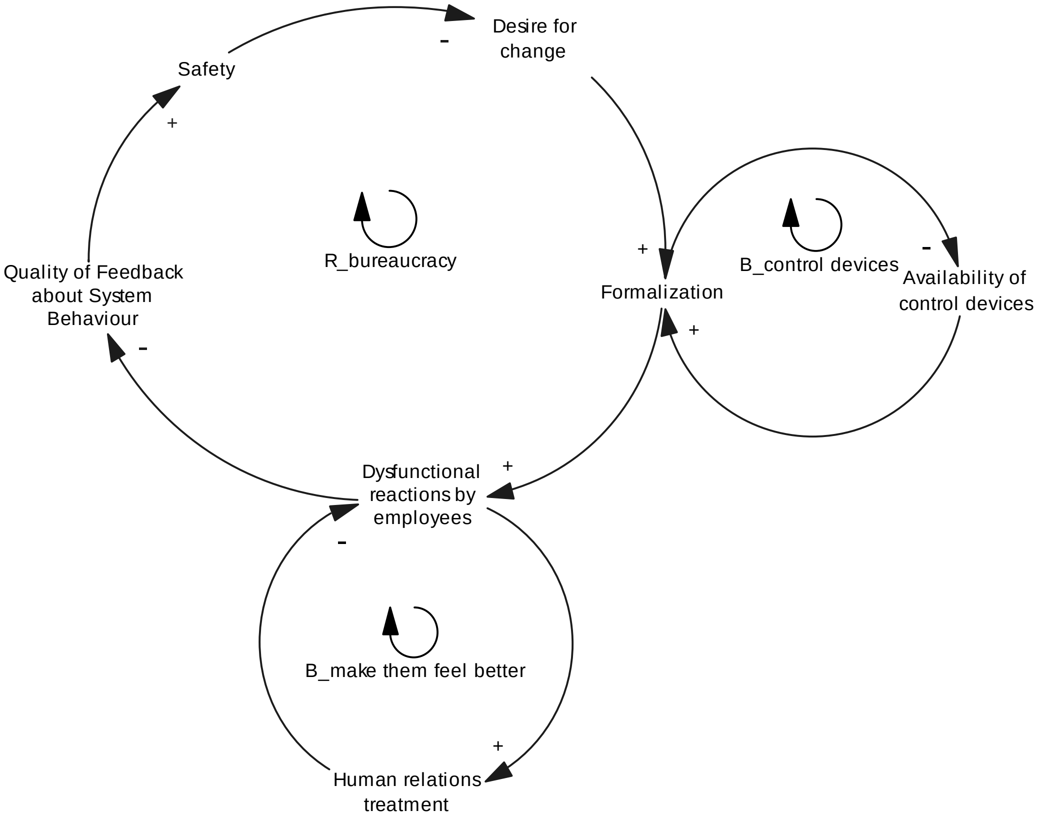

Safety
-
Desire for change
+
Quality of Feedback about System Behaviour
R_bureaucracy
+
Formalization
B_control devices
-
Availability of control devices
+
-
Dysfunctional reactions by employees
+
-
B_make them feel better
+
Human relations treatment

Figure 4.25 Vicious Cycle of Bureaucracy

One way of counteracting apathy and other dysfunctional reactions in organizations is to apply human relations treatments, such as employee recognition programs and social events [Argyris and Schön, 1978]. While such devices can lower apathy to some degree, they are not as effective as lowering the level of bureaucracy ($B_{\text{make them feel better}}$).

Bureaucracy can have a detrimental effect on safety when it decreases the ability to surface and resolve issues. Formal methods of operation and strict hierarchies limit communication, as discussed in Appendix A. When information is passed up hierarchies it is often distorted, depending on the interests of managers and the way they interpret the information. Information (e.g., concerns about safety) may even be completely silenced as it passes up the chain of command. Employees may not feel comfortable going around a superior who does not listen to their concerns. Managers may find it difficult to build trust with their subordinates. Employees or organizational units may also promote bureaucratic barriers to communication [Adler and Borys, 1996; Carroll et al., 2002] for a variety of reasons such as to decrease information overload or to assert their independence (see also Section 4.3.3).

Bureaucratic requirements may limit the time available for ‘real’ technical work, frustrating efforts to make progress or improve understanding of the system. When Morton Thiokol set up a team to investigate the problem of O-ring erosion before the Challenger disaster, the team’s progress was impeded by bureaucratic obstacles in the form of administrative delays and lack of cooperation from managers at Thiokol [Winsor, 1988]. See also Section 4.3.1 for a discussion of bureaucracy at NASA.

4.2.5 Challenges of Successfully Addressing Root Causes

The next two archetypes illustrate the problems associated with improving safety even when the appropriate solutions have been identified.

The Short-Term Performance Trap

Section 4.1 discussed how taking a short-term view leads to an apparent trade-off between performance and safety. The *short-term performance trap* shows how this short-term orientation can result in a situation where both safety and performance decline (Figure 4.26). It is based on a model developed by Repenning and Sterman to explain why organizations are often unable to reap the benefits of process improvement programs [Repenning and Sterman, 2001]. They suggested that this inability had little to do with the specific improvement tools, but was rather a systemic problem arising from the interaction between the organization‘s structure, management and employees, and the process improvement programs.

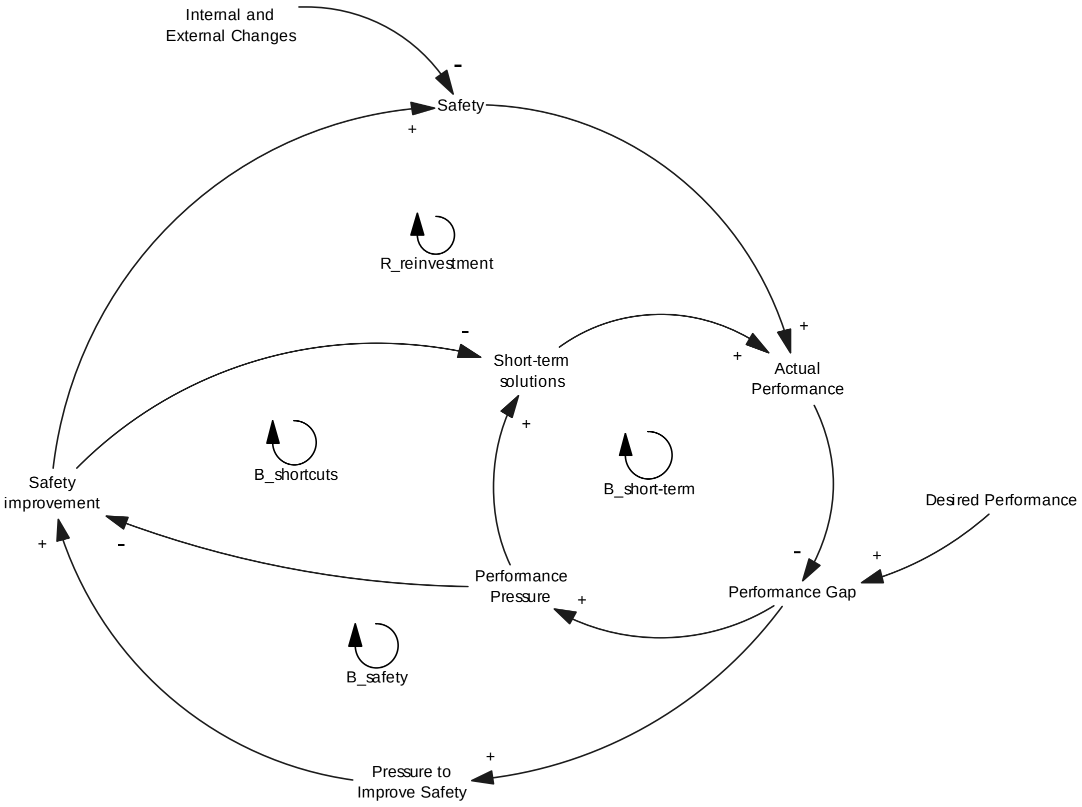


Figure 4.26 The Short-Term Performance Trap[a]

[a] Adapted from [Repenning and Sterman, 2001]

The trap works as follows: The safety of an organization and the systems it operates decreases as infrastructure wears down, designs and processes become obsolete, or skilled employees leave the organization. Safety can be increased by investing in infrastructure upgrades, process improvement, employee training, and so forth.

The *actual performance* of the organization can be increased in two ways: by focussing on short-term solutions such as cutting corners and fixing symptoms, or by focussing on long-term solutions, such as improving safety[1]. Safety improvements improve performance in the long run because they make accidents less likely. Catastrophic accidents often result in a suspension or cancellation of operations. Both Shuttle accidents resulted in immediate suspension of all launches pending investigation of the accident and remedial steps. In the case of Challenger this suspension lasted two years; the suspension following Columbia had not been lifted at the time of writing.

Consider first the option of focussing on short-term performance. *Performance pressure* arises as a result of a *performance gap* between actual and *desired performance*. The greater the gap, the greater the pressure. As a result of this pressure, reliance on *short-term solutions* increases, leading to an improvement in short-term actual performance ($B_{\text{short-term}}$).

Now consider the option of focussing on long-term performance and improving safety. In this case the performance gap creates *pressure to improve safety.* This pressure results in increased *time spent on safety improvement,* which eventually results in increased *safety* and hence improved long-term actual performance ($B_{\text{work smarter}}$).

Short-term solutions and improving (or even maintaining) safety compete for employees' time and company resources. By placing an emphasis on short-term performance and by tacitly accepting short cuts, managers send an implicit message that safety is less important, making it difficult for employees to focus on safety. The reinvestment loop ($R_{\text{reinvestment}}$) is a reinforcing loop that amplifies whichever loop ($B_{\text{short-term}}$ or B_{safety}) is dominant, further strengthening its dominance. For example, if more emphasis is placed on short-term solutions, less effort is expended on safety improvement, increasing the likelihood of accidents and incidents. A common response is to place even more emphasis on short-term solutions in order to narrow the performance gap quickly. Conversely, if more emphasis is placed on safety improvement, accidents and incidents are less likely. Fewer

[1] Of course safety improvement is not the only way to that long-term performance can, or should, be improved. Other techniques, such as process improvement, should also be used where appropriate.

accidents and incidents means fewer interruptions to normal work and therefore frees up more time to work on safety improvement. For example, by making an effort to identify and solve the root causes of a problem, the need for symptomatic fixes is eliminated.

The reinvestment loop typically works to reinforce the short-term loop and not the safety improvement loop because focussing on short-term solutions and decreasing emphasis on safety usually yield increased performance in the short term. Spending less time on safety means more time is available to work. Short cuts like avoiding maintenance and cutting training frees up time to do work, thus increasing performance. Symptomatic fixes mask underlying problems and allow continued operation of the system despite underlying problems. Short cuts and fixes are tempting because they show immediate improvements in productivity but their negative effect on safety usually only becomes apparent after some time. Because short-term solutions yield immediate benefits, managers may be blinded to alternative considerations. The shortcuts loop ($B_{shortcuts}$) is a balancing loop that allows increased performance at the cost of cutting corners.

It is difficult for organizations to avoid falling into the short-term performance trap for two main reasons. First, short-term solutions usually show immediate results and activate the reinvestment loop in the negative sense, making investments in improving safety less likely. Second, it is far easier to take short cuts and fix symptoms than it is to understand the root causes of problems.

Employee Commitment

Unintended Side Effects and *Fixing Symptoms* have shown that identifying root causes and devising and implementing appropriate solutions is difficult. One of the problems with implementing safety improvement programs is obtaining and maintaining employee commitment. The *Employee Commitment* archetype[1] illustrates the potential pitfalls associated with obtaining and maintaining employee commitment (Figure 4.27).

[1] This archetype is based on concepts in [Keating et al., 1999].

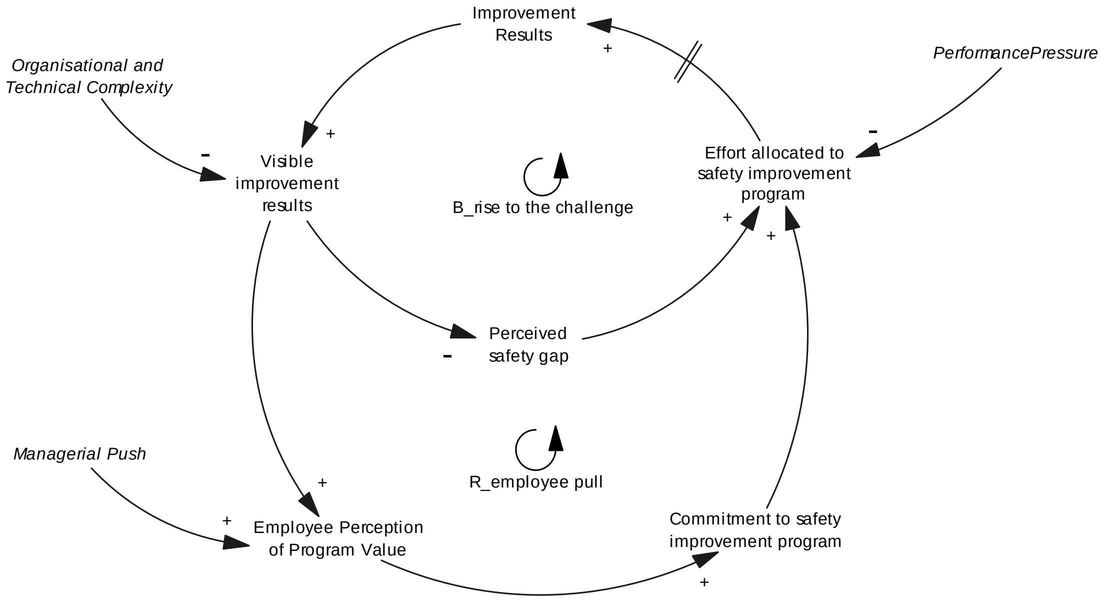

Improvement Results
Organisational and Technical Complexity
PerformancePressure
Visible improvement results
B_rise to the challenge
Effort allocated to safety improvement program
Perceived safety gap
Managerial Push
R_employee pull
Employee Perception of Program Value
Commitment to safety improvement program
+
-

Figure 4.27 Employee Commitment

There are two sources of employee commitment to improvement programs, managerial *push* and employee *pull* [Shiba et al., 1993]. Managerial push refers to efforts to promote improvement efforts or force employee participation. Examples of managerial push include inspirational speeches and literature, mandatory participation in training and workshops, financial incentives, and performance review criteria based on observed improvement. Thus, increased *Managerial Push* drives increased *Employee Perception of Program Value*.

Employee pull occurs when employees understand the benefits of improvement and commit themselves to improvement efforts independent of, and occasionally in spite of, managerial attitudes [Keating et al., 1999]. Employee perception of program value increases in response to *visible improvement results*. This perception sets off increased *commitment to improvement program*, resulting in increased *effort allocated to improvement program*. If the program is appropriately designed, the result is an eventual increase in *improvement results* and in *visible improvement results*, leading once again to increased *employee perception of program value (*$\mathrm{B}_{\text{rise to the challenge}}$). Employee pull can therefore create a valuable reinforcing loop of increasing commitment and increasing improvement results ($R_{\text{employee pull}}$). Note that removing managerial push leaves the employee pull loop intact, while removing employee commitment destroys the loop. Managerial push can adjust employee perception of program value, but a program cannot succeed without employee commitment.

Safety improvement programs do not always have highly or immediately visible results (see *Eroding Safety*). In particular, the *organizational and technical complexity* of a system make it more difficult to devise, implement, and observe the results of improvement programs. In cases where results have low visibility, are ambiguous, or take a long time to become apparent, employees may be less committed than the program‘s ‘goodness’ warrants. Furthermore, many improvement programs show a ‘worse-before-better’ pattern, whereby performance deteriorates before it improves. For example, safety measures such as increased inspection or stricter design verification procedures directly impinge on short-term productivity (see *Short-term Performance Trap*). In such cases, extra effort should be put into educating employees about the benefits of the program, including emphasising the long-term nature of the expected pay-offs.

4.2.6 Tailoring and Expanding the Archetypes

The safety archetypes represent the basic characteristics of some of the most common risk dynamics. Organizations can obtain the most value from the archetypes by tailoring them to their specific problems, as shown in the example applications (e.g. *Disappointing Safety Programs*). Organizations can also construct additional archetypes to understand their specific problems.

The archetypes have emphasised the importance of identifying the fundamental solution and obtaining employee commitment. One way to facilitate both finding and implementing the solution is to involve employees in the generation of the dynamic models as much as possible. Employees may have specific knowledge of the problem and insights into solutions. In addition, by involving employees in the process, they are more likely to be committed to subsequent solutions.

The next section presents some examples of archetypal behaviour at NASA, illustrating the process of tailoring and generating additional archetypes.

4.3 Organizational Risk Dynamics at NASA

This section illustrates the construction of archetypes to improve understanding of safety problems at NASA, specifically in the Space Shuttle Program. The archetypes are examples of problem behaviours at NASA and focus on situations where feedback makes the situation worse. They are not intended to be a complete explanation of problems at NASA. The archetypes are based on a history of NASA [McCurdy, 1993] and on the Challenger [Rogers, 1986] and Columbia accident reports [Gehman, 2003].

On a complex program such as the Space Shuttle problems are inevitable. For example, it is impossible to foresee all the possible interactions between shuttle components, and which of those will cause problems. The trick is to surface and resolve issues before they turn into catastrophes. The following examples identify some mechanisms that decrease NASA's *ability to surface and resolve issues*.

The first example shows how *Vicious Circles of Bureaucracy* can be tailored to understand the problems of bureaucracy at NASA. The next three examples illustrate specific problem behaviours on the shuttle program.

4.3.1 Growing Bureaucracy

Figure 4.28, based on *Vicious Circles of Bureaucracy*, illustrates how NASA's bureaucracy grew and changed in a way that decreased the ability to surface and resolve issues [McCurdy, 1993, pp. 111-118].

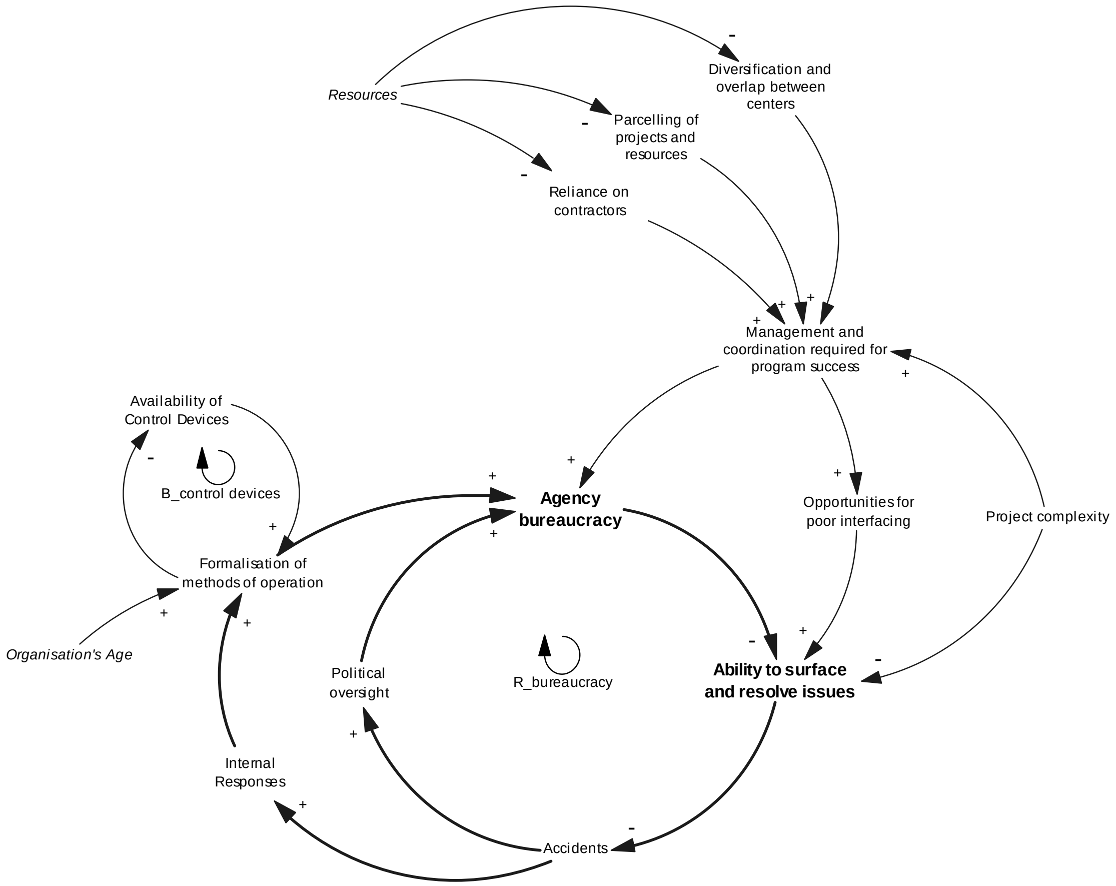

Resources
Diversification and overlap between centers
Parcelling of projects and resources
Reliance on contractors
Management and coordination required for program success
Availability of Control Devices
B_control devices
Agency bureaucracy
Opportunities for poor interfacing
Project complexity
Formalisation of methods of operation
Organisation's Age
Political oversight
R_bureaucracy
Ability to surface and resolve issues
Internal Responses
Accidents
+
-

Figure 4.28 Growing Bureaucracy

The agency bureaucracy, as indicated by the relative number of administrative staff, expanded significantly over the years since NASA was founded. In 1961, approximately six per cent of NASA staff was classified as professional administrators. By 1991, this percentage had grown to approximately eighteen per cent [McCurdy, 1993, p. 116, Fig. 7]. This increase in administrative staff was accompanied by a growing perception within NASA that bureaucracy made it more difficult to accomplish technical tasks [McCurdy, 1993, p. 117].

There are three main sources of bureaucratic growth [McCurdy, 1993, pp. 111-118]. First, organizations tend to become more bureaucratic over time because initially informal or ad hoc *methods of operation are formalised*. Such formalization can be useful because it decreases uncertainty about how to act and results in increased predictability. Second, as *programs become more complex*, informal management and planning is no longer sufficient. In NASA's case, decreasing resources further increased the *management and coordination required for program success*. Increasing reliance on contractors, parcelling of projects and resources between centers, and diversification and overlap between centers all worked to increase the management burden. *Opportunities for poor interfacing* increased, decreasing the *ability to surface and resolve issues*. Third, as political oversight increases, organizations are forced to deploy more administrative staff to handle the burden. In NASA's case, political oversight grew as government grew bigger and committee structures became more democratic [McCurdy, 1993] and in response to the Apollo and Shuttle accidents.

Increased public oversight and internal changes in response to accidents increases the bureaucratic burden and may therefore perversely make accidents more likely ($B_{bureaucracy}$).

It is interesting to note that there are no negative arrows going into bureaucracy. As many organizations, and especially governments, have found, growing bureaucracies is far easier than shrinking them.

4.3.2 Contracting the Culture Away

Ever increasing levels of contracting out had a profound negative effect on the technical culture (see Appendix B) at NASA [McCurdy, 1993], as illustrated by Figure 4.29. Increasing *reliance on contractors* meant that the *in-house technical capability* declined, requiring further reliance on contractors ($R_{losing\ the\ culture}$). The result is a continuing decrease in in-house technical capability, which weakens the *technical culture*, and decreases the *ability to surface and resolve issues*.

Any agency that attempts challenges like the moon landings and the Space Shuttle has to rely on contractors to some extent [McCurdy, 1993]. It is possible to use contractors with-

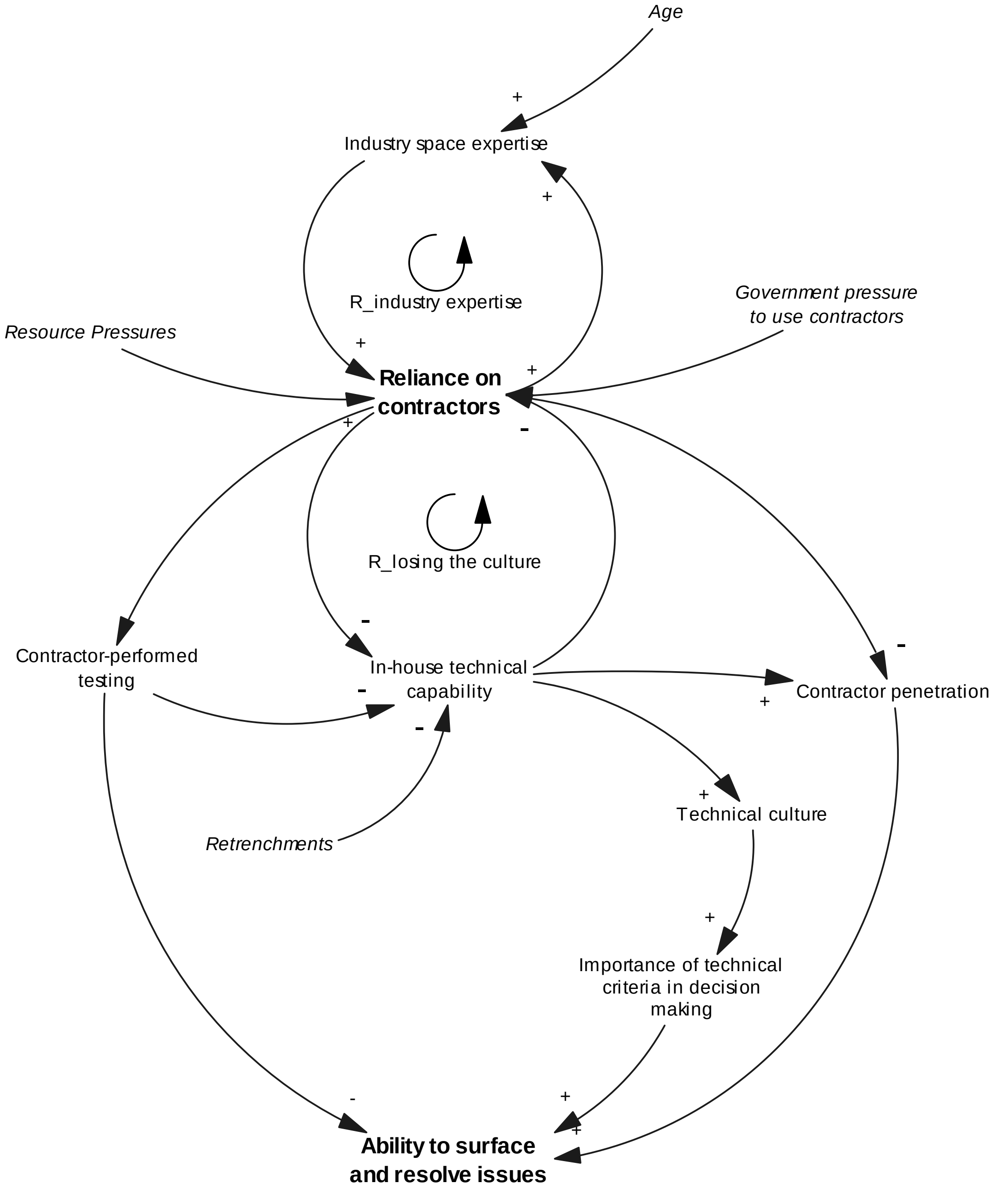


Figure 4.29 Contracting the Culture Away

out compromising safety, but the nature and extent of the contracting relationships must be carefully managed. NASA has used contractors since the days of Apollo but the relationship between NASA and its contractors changed significantly over time. In the 1960s, NASA took great pride in its in-house technical capability, from technicians to scientists and engineers. They developed a tradition of *contractor penetration*, double-checking all contractor work, and retaining control of functions like spacecraft control and mission planning.

Following the moon landings NASA was forced to make severe *retrenchments*, which focussed disproportionately on clerical and technical staff [McCurdy, 1993]. While NASA retained its scientists and engineers, the retrenchments curtailed the agency's ability to perform in-house technical work. NASA came to rely more and more on contractors and NASA engineers spent more of their time managing contractors and less time doing actual engineering, such as testing.

When NASA first embarked on space exploration, *industry space expertise* was limited [McCurdy, 1993]. NASA was therefore able to both exert authority over contractors, and resist government pressure to use contractors. Over time, space industry expertise increased, partially as a result of doing contract work for NASA and the DoD ($R_{\text{industry experience}}$). By the 1980s the space industry was big enough to exert strong political influence and preclude a return to in-house capability.

In 1984 the work of preparing and flying the space shuttle was assigned to a consortium of aerospace companies organized under the United Space Alliance. The agency that had insisted on knowing everything about its spacecraft during the Apollo days relinquished control of the most complex spacecraft ever built to contractors. This step placed NASA engineers and technicians one step further from the shuttle and further decreased their ability to surface and resolve issues.

While contractors are of course concerned for the safety of the crew and shuttle, they are ultimately driven by financial concerns. Many contractors rely on NASA for a large portion of their business and must therefore maintain good relations with NASA. Causing frequent launch delays creates friction and may place future contracts in jeopardy. NASA emphasised the importance of avoiding delays by providing contractors with on-time launch incentives. Contractors therefore have less incentive to look for problems or be aggressive about delaying launches when they are not absolutely sure that there is a problem.

The fatal launch decision on the eve of the Challenger disaster can be partly attributed to Morton Thiokol‘s reluctance to cause further launch delays. There had already been several launch delays when Marshall and Morton Thiokol teams met to decide whether to launch. Thiokol engineers expressed concerns about launching in cold weather but when Thiokol suggested delaying the launch, the Marshall manager famously asked, “My god, Thiokol, when do you want me to launch, next April?” [Rogers, 1986]. Subsequently Thiokol management reversed its position and recommended the launch at the urging of NASA and contrary to the views of its engineers in order to accommodate a major customer [Rogers, 1986]. Pressure to launch coupled with contractors’ concerns about future contracts can have an adverse effect on safety.

4.3.3 Ineffective Leadership Structure

Lead centers were intended to be a way of decreasing the management load at NASA headquarters while simultaneously retaining centralised control of programs. McCurdy suggests that the prevailing conditions of increasing program complexity (shuttle and space station development) and decreasing resources made this approach inappropriate. It weakened program management and exacerbated rivalry and mistrust between centers [McCurdy, 1993, p. 127], as shown in Figure 4.30.

As industry space expertise increased, there were more questions about whether all the NASA centers were necessary. Centers had to look after their own survival. Some centers diversified, moving into different parts of the space program or developing clients outside NASA. The Marshall Center took charge of the development of Skylab while the Lewis Center moved into non-space related research, developing expertise in diverse areas like air quality monitoring and power generation. In some cases centers developed overlapping areas of expertise, resulting in direct competition. For example, Marshall and Goddard

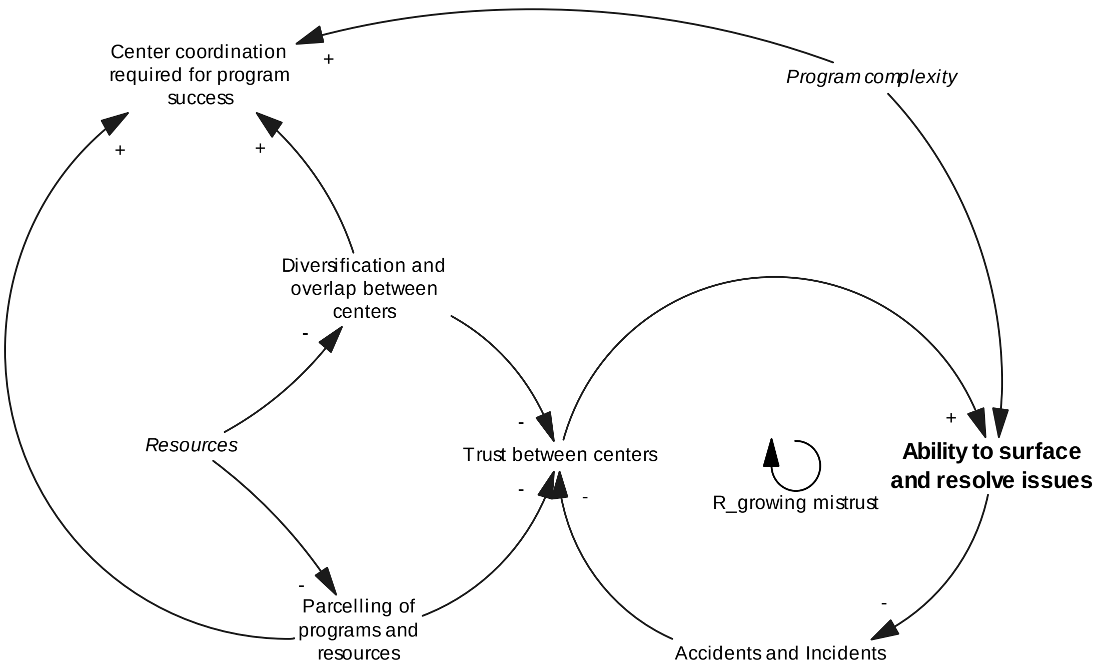

Center coordination required for program success
+
Program complexity
+
+
Diversification and overlap between centers
-
Resources
-
Trust between centers
-
-
R_growing mistrust
+
Ability to surface and resolve issues
-
Parcelling of programs and resources
-
Accidents and Incidents

Figure 4.30 Ineffective Leadership Structure

competed to develop the Hubble Space Telescope. Competition can be good but it must be carefully managed so that it encourages rather than suppresses progress. For example, when more than one center works on the same project, lines of responsibility must be clearly drawn. In the case of Hubble, in May 1972 both centers were granted an equal share of responsibility, which created permanent management problems. Work on the space station Freedom was parcelled among four centers, with Johnson taking overall responsibility for managing itself and the three other centers. Lines of responsibility were unclear leading to bitter conflict between centers and the collapse of the scheme after only two years [McCurdy, 1993]. James C. Welch, who became NASA's Hubble program manager in 1983 noted that, "No one really felt like they had the authority to go in and direct all other parts of the program to conform" [Capers and Lipton, 1993]. Parcelling of resources and responsibilities between the centers meant they were effectively competing against each other and therefore had little reason to be supportive or cooperative.

Greater *program complexity* and *parcelling of programs and resources* means that greater coordination is required between centers for programs to succeed. But parcelling and overlap places centers in competition with each other (to maximize its likelihood of survival each center needs to obtain as much of each project as it can handle), which increases mistrust between centers and so decreases communication and coordination.

As trust between centers declines, technical criteria become less important than bureaucratic and political criteria, centers hold problems back, and the lead center is unable to exert its authority. The result is a decreased *ability to surface and resolve issues*, increasing the likelihood of accidents and incidents. The finger pointing that invariably results when something goes wrong decreases trust further, resulting in a vicious circle of growing mistrust and decreasing transparency and cooperation ($R_{\text{growing mistrust}}$). Expecting centers to follow the lead of a lead center that they do not trust is naive and places unfair expectations on the lead center.

4.3.4 Going Operational

The decision to declare the shuttle “fully operational” after only four flights helped to set the stage for the Challenger accident. Figure 4.31 illustrates why this decision was made so early and what the effects on safety were.

As organizations age, they become less comfortable with uncertainty and start preferring routine operations to risky ventures [McCurdy, 1993]. In the early 1980s NASA was facing a future of highly uncertain funding. By shifting from an emphasis on research and development to an emphasis on space operations, NASA hoped to decrease this uncertainty. The space station was one potential source of guaranteed, long-term, funding but its construction (and approval) depended on a fully operational and reliable shuttle. At the same time, launch competition had suddenly arrived in the form of the European Space Agency’s Ariane rocket. NASA was in danger of losing its status as the primary means of accessing space. By declaring the shuttle operational, NASA cleared the way for space station approval and increased the predictability of its funding. Decreased appetite for

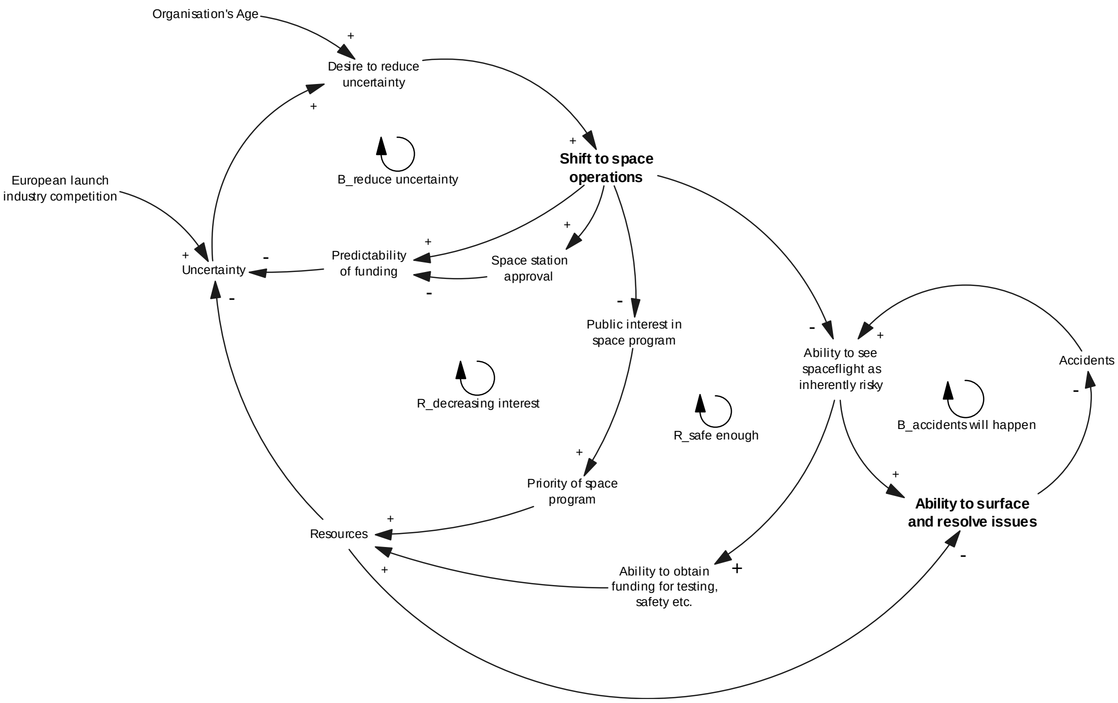

Organisation's Age
+
Desire to reduce uncertainty
+
B_reduce uncertainty
+
Shift to space operations
European launch industry competition
+
Uncertainty
-
-
Predictability of funding
+
-
+
Space station approval
-
Public interest in space program
-
+
Ability to see spaceflight as inherently risky
Accidents
-
R_decreasing interest
R_safe enough
B_accidents will happen
+
Priority of space program
+
Ability to surface and resolve issues
+
Resources
+
Ability to obtain funding for testing, safety etc.
+
-

Figure 4.31 Going operational

uncertainty as the agency aged, coupled with the pressure of decreasing resources and increasing competition, pushed NASA away from research and toward space operations ($B_{\text{reduce uncertainty}}$).

The shift towards space operations instigated a number of feedback loops that worked to decrease safety. First, the emphasis on the continuous and routine nature of shuttle missions decreased public interest in the space program. Shuttle missions were not as exciting as lunar landings. As public interest decreased, the political priority of the space program decreased, leading to reduced funding and greater uncertainty. The strategy that was supposed to secure funding ironically decreased politicians' willingness to fund the agency ($B_{\text{decreasing interest}}$). Decreasing resources decreased the ability to surface and resolve issues and therefore made accidents more likely.

Second, by shifting to space operations, NASA decreased the ability to see spaceflight as inherently risky. The perception that the shuttle was safe and operational made it more difficult to obtain funding for safety and testing ($R_{\text{safe enough}}$).

Third, the shift to operations made it difficult for many NASA and contractor employees to continue seeing the shuttle as essentially an experimental vehicle. While it appears that many lower level employees were well aware of the risks, the investigations of both the Challenger and Columbia accidents suggest that management did not share this view ($B_{\text{accidents will happen}}$).

4.3.5 Short-Term Cost Savings

The space shuttle is an example of how design and management choices made to save money in the short term incurred severe cost and safety penalties in the long term, as shown in Figure 4.32.

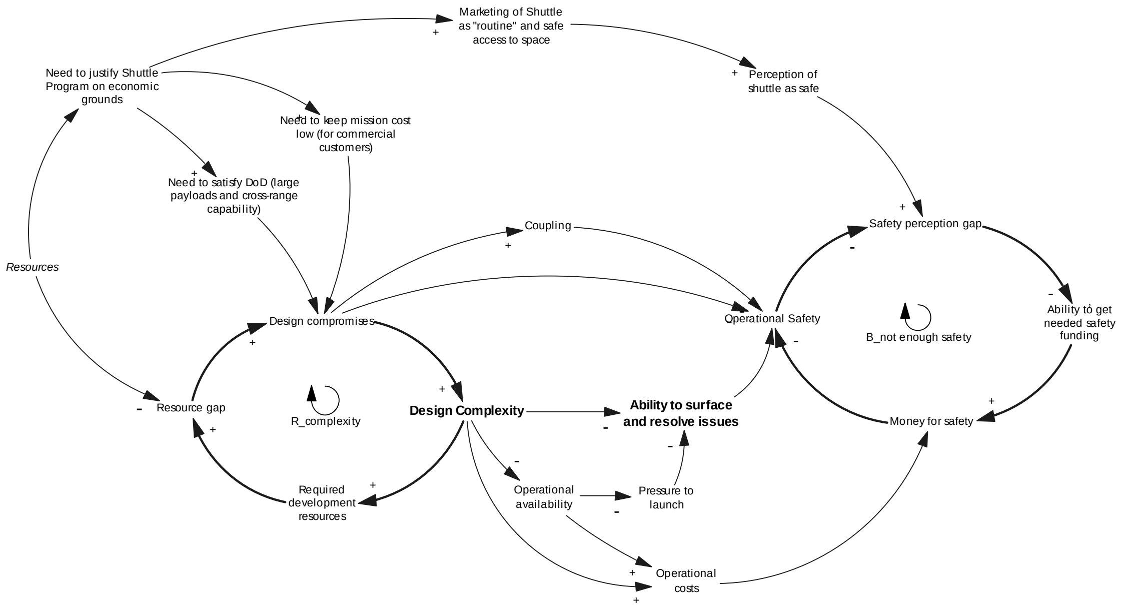

Marketing of Shuttle as "routine" and safe access to space
Need to justify Shuttle Program on economic grounds
Need to keep mission cost low (for commercial customers)
Perception of shuttle as safe
Need to satisfy DoD (large payloads and cross-range capability)
Coupling
Safety perception gap
Resources
Design compromises
Operational Safety
B_not enough safety
Ability to get needed safety funding
Resource gap
R_complexity
Design Complexity
Ability to surface and resolve issues
Money for safety
Required development resources
Operational availability
Pressure to launch
Operational costs

Figure 4.32 Short-Term Cost Savings

Compromises on the shuttle design began almost as soon as the shuttle concept was put forward. Initially, NASA proposed the shuttle as a fully reusable vehicle that would provide “routine and low cost manned access to space” [Gehman, 2003, p. 22]. But there was little support from government, which had questions about the value of manned space flight, especially given the high cost of shuttle development. NASA was therefore forced to justify the shuttle on economic grounds. NASA argued that if the shuttle launched all commercial and government payloads, and that if it were fully reusable, the total cost of launching and maintaining satellites would be dramatically reduced. The proposed savings would only be realised if the Shuttle made approximately fifty flights a year, including launching all Department of Defense (DoD) satellites.

Attempting to satisfy the DoD and commercial customers simultaneously created complex requirements. Commercial customers required low cost launches, while the DoD required a large payload bay and the ability to perform large “cross-range” manoeuvres. Satisfying these complex requirements, while keeping costs low, required “a revolution in space technology” [Gehman, 2003, p. 22]. But revolutions do not come cheaply, and numerous compromises were made to keep development costs low and decrease the *resource gap* between required and available resources. Compromises tend to increase *design complexity* (especially when they come later in development process), thus increasing the *required development resources*. The result is that the compromises intended to decrease development or operational costs may perversely contribute to the *resource gap* ($R_{complexity}$).

Some design compromises directly detract from safety, while the effect of other types of compromises is more subtle. Compromises that affect safety can be grouped into three types:

Compromises that directly decrease safety. In some cases compromises were made that directly decreased safety. NASA had to choose between solid and liquid rocket engines and boosters for the shuttle. Liquid rocket engines are safer because unlike solid rocket engines they can be shut down after ignition. Solid rocket engines, on the other hand, are cheaper to develop. Administrator James F. Fletcher openly acknowledged that the choice of solid rocket engines was based on a “trade-off between future benefits and earlier savings in the immediate years ahead: liquid boosters have lower development costs” [Logsdon, 1986]. NASA therefore decided to design the shuttle with liquid rocket engines and solid rocket boosters.

Conflicting requirements mean that it is not always possible to select the design that maximises safety. The silver lining in this type of compromise is that the effect on safety is obvious and can be more easily assessed than the effect of less direct compromises. When this type of compromise cannot be avoided, its impact on risk should be considered and communicated to the relevant decision makers. In the above example, solid fuel was used only on the boosters, which are jettisoned shortly after launch. By avoiding using solid fuel on the shuttle engines, the risk of an accident involving solid rocket malfunctions was reduced. The design of the boosters themselves, however, left much to be desired, as was tragically revealed by the Challenger accident.

A decision was made to section the solid rocket boosters for easier transportation from the solid rocket contractor‘s facilities in Utah. This decision was not one that necessitated a decrease in safety, if the joints between sections were properly designed. Unfortunately the design of the joints, coupled with incomplete characterization of O-ring performance[1], was such that a joint failure was almost inevitable, given low enough temperatures. Subsequent redesign of the joints showed that a safer design was indeed possible.

Compromises that increase coupling. By making designs more tightly coupled (see Chapter 4), designers can extract more performance for similar cost. Separating components reduces the likelihood of accidents, but a performance penalty is extracted. As one engineer noted:

> “You‘re letting some opportunities for increasing your performance get away when you [separate components]... The trend is to greater complexity in order to get more bang for

[1] For example, there were no launch criteria that specified safe launch temperatures for O-rings.

> the buck... In a case like the shuttle, if you don‘t take advantage of some of the opportunities to improve performance, you‘ll never get off the ground.” [McCurdy, 1993, p. 152]

For example, a less coupled and safer design would consist of an orbiter and a separate launch vehicle. But recovering and refurbishing the engines would be difficult. A decision was therefore made to incorporate the engines into the shuttle and make the rocket boosters reusable.

Compromises that increase complexity. Some compromises do not directly involve safety, but have a cumulative negative effect on safety because they increase the complexity of the system. As systems become more complex, they become more expensive to operate, more difficult to understand, and the potential for unforeseen dysfunctional interactions increases. Increasing operational cost can have an insidious effect on safety. Despite all protestations to the contrary, spending on safety is often seen as a luxury, especially when funding is restricted and there is a history of successful missions. For example, the Kraft Report infamously characterised the shuttle as “mature” and dismissed concerns from credible sources as being part of an unnecessary “safety shield conspiracy” [Kraft, 1995]. Another side-effect of increasing complexity is that preparing shuttles for launch becomes more difficult and it is more likely that problems will arise that delay launches. Frequent launch delays increase the pressure to launch on any given occasion, thereby decreasing the ability to surface and resolve issues.

NASA made one other big compromise, that rivals the design compromises for its impact on safety. When they marketed the shuttle as “safe” and “routine”, they created a perception that the shuttle was safe. This perception was at odds with reality, creating a *safety perception gap* that compromised NASA‘s *ability to obtain safety funding* ($B_{\text{not enough safety}}$). This perception has been so enduring that NASA has never obtained the necessary funds to ensure shuttle safety. The CAIB notes:

> “In the end, the greatest compromise NASA made was not so much with any particular element of the technical design, but rather with the premise of the vehicle itself. NASA promised it could develop a Shuttle that would be launched almost on demand and would fly many missions each year. Throughout the history of the program, *a gap has persisted between the rhetoric NASA has used to market the Space Shuttle and operational reality,*

> leading to an enduring image of the Shuttle as capable of safely and routinely carrying out missions with little risk.” [Gehman, 2003, p. 23]

***

The preceding sections have presented examples of behavioural dynamics at NASA that contributed to deteriorating safety. Similar dynamics can be developed for other industries.

4.4 Summary

Understanding risk requires an appreciation of how and why risk changes over time. This chapter has discussed the dynamics of risk in organizations, with particular emphasis on the impact of organizational factors.

The apparent conflict between performance and safety was shown to result from the different time horizons applying to performance and safety. Performance is measured in the short term, while safety is indirectly observed over the long term. A short-term view creates the impression that safety and performance necessarily conflict. Expanding the time horizon attenuates the tension. By increasing awareness of the often implicit trade-offs between safety and performance, organizations can avoid decisions that increase risk.

Accidents in diverse industries, while unique in their technical aspects, often exhibit common patterns of organizational behaviour. This chapter identified several such patterns, or archetypes, and demonstrated their application in diverse industries. In addition, NASA specific archetypes were developed based on investigations into the Challenger and Columbia accidents.

Chapter 5

A NEW APPROACH TO RISK ANALYSIS

I am a devoted admirer of theory, hypothesis, formula, and every other emanation of pure intellect which keeps erring man straight among the stumbling blocks and quagmires of matter-of-fact observations.

G. B. Airy, Director of the Royal Observatory, 1949

"The main benefit of estimating risk lies in the achievement of a detailed understanding of the engineered system."

Royal Society Report, 1992

This chapter introduces an approach to risk analysis that is applicable to modern, complex socio-technical systems. The proposed approach goes beyond event-based models to include risks that do not solely arise from component or subsystem failures and incorporates the impact of technical, human and organizational factors on risk. By taking an explicit lifecycle view of systems, the approach (1) enables the early identification of risks and risk mitigation strategies; (2) aids in the allocation of resources to best manage risk; and (3) provides for the continuous monitoring of risk throughout the system lifecycle. In addition, the approach emphasizes and enables the participation of members at all levels of the organization as well as other stakeholders in order to best identify, assess, and manage risks.

5.1 Continuous Participative Risk Management

Risk management consists, on the one hand, of proving that an initial system design satisfies safety requirements, and, on the other hand, of best allocating limited resources to maintain risk at an acceptable level[1] throughout the system's lifecycle[2]. Managing risk effectively over the entire lifecycle requires first that risks be properly identified and assessed, and second that impacts of decisions in the present on the future behaviour of the system be considered and understood. This section discusses one way of effectively addressing risk management throughout a system's lifecycle: Continuous Participative Risk Management (CPRM). It is 'continuous' because the process of risk identification, assessment, and mitigation should continue throughout the system lifecycle and not be a one-time effort; and it is 'participative' because inputs from members at all levels of the organization are needed for an appropriate and extensive risk management effort. Figure 5.1 summarizes the motivation for CPRM.

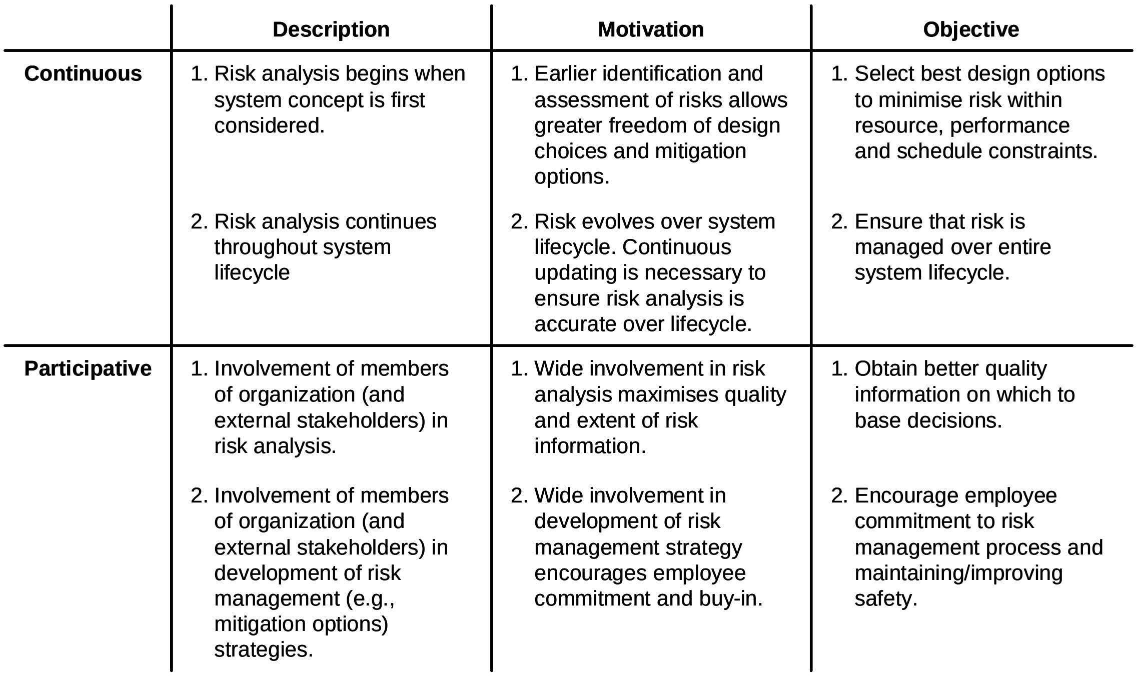
5.1.1 The Importance of Continuous Risk Management

There are two aspects to continuous risk management. First, initiating risk management when a new system is first considered allows the early identification and assessment of risks. The earlier risks are understood the easier it is to develop effective and cost-effective strategies to eliminate or mitigate the risks. When risks are identified late in system development, significant changes in the system design may no longer be feasible. 'Band-aid' and after-the-fact fixes for risks tend to be less effective and more expensive. Risk assessment tools like probabilistic risk assessment that require a near-complete system design are not a good basis for continuous risk management of novel systems.

Second, continuing risk management throughout a system's lifecycle until its retirement allows operators to actively manage risk and maintain it at an acceptable level throughout the system's lifecycle. Risk is not constant: new risks may arise, and old risks may change.

[1] The question of acceptability depends on various factors such as public opinion and regulatory standards. The acceptability of risk is context dependent: Risks that are deemed acceptable in one context may be unacceptable in another context. For example, society accepts much higher risks for military aviation than for civil aviation. The interested reader is referred to the extensive literature, for example [Fischhoff et al., 1983].

[2] Here system lifecycle is considered in the widest sense and extends as far as the consequences of realized system risks may persist.



	Description	Motivation	Objective
Continuous	1. Risk analysis begins when system concept is first considered.	1. Earlier identification and assessment of risks allows greater freedom of design choices and mitigation options.	1. Select best design options to minimise risk within resource, performance and schedule constraints.
	2. Risk analysis continues throughout system lifecycle	2. Risk evolves over system lifecycle. Continuous updating is necessary to ensure risk analysis is accurate over lifecycle.	2. Ensure that risk is managed over entire system lifecycle.
Participative	1. Involvement of members of organization (and external stakeholders) in risk analysis.	1. Wide involvement in risk analysis maximises quality and extent of risk information.	1. Obtain better quality information on which to base decisions.
	2. Involvement of members of organization (and external stakeholders) in development of risk management (e.g., mitigation options) strategies.	2. Wide involvement in development of risk management strategy encourages employee commitment and buy-in.	2. Encourage employee commitment to risk management process and maintaining/improving safety.

Figure 5.1 Continuous Participative Risk Management

In addition, because people tend to overemphasize near-term risks over long-term risks, dysfunctionality often creeps into decision making (for example, saving money or time in development while dramatically increasing maintenance costs). Continuous risk management allows operators to not merely to minimize risk at a particular stage, but rather to minimize the total risk across the entire system lifecycle. Hence the case for a *continuous* participative risk management approach.

5.1.2 The Importance of Participative Risk Management

Consider now the importance of wide participation in risk management in general, and risk analysis in particular. Figure 5.2 shows the different groups involved in system design and risk analysis. Each person has a unique background, interacts with different aspects of the system to different extents, and therefore develops a unique understanding, or mental model, of the system. For example, electrical engineers and structural engineers will have different perspectives on a system.

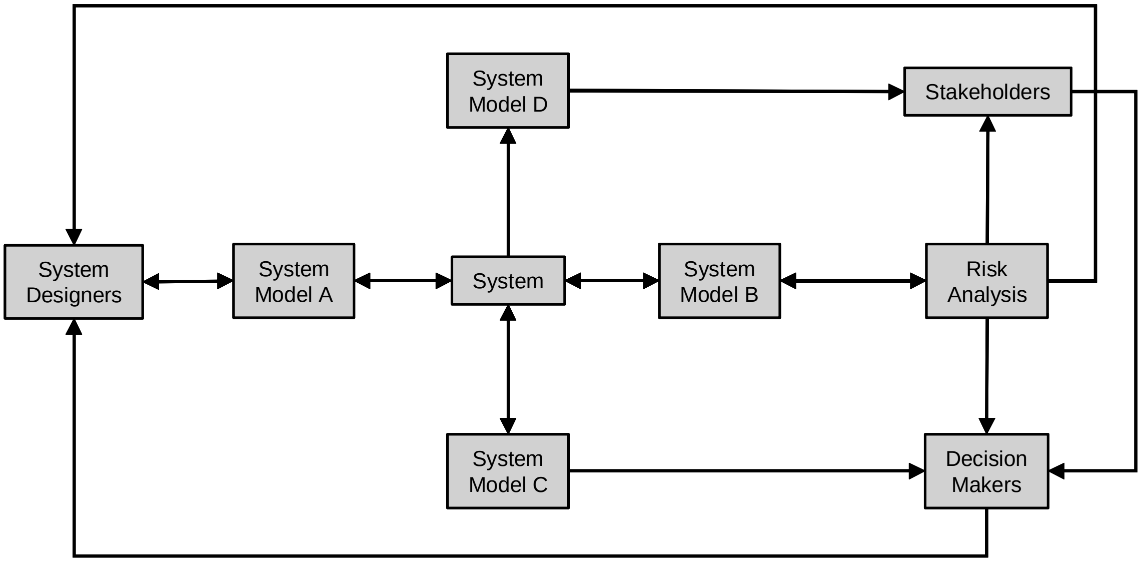


Figure 5.2 Views on a System: The Standard Approach to Risk Analysis

When risk is analyzed in an isolated process that does not involve members at all levels of the organization, risk analysts are likely to be exposed only to limited viewpoints on what the risks are and how important they are. They are therefore more likely to form an inaccurate representation of risks. In addition, such an isolated process of risk analysis misses a valuable opportunity for developing stakeholder commitment to the risk management approach.

Different perspectives on a system are useful for risk analysis because they allow different aspects of system behaviour to be identified. However, while different perspectives on a system are useful, it is necessary to develop a shared understanding of the *risk* associated with a system. Without such a shared understanding of risk it can be impossible to reach consensus on risk-related decisions or to obtain commitment to risk management strategies (e.g., use of safety procedures).

Figure 5.3 shows an alternative approach, which draws designers, decision-makers, and other stakeholders (e.g., operators) into the process from the outset. Whereas the standard approach (Figure 5.2) keeps stakeholders, risk analysts, system designers, and decision makers separate, this approach emphasizes the shared development of models of the system by encouraging the different groups to work together. In this approach, information is shared between all the groups, thus maximizing the shared knowledge of the system, facilitating the development of a shared understanding of risk and how it should be managed, and encouraging commitment to the risk management strategy.

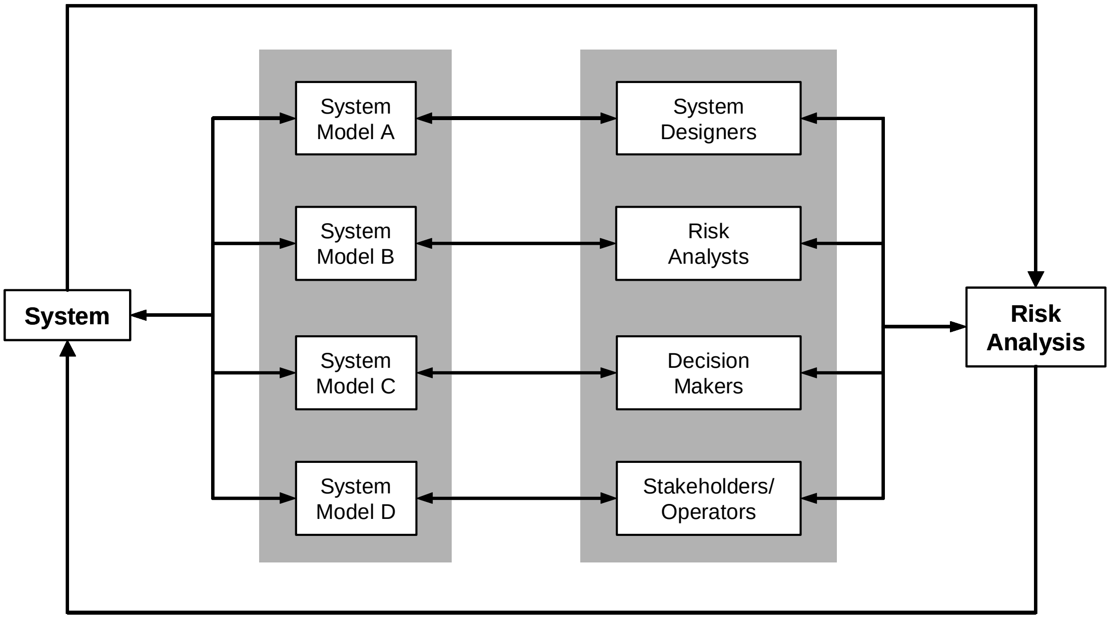


Figure 5.3 Participative risk management

Figure 5.4 shows four conceptual phases in CPRM: pre-analysis, risk analysis, risk-informed decision making, and post-analysis. The phases are separated here for convenience; in practice they overlap and inform each other. The first (pre-analysis) and last (post-analysis) phases are high-level and are to do with how the organization views, analyses, and responds to risk. The pre-analysis phase includes the selection of risk analysis methods, determination of system boundaries, and stakeholder analysis. The post-analysis phase reviews the results of risk management and adjusts the approach where necessary. The middle two phases look at particular systems, the risk associated with these systems, and how the risk can be mitigated. The risk analysis phase identifies and assesses risks for a particular system. The results of this phase are used to inform design and operating decisions in the decision making phase. Note that the items in each phase may already be performed as part of existing risk analysis techniques. They are presented here for completeness and to illustrate the concept of continuous participative risk management.

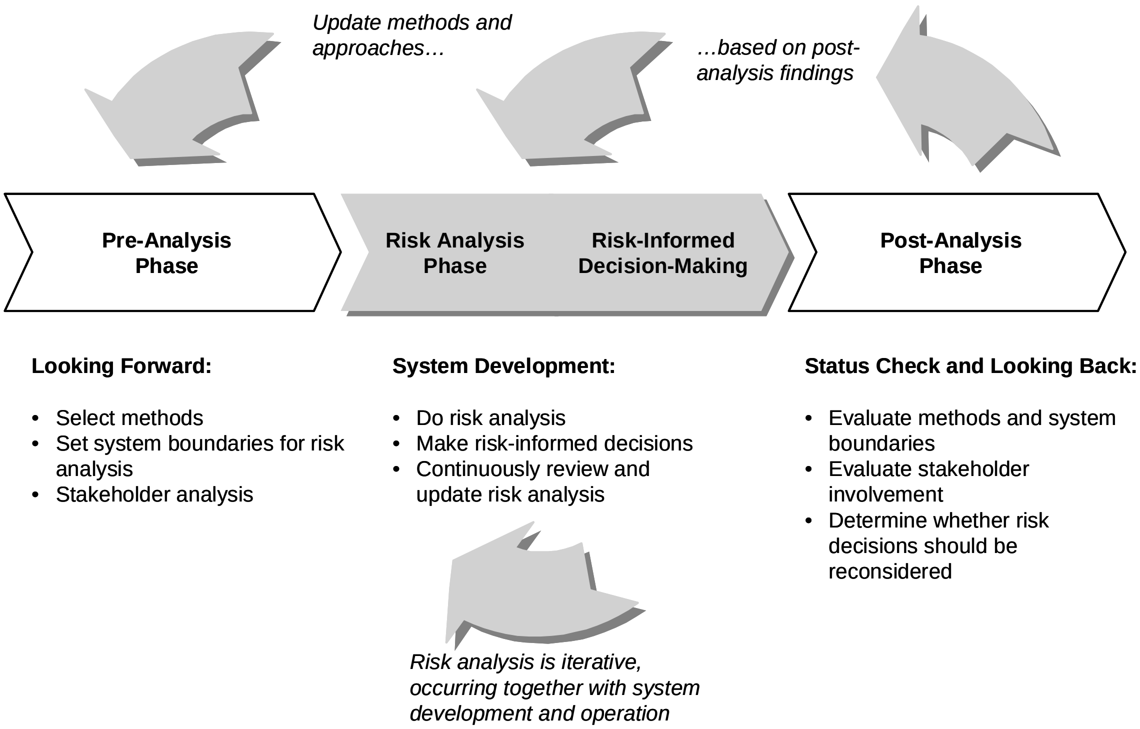

Update methods and approaches…
…based on post-analysis findings
Pre-Analysis Phase
Risk Analysis Phase
Risk-Informed Decision-Making
Post-Analysis Phase
Looking Forward:
• Select methods
• Set system boundaries for risk analysis
• Stakeholder analysis
System Development:
• Do risk analysis
• Make risk-informed decisions
• Continuously review and update risk analysis
Status Check and Looking Back:
• Evaluate methods and system boundaries
• Evaluate stakeholder involvement
• Determine whether risk decisions should be reconsidered
Risk analysis is iterative, occurring together with system development and operation

Figure 5.4 Phases in Continuous Participative Risk Management

This section developed and advocated the concept of continuous participative risk management (CPRM), which takes an explicit lifecycle view of systems and incorporates information from all members of an organization. Risk analysis is one part of risk management. Section 5.4 introduces an approach to risk analysis that enables and encourages continuous participative risk management. First, though, it is necessary to review the process that underlies the proposes risk analysis approach, STAMP-Based Hazard Analysis [Leveson, 2003].

5.2 Review of STAMP-Based Hazard Analysis

This section provides a brief overview of STPA [Leveson, 2003; Dulac and Leveson, 2004; Daouk et al., 2004]. The purpose of hazard analysis is (1) to identify system hazards and the related safety constraints required to maintain risk at an acceptable level; and (2) to determine how these constraints could be violated and use this information to eliminate, reduce, or control the hazards. The hazard analysis process is illustrated in Figure 5.5. An example of a hazard analysis is presented in Section 5.11.

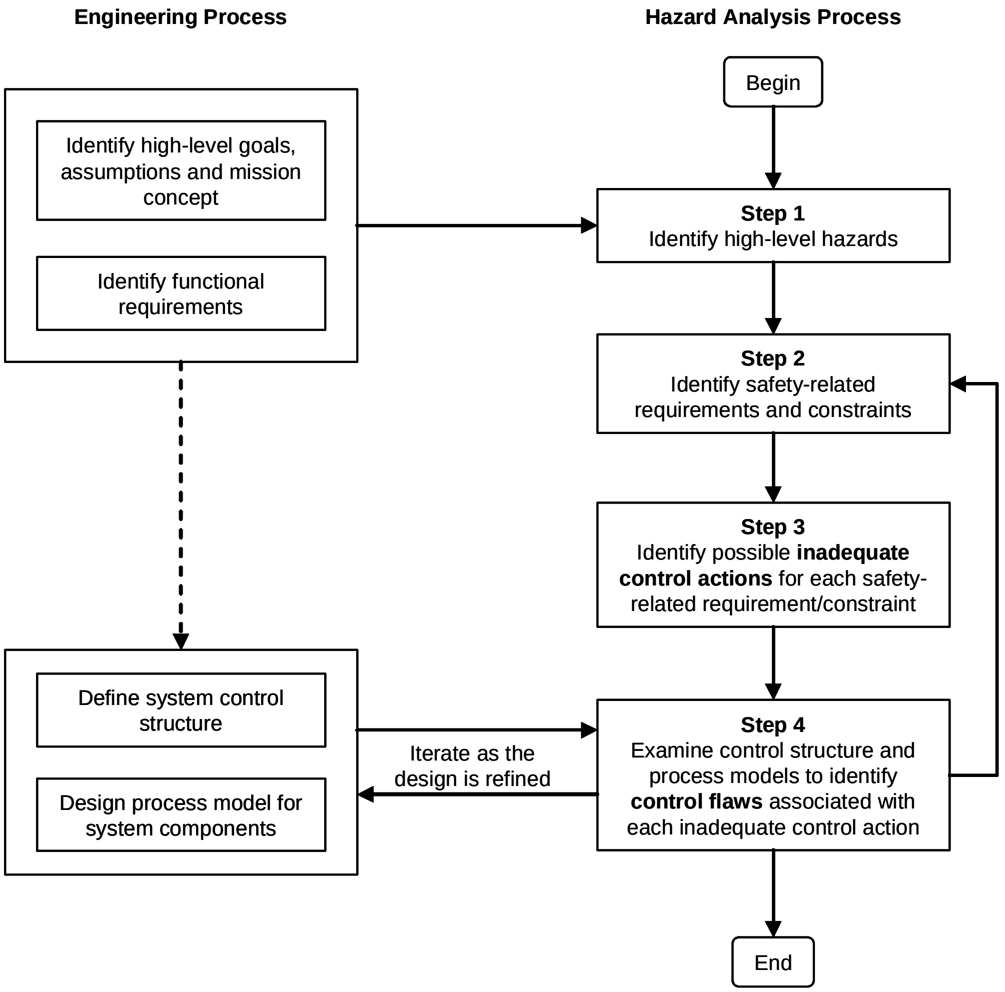

Engineering Process
Hazard Analysis Process
Begin
Identify high-level goals, assumptions and mission concept
Identify functional requirements
Step 1
Identify high-level hazards
Step 2
Identify safety-related requirements and constraints
Step 3
Identify possible inadequate control actions for each safety-related requirement/constraint
Define system control structure
Design process model for system components
Iterate as the design is refined
Step 4
Examine control structure and process models to identify control flaws associated with each inadequate control action
End

Figure 5.5 STAMP-Based Hazard Analysis[a]

[a] Adapted from [Daouk et al., 2004]. See Chapter 1 for a description of STAMP.

The four steps in the hazard analysis process (right-hand side of Figure 5.5) are briefly discussed below.

Step 1: Identify and Characterize High-Level Hazards

The first step in hazard analysis is to identify and characterize the high-level system hazards. Hazard identification is part art, part science. Analysts base identification of hazards on personal past experience, hazard identification lists that codify past organizational or industry experience, and technical expertise. Engaging employees at different levels of the organization in the process of hazard identification can be useful because they are likely to have different perspectives and may identify hazards that are not apparent to the risk analysts. The literature on hazard identification is extensive, see for example [Rasmussen and Whetton, 1997] for an approach to hazard identification in socio-technical systems.

Step 2: Identify System-Level Safety Requirements and Constraints

The second step identifies a preliminary set of system-level safety requirements and constraints that addresses the high-level hazards identified in Step 1. Safety requirements and constraints are identified by analyzing the ways that the hazard could occur. These requirements and constraints must be augmented and refined as the engineering design and hazard analysis processes proceed and more information becomes available. In the next steps, the constraints identified here are used to identify possible inadequate control actions and control flaws that could lead to the hazard and associated accident. Additional safety constraints may be identified as the system design and risk analysis proceed.

Step 3: Identify Possible Inadequate Control Actions

The third step identifies how the safety requirements and constraints identified in step two could be violated. In the STAMP model, safety requirements and constraints are violated when the control actions intended to enforce them are inadequate. There are four general ways in which a controller action may be inadequate [Leveson, 2003]:

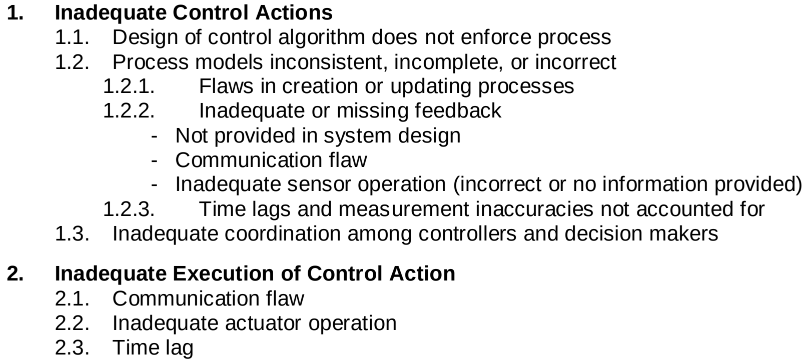
1. A required control action is not provided;

2. An incorrect or unsafe control action is provided;
3. A potentially correct or adequate control action is provided at the wrong time (e.g., too late); or
4. A potentially correct or adequate control action is stopped too early or too late.

Inadequate control actions can arise at all levels of the system and supporting organizational structure. In some cases inadequate control actions may already be addressed by existing constraints. The next step identifies the control flaws that could lead to the inadequate control actions.

Step 4: Identify Possible Control Flaws and Design Options

The fourth step uses both the system control structure and process models developed in the engineering process (left-hand side of Figure 5.5) to identify scenarios (control flaws) in which the inadequate control actions could arise. Figure 5.6 shows a classification of control flaws developed by [Leveson, 2004a]. Control flaws that could lead to inadequate control actions can be identified using the hierarchical control structure and process control models together with inputs from the underlying engineering and social science disciplines. For example, a structural analysis may be used to identify parts of an aircraft that are prone to metal fatigue.



1. Inadequate Control Actions
- 1.1. Design of control algorithm does not enforce process
- 1.2. Process models inconsistent, incomplete, or incorrect
  - 1.2.1. Flaws in creation or updating processes
  - 1.2.2. Inadequate or missing feedback
    - Not provided in system design
    - Communication flaw
    - Inadequate sensor operation (incorrect or no information provided)
  - 1.2.3. Time lags and measurement inaccuracies not accounted for
- 1.3. Inadequate coordination among controllers and decision makers

2. Inadequate Execution of Control Action
- 2.1. Communication flaw
- 2.2. Inadequate actuator operation
- 2.3. Time lag

Figure 5.6 Classification of Control Flaws [Leveson, 2004a]

The control flaws are used to generate new safety requirements and constraints, and also to inform design decisions based on an assessment of the risk associated with each hazard (see Step 1). Each identified control flaw is examined to identify whether (1) an existing safety requirement or constraint already addresses the control flaw; (2) whether additional safety requirements and constraints are required; and (3) to identify and inform design options that could eliminate, mitigate, or control the associated hazard.

The hazard analysis approach discussed here is used in a new approach to risk analysis, as described in the Section 5.4.

5.3 Quantitative and Qualitative Risk Assessments

Quantitative assessments of risks provide a convenient basis for decision making, provided they are accurate and are trusted by decision makers. Unfortunately, as discussed in Chapter 2, obtaining accurate numerical estimates of probability is difficult and often impossible. Even proponents of Probabilistic Risk Assessment note that PRA estimates should not be used in an absolute sense but only to rank risks and identify the most important contributors to these risks [Apostolakis, 2004].

In addition, research has shown that decision makers are often uncomfortable with numerical probabilities. For example, studies of the chemical industry following the Bhopal accident indicate that when organizations are really concerned about safety they discard probability estimates and focus instead on worst-case consequences [Bowman and Kunreuther, 1988; Kunreuther and Bowman, 1997]. Prior to the Bhopal accident the organizations in the above-mentioned studies used standard risk assessment techniques such as fault trees. At least one of the organizations used a strategy by which events with probabilities below a certain threshold were assumed to effectively have zero probability of occurrence. After the accident the organizations turned to worst-case scenario analyses and attempted to reduce the chances of such events as much as possible. Even when risk assessments included probability estimates these were discarded. For example, one manager noted [Kunreuther and Meszaros, 1997]:

> *"We can't deal with probabilities. We don't know what an acceptable probability is. If it's one in 28,000 years but the next year is the year, you are in big trouble."*

This general insensitivity, and sometimes antipathy, to probability estimates may also be partly due to the difficulties of making decisions in real contexts [March and Shapira, 1987]. Real decision-making situations are complex and ambiguous. Accurately estimating the probabilities of outcomes in the case of complex socio-technical systems is difficult and often impossible, and more importantly, many managers know this. It therefore makes sense for them to place less faith in probability estimates and focus on outcomes instead. One way of increasing decision makers' confidence in quantitative estimates is to document assumptions, methods, and data sources and present these together with numerical estimates.

In some cases, it may be possible to calculate probabilities quantitatively based on technical analyses or historical data. For example, structural analyses may be used to estimate the time to failure of mechanical structures.

In cases where it is not possible to unambiguously and accurately calculate probabilities, it may be better to rely on coarse quantitative estimates, such as those used in risk matrices (see Chapter 2). Quantitative estimates can be used both in an absolute sense, to indicate an analyst's sense of the probability of a particular hazard/accident, and in a relative sense, to indicate which risks are considered more or less likely than others. Thus, for example, hazards and risks might be characterized as very unlikely, somewhat likely, or very likely. When there is no information about the probability of the hazard or risk, classifying them as unknown avoids assigning arbitrary qualitative or quantitative estimates.

5.4 A New Approach to Risk Analysis

This section introduces a new approach to risk analysis that enables concurrent and integrated development of safety constraints on the one hand, and system and organizational design on the other hand. In this approach, consideration of hazards, together with other project requirements (e.g., performance, cost, schedule), drives the development of the system design as well as the definition of the supporting organizational structure. The approach allows the incorporation of risk considerations into decision making from the beginning of system development, and is inherently integrative: technical, human, and organizational factors affecting risk are included. The hazard and risk analyses occur in parallel with the system engineering process (see Figure 5.7).

The risk analysis approach here presented builds on Leveson's hazard analysis technique [Leveson, 2003] and extends it to include the risk implications of different design options and inform trade-off decisions. Figure 5.7 shows how the risk analysis adds to and complements STPA.

The risk analysis consists of the following steps as briefly described below: initial risk assessment, evaluation of design options, and residual risk assessment. The next three sections discuss each step in detail.

Step 1: Initial Estimate of High-Level Risks. In this step, initial estimates of risks are formed to facilitate ranking and to determine whether the risk level falls within acceptable limits. As in traditional system safety engineering, hazards are characterized by their probability and the range of possible severity or damage associated with the hazard, as discussed in Chapter 1 (and shown in Figure 5.8). The probability of the associated accident is a function of the probability of the hazard, the hazard duration or exposure, and the probability of the hazard leading to an accident.

Step 2: Evaluate Design Options. In this step, different design options are evaluated with regards to their impact on risk, as discussed in detail Section 5.6. Different design options may be available to implement the constraints and mitigate the control flaws associated with a particular hazard. For the purpose of highlighting trade-offs and informing decision making (see Step 3), the design options are characterized according to scope, type, effectiveness, stability, and observability. These evaluations should be updated and refined as the design is developed and more detailed information becomes available. Note also that design options may give rise to new control flaws.

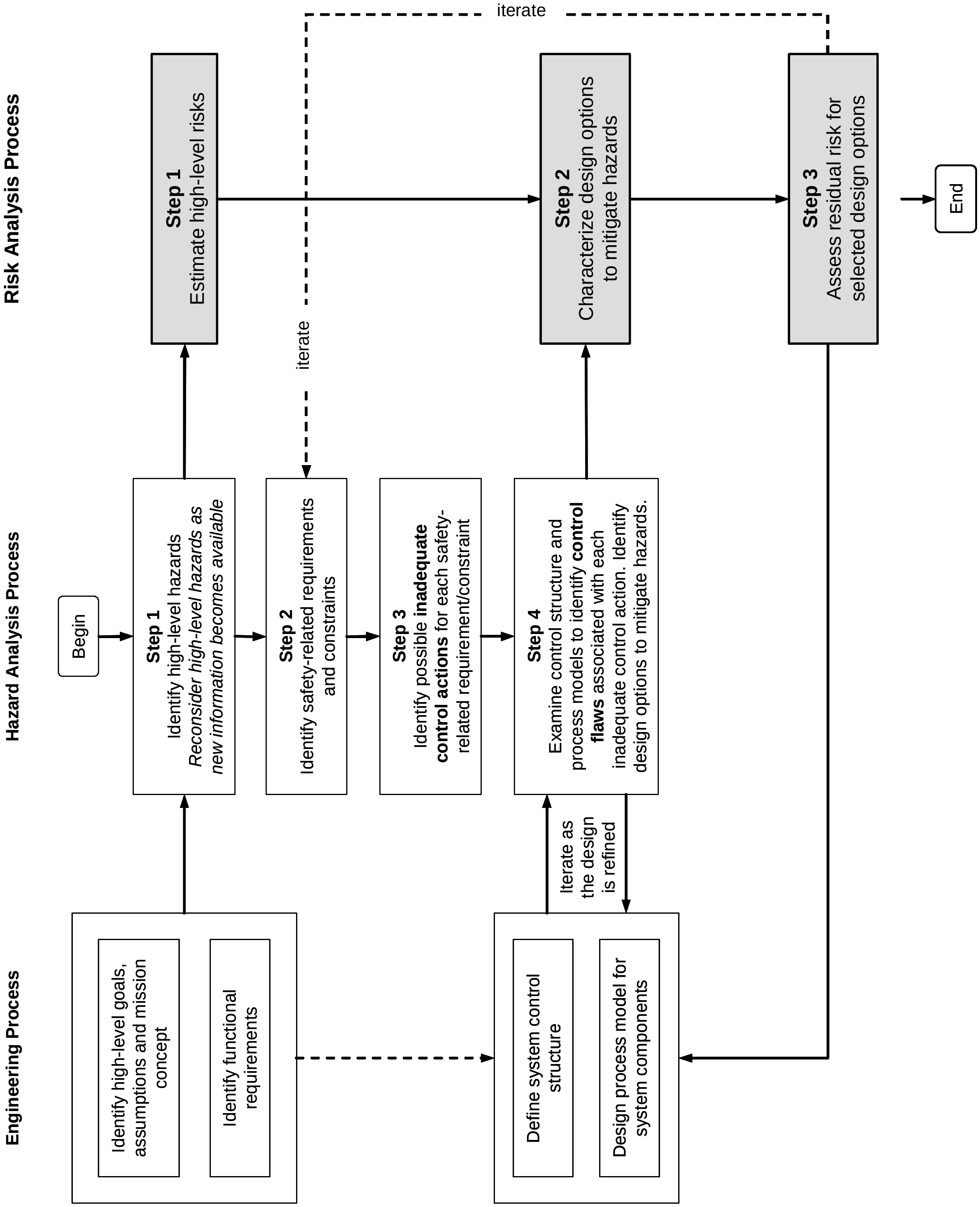

Risk Analysis Process
Hazard Analysis Process
Engineering Process
iterate
iterate
Step 1
Estimate high-level risks
Step 2
Characterize design options to mitigate hazards
Step 3
Assess residual risk for selected design options
End
Begin
Step 1
Identify high-level hazards
Reconsider high-level hazards as new information becomes available
Step 2
Identify safety-related requirements and constraints
Step 3
Identify possible inadequate control actions for each safety-related requirement/constraint
Step 4
Examine control structure and process models to identify control flaws associated with each inadequate control action. Identify design options to mitigate hazards.
Iterate as the design is refined
Identify high-level goals, assumptions and mission concept
Identify functional requirements
Define system control structure
Design process model for system components

Figure 5.7 Design, Hazard Analysis, and Risk Analysis Processes

Step 3: Residual Risk Assessment. In this step, the residual risks that remain once particular design options are selected are assessed, as discussed in detail in Section 5.7. Design options modify the risk profile formed in Step 1, by eliminating or reducing the probabilities of hazards and/or accidents, or by limiting the consequences of accidents. The residual risk assessment can be used to (1) estimate the residual risks for individual hazards given particular sets of design options; (2) estimate the overall risk of the system; and (3) make the safety case by presenting the approaches taken/not taken to mitigate hazards.

The initial risk analysis is complete when the detailed design of the all system components is complete. The analysis must be updated whenever changes in the system are considered, or whenever changes in the operating environment occur. Changes in the system or its environment may give rise to new hazards. Finally, the possibility that all hazards have not been identified must be considered. Continued vigilance is necessary to ensure an exhaustive identification of hazards.

This section has presented a high-level overview of the proposed risk analysis approach. The next three sections discuss the steps in the risk analysis in detail. Section 5.5 discusses the initial assessment of risks. Section 5.6 focusses on the evaluation of design options. Section 5.7 discusses the assessment of residual risk in detail. Section 5.8 develops a more elaborate example, based on the Space Shuttle Columbia accident.

5.5 Initial Risk Assessment (Step 1)

In Step 1 of the risk analysis, initial estimates of the risks are formed (1) to facilitate ranking of these risks and (2) to determine whether the initial estimate of risk falls within an acceptable range to continue system development.

This section first presents the parameters that will be used to represent risk. These parameters are common to most risk analysis methodologies and are presented here for completeness. Next, an example is used to illustrate the process of initial risk assessment.

Hazards are characterized by the probability, $p(H))$, of their occurrence and by the range of possible consequences, $\mathbf{C}$, of accidents associated with the hazard, as shown in Figure 5.8 [Leveson, 1995]. The probability of an accident is a function of the probability of the hazard, the hazard duration or exposure, and the probability of the hazard leading to an accident, $p(A|H)$. The hazard duration or exposure may affect the probability of an accident. In general, the longer a hazard persists, the more likely it is to lead to an accident.

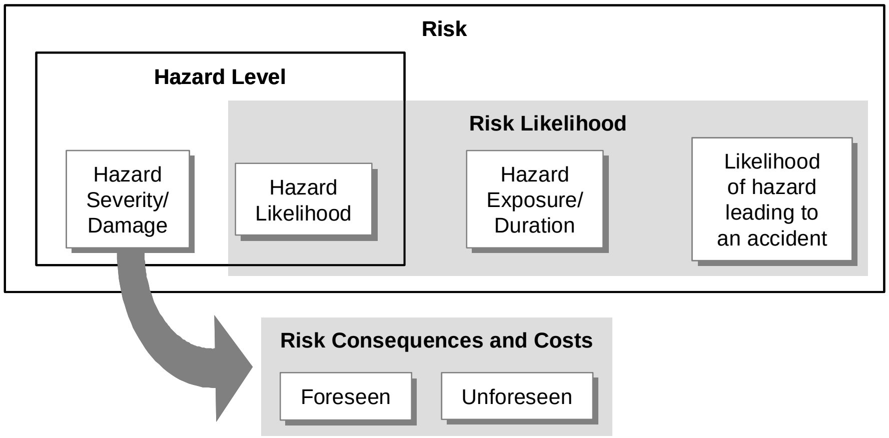

Risk
Hazard Level
Risk Likelihood
Hazard Severity/ Damage
Hazard Likelihood
Hazard Exposure/ Duration
Likelihood of hazard leading to an accident
Risk Consequences and Costs
Foreseen
Unforeseen

Figure 5.8 Components of Risk. Adapted from [Leveson, 1995]

Based on these definitions of hazard and risk, each risk is characterized here by three components: (1) the probability, $p(H)$, of the hazard; (2) the conditional probability, $p(A|H)$, of an accident given the hazard (which is defined here to include the effect of hazard exposure); and (3) the consequence vector $\mathbf{C}$. This characterization is captured in Eq. (5.1):

$$\mathbf{R} = [p(H), p(A|H), \mathbf{C}] \tag{5.1}$$

This characterization of risk can also be represented graphically using risk matrices, as discussed in Chapter 2.

The initial estimate of the probability of a hazard H takes the form:

$$p_o(H) \in \{\text{Quantitative Estimate, Qualitative Estimate}, \text{Unknown}\} \tag{5.2}$$

For a completely new system design, the probabilities of all hazards and accidents are initially classified as unknown, to indicate that nothing is yet known about the design or the risks. As the design is developed and more information becomes available probabilities can be estimated and updated based on the hazard analysis and the design options selected to address the hazard, as discussed later.

The proposed process of risk analysis in general, and initial risk assessment in particular, are best demonstrated by means of a simple example, as shown next. The hazard analysis for this example is presented in Section 5.11 for reference. Section 5.8 presents a more complex example.

Initial Risk Assessment: Gas Leak. Consider the following hypothetical scenario. An office building is being designed, and one concern is that gas leaks in offices could be set off by ignition sources such as matches or electric sparks. For the sake of this example, it is assumed that ignition sources are beyond the control of the building developers. In this case, the hazard, risk, and accident are, respectively:

- H: Gas leak in the office.
- R: Probability and consequences of gas-leak induced fire in the office.
- A: Fire in the office and possibly the rest of the building.

The hazard has been narrowly defined to keep the example relatively simple. In a real risk analysis, the hazard would be more broadly defined to include any areas where ignition sources may occur, such as storage rooms and restrooms.

The probability of the hazard is the probability that a leak arises in the gas distribution system. Note that this probability depends on various factors, such as the quality of materials used to construct the gas distribution system, the quality of the installation, the frequency of inspections, and so forth. The damage associated with the hazard is the damage caused by the fire to the building, and the possible injuries and loss of life. The hazard exposure is the time for which a gas leak coincides with an ignition source such as a flame from a match. Because a spark need only coincide with a gas leak for an instant, the probability of an accident given a gas leak is the same as the probability of a spark during a gas leak. Finally, the immediate consequences of the accident are the damage to the building and injury and/or loss of life. Other consequences include the impact of the damage to the building on the organization that uses the building (e.g., productive time lost due to relocation efforts). Figure 5.9 shows how the gas leak hazard is related to the risk of a fire in the office and building.

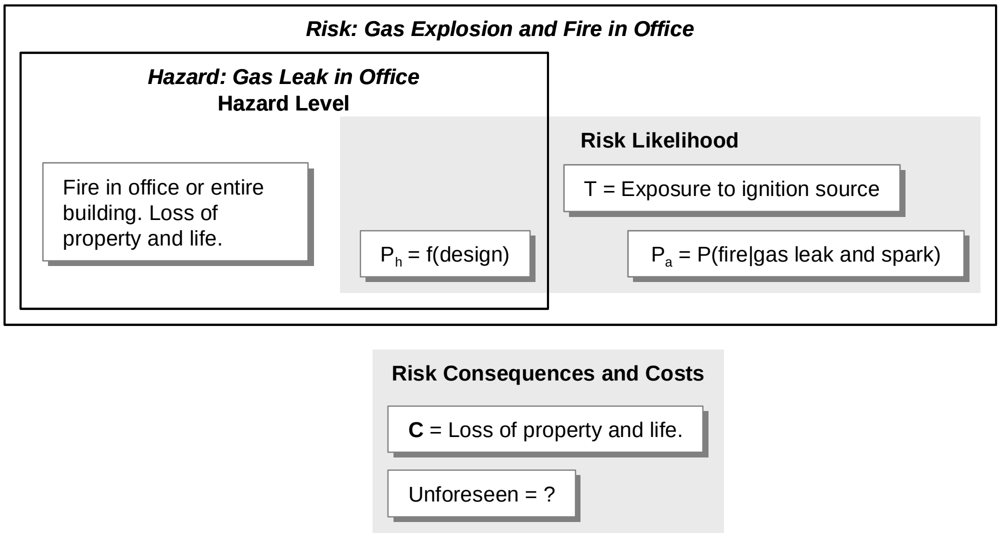

Risk: Gas Explosion and Fire in Office
Hazard: Gas Leak in Office
Hazard Level
Fire in office or entire building. Loss of property and life.
P_h = f(design)
Risk Likelihood
T = Exposure to ignition source
P_a = P(fire|gas leak and spark)
Risk Consequences and Costs
C = Loss of property and life.
Unforeseen = ?

Figure 5.9 Gas Leak Hazard and Fire Risk

The risk of an accident associated with a gas leak fire can be summarized as:

$$\mathbf{R}(\text{gas leak fire}) = \begin{bmatrix} p(H) = f(\text{gas distribution system}) \\ p(A|H) = p(spark) \\ \mathbf{C}(\text{fire}) \end{bmatrix} \tag{5.3}$$

where

$$\mathbf{C}(\text{gas leak fire}) = \begin{bmatrix} \text{damage to property} \\ \text{injury and loss of life} \\ \text{etc.} \end{bmatrix} \tag{5.4}$$

In each case, quantitative or qualitative estimates are obtained by consulting the applicable domain experts, performing tests and analyses, and so forth.

5.6 Evaluation of Design Options (Step 2)

Design options are used to reduce risk by controlling hazards. Design options to control hazards may be developed in Step 1 of the hazard analysis. Additional design options to address control flaws may also be developed in Step 4 of the hazard analysis. Several design options may be available to address a hazard and/or the associated control flaws.

This section discusses the evaluation of these design options from a risk perspective (Step 2 of the risk analysis). These evaluations are used to assess the residual risk for different design options, or sets of design options (Step 3 of the risk analysis). The residual risk assessment can be used, together with other considerations such as the performance and cost implications of the design options, to inform design decisions.

For the purpose of risk assessment, the design options are characterized according to the following attributes: scope, type, effectiveness, stability, and observability. These attributes are briefly discussed in the following subsections. Note that these characterizations are not necessarily a new way of looking at systems, but are presented here because of their implications for risk.

5.6.1 Scope

The scope of a design option indicates the extent of the mitigation potentially achieved by the design option. A design option can address the hazard, one or more inadequate control actions, or one or more control flaws. Design options that address the hazard will in general tend to have a greater risk reduction effect than those that address inadequate control actions, which in turn have a greater risk reduction effect than those that address control flaws. For example, a design option that eliminates a hazard has a greater risk reduction effect than one that eliminates a control flaw that could lead to that hazard.

The design options suggested in the gas leak example all directly address the hazard (listed below by decreasing impact on risk reduction):

- D1.1: Remove gas line from offices/do not install gas lines in offices—hazard eliminated.
- D1.2: Use leak-resistant tubing and joints—probability of hazard reduced.
- D1.3: Periodically inspect gas lines and joints for damage and/or leaks—probability of hazard reduced.
- D1.4: Use leak sensors and shut off gas in event of gas leaks—hazard controlled.
- D1.5: Detect gas leaks and evacuate building—damage minimized.
- D1.6: Detect explosions/ fires and activate building sprinkler system—damage minimized.

The scope parameter is used for convenience in the risk assessment, where it guides the assessment of probabilities and/or consequences.

5.6.2 Type

The type of the design option indicates the type of mitigation potentially achieved by the design option. There are four complementary approaches to mitigating hazards, the inadequate control actions (ICAs) that could lead to the hazards, or the underlying control flaws (CFs), listed here in order of preference [Leveson, 1995, Ch. 16]:

1. Eliminate Hazard, Inadequate Control Action, or Control Flaw. It may be possible to eliminate the hazard, inadequate control actions or control flaws by making certain design decisions. It will not always be practical to eliminate a hazard, inadequate control action, or control flaw. For example, some elimination options may be judged to be infeasible because they are too expensive, too difficult to implement, conflict with essential performance goals, or contribute to existing or additional hazards. For the gas leak example, D1.1 (remove gas line from offices/do not install gas lines in offices) eliminates the hazard.

If a hazard is eliminated from the design, no further action in the risk analysis is necessary. Monitoring the system as it evolves over time can reveal if the hazard arises again. If the hazard cannot be eliminated (because no options are available or because the available options are judged to be infeasible), the next three approaches can be used singly or in combination to decrease the probability of the hazard, control the hazard, or minimize the damage resulting from the hazard.

2. Reduce Probability. It may be possible to reduce the likelihood of the hazard, inadequate control action, or control flaw. For the gas leak example, D1.2 (use leak-resistant tubing and joints) and D1.3 (periodically inspect gas lines and joints for damage and/or leaks) reduce the probability of the hazard.

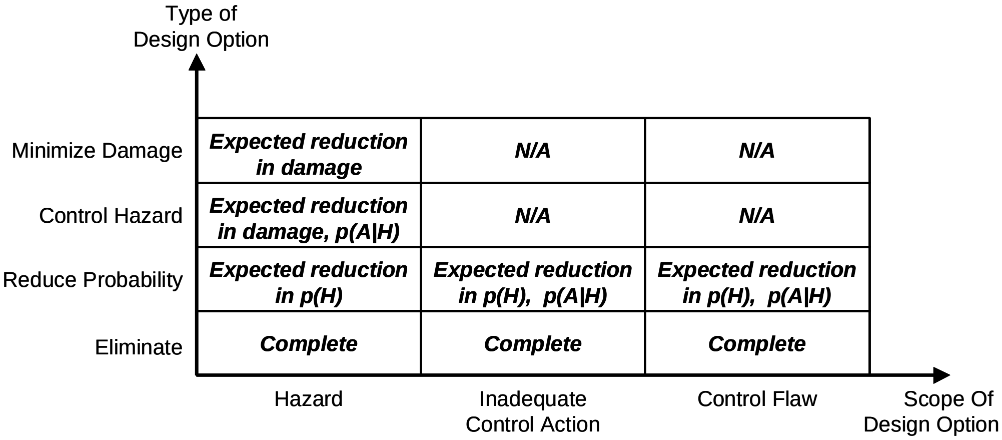
3. Control Hazard. It may be possible to control the hazard if it does occur. For the gas leak example, D1.4 (use leak sensors and shut down gas in event of gas leaks) controls the hazard.

4. Minimize Damage. It may be possible to minimize the damage resulting from the hazard. For the gas leak example, D1.5 (detect gas leaks and evacuate building) and D1.6 (detect explosions/ fires and activate building sprinkler system) minimize the damage associated with the hazard.

Design options that eliminate hazards have the highest risk reduction potential (see also Figure 5.14). In contrast, design options that reduce the probability of a control flaw or minimize damage have lower risk reduction potential. The notion of design option, discussed next, further expands on the risk reduction potential of a design option.

5.6.3 Effectiveness

The effectiveness parameter indicates how effective the design option is expected to be at a particular scope and type of mitigation effort. The characterization of effectiveness depends on the type of design option, as shown in italics in Figure 5.10. Thus, a design option that eliminates a hazard, inadequate control action, or control flaw, is by definition fully effective. The effectiveness of a design option that reduces the probability of a hazard, inadequate control action, or control flaw is the reduction in probability obtained with the design option. In the gas leak example, different types of gas lines may have different probabilities of leaking. The effectiveness of a design option that controls a hazard is the expected reduction in damage or the expected reduction in the probability of an accident, given the hazard. For example, the reduction in accident probability obtained by shutting gas leaks off depends on how quickly leaks are shut off. Finally, the effectiveness of a design option that minimizes damage is the expected reduction in damage. For example, the expected reduction in injuries resulting from a gas leak fire depends on how quickly the evacuation alarm is sounded, and on how well the building occupants are trained in evacuation procedures.



	Hazard	Inadequate Control Action	Control Flaw
Minimize Damage	***Expected reduction in damage***	***N/A***	***N/A***
Control Hazard	***Expected reduction in damage, p(A|H)***	***N/A***	***N/A***
Reduce Probability	***Expected reduction in p(H)***	***Expected reduction in p(H), p(A|H)***	***Expected reduction in p(H), p(A|H)***
Eliminate	***Complete***	***Complete***	***Complete***

Figure 5.10 Effectiveness of Design Options

The effectiveness of each design option can be estimated by risk analysts in concert with domain experts. For example, traditional risk analysis techniques such as fault and event trees may be used once the design is complete to help determine the ideal residual probabilities (see Section 5.7) of hazards or accidents where the systems are simple and do not involve software. When it is not possible to evaluate the effectiveness quantitatively, qualitative assessments can be used instead.

For the gas leak example introduced on page 168 (H: Gas leak in the office), the effectiveness of the proposed design options can be evaluated as follows, using the inadequate control actions and control flaws identified in Steps 2 and 3 of the hazard analysis (see Section 5.11) as a guide:

- D1.1: Remove gas line from offices/do not install gas lines in offices.

  Scope: Hazard; Type: Eliminate

  Effectiveness: Complete—Removing gas lines from offices eliminates the hazard. Although several inadequate control actions were identified that could limit the how well this option is implemented, they can be mitigated quite easily by means of inspections.

- D1.2: Use leak-resistant tubing and joints;

  Scope: Hazard; Type: Reduce probability

  Effectiveness: Reduced probability of hazard—Reduction can be estimated based on technical specifications of tubing and joints.

- D1.3: Periodically inspect gas lines and joints for damage and/or leaks;

  Scope: Hazard; Type: Reduce probability

  Effectiveness: Reduced probability of hazard—Reduction can be estimated based of technical specifications of tubing and joints, maintenance schedule, and human factors analyses.

- D1.4: Use leak sensors and shut off gas in event of gas leaks;

  Scope: Hazard; Type: Control hazard

  Effectiveness: Reduced damage—The reduction in damage depends on how rapidly the gas is shut off.

- D1.5: Detect gas leaks and evacuate building;

  Scope: Hazard; Type: Minimize damage

Effectiveness: Reduced damage—The reduction in damage depends on how rapidly the gas leak is detected and how effective the evacuation procedures are.

- D1.6: Detect explosions/ fires and activate building sprinkler system.

  Scope: Hazard; Type: Minimize damage

  Effectiveness: Reduced damage—The reduction in damage depends on how rapidly the gas leak is detected and how effective the building sprinkler system is.

Subsequent iterations of the risk analysis may yield information that changes the design option evaluation. For example, they may reveal that a design option creates additional inadequate control actions that could contribute to the present or other hazards.

***

The first three design option parameters, scope, type, and effectiveness, are used to evaluate the risk reduction potential of a design option. Design options may however be improperly implemented, or decline in effectiveness over time, thus leading to an increase in risk.

The next two design option parameters, stability and observability, are used to identify potential reasons for a decline in effectiveness, and to assess how difficult it is to determine whether a design option is properly implemented and whether it continues to be effective. Determining the ability of design options to enforce constraints over the lifetime of the system (stability) can help decision makers to make decisions that address risk over the lifetime of the system. In addition, un-noted declining constraints can result in increased risk. Considering the ease with which declining constraints can be observed (observability) further aids in informed decision making about risk over the system lifetime. These two parameters also provide a guide as to which aspects of the system should be monitored for signs of increasing risk over time.

5.6.4 Stability

The stability of the design option indicates, where applicable, how rapidly the design option effectiveness may decline over the system's lifetime. The stability of design options indicates the degree of continued vigilance that is required to ensure the design option remains effective. The stability parameter is used in the risk assessment to identify potential areas of concern where risk may increase over time.

Design options that do not require any further attention such as inspections or preventative maintenance and are not subject to physical decline over time have maximum stability. Thus, a design option that eliminates a hazard, inadequate control action, or control flaw has maximum stability.

For other design options, the stability is the expected time for which the design option is expected to remain effective. Where the design option consists of physical elements, the stability is captured by the overall mean-time-to-failure (MTTF) of these elements. For example, brake pads with a higher MTTF have higher stability than those with lower MTTF. Low MTTF brake pads will require more frequent inspections and replacements to ensure that they continue functioning as desired. Design options that require continued attention may become less effective over time if the quality of maintenance and inspection declines.

The effectiveness of procedures in mitigating hazards is likely to decline over time unless active efforts are made to ensure continued compliance. For example, safety procedures often include steps that are intended to address potential problems. Under normal conditions, it may be possible to omit these steps without serious consequences [Leveson, 1995]. Employees may become complacent when they note that ignoring steps or performing them sloppily usually does not have immediately observable effects on safety.

In practice it will usually be impossible to determine a specific time at which procedure compliance is expected to decline. Noting in the risk analysis documentation that procedure compliance may decline over time if active measures are not taken to counteract this decline serves two purposes. First, it provides designers with another factor that affects the continued risk-reduction potential of design options. Second, it indicates to risk managers areas where a decline in effectiveness, and hence and increase in risk, is possible.

For the gas leak example, the stability of the design options is as follows:

- D1.1: Remove gas line from offices/do not install gas lines in offices.

  Stability: Maximum—Hazard is eliminated. Gas lines will remain where placed unless active effort is made to move them (e.g., building refurbishment).
- D1.2: Use leak-resistant tubing and joints.

  Stability: Subject to Organizational Factors—Gas lines and joints may be replaced with lower quality items during periodic maintenance.
- D1.3: Periodically inspect gas lines and joints for damage and/or leaks.
- Stability: Subject to Organizational Factors—Inspections may become less frequent/thorough over time.
- D1.4: Use leak sensors and shut down gas in event of gas leaks.

  Stability: MTTF of leak sensors. Periodic inspection and maintenance necessary to prevent deterioration of sensors.
- D1.5: Detect gas leaks and evacuate building.

  Stability: MTTF of leak sensors. Periodic inspection and maintenance necessary to prevent deterioration of sensors.

  Evacuation procedure must be practised at periodic intervals to ensure it is effective.
- D1.6: Detect explosions/ fires and activate building sprinkler system.

  Stability: MTTF of fire sensors and sprinkler systems. Periodic inspection and maintenance necessary to prevent deterioration of sensors and sprinkler system.

5.6.5 Observability

The observability parameter indicates how easy it is to determine (1) whether a design option is implemented properly and (2) how effective it is in practice at mitigating the hazard. The observability parameter is used in the risk assessment to identify potential areas of concern where it may be difficult to observe an increase in risk over time. When it is difficult to determine whether a design option is ineffective, the risk level may be higher than initially assessed.

Monitoring design options generally consists of both technical and organizational aspects. For example, special equipment may be needed to examine an aircraft for cracks in the fuselage, but inspection procedures must be followed to ensure that inspection is carried out properly and at the required time intervals. The observability parameter therefore includes both the technical and organizational challenges of evaluating the actual effectiveness of design options. Technical challenges depend on the technical features of the design option and can be identified and evaluated by domain experts.

The ease with which procedures can be monitored and non-compliance detected, depends on various factors. For example, more complicated procedures are in general more difficult to monitor than simple procedures because there are more ways for them to be violated, either intentionally or unintentionally. Also, procedures that require specialized skills cannot be effectively monitored by personnel who do not have the same specialized skills. Such procedures therefore have lower observability than procedures that require less specialized skills.

For the gas leak example, the observability of the design options is as follows:

- D1.1: Remove gas line from offices/do not install gas lines in offices.
  Observability: Simple inspections can determine placement of gas lines.
- D1.2: Use leak-resistant tubing and joints.

  Observability: Simple inspections can ensure designs specify appropriate equipment. However, continued vigilance necessary to ensure any replacement meet requirements.
- D1.3: Periodically inspect gas lines and joints for damage and/or leaks.

  Observability: May be difficult to determine whether inspectors perform inspections properly.
- D1.4: Use leak sensors and shut down gas in event of gas leaks;

  Observability: Inspection and analysis of the design can be used to determine whether it is correct. Testing of the installed leak sensors and shut down valves can determine whether this option is effective. Inspections required to detect deterioration of gas sensors. May be difficult to determine if inspectors perform inspections properly.

- D1.5: Detect gas leaks and evacuate building;

  Observability: Inspections required to detect deterioration of gas sensors. May be difficult to determine if inspectors perform inspections properly. Evacuation procedures must be tested in dry-runs to determine whether they are effective.

- D1.6: Detect explosions/ fires and activate building sprinkler system.

  Observability: Inspections required to detect deterioration of sensors or emergency response systems. May be difficult to determine if inspectors perform inspections properly.

5.7 Residual Risk Assessment (Step 3)

This section shows how the initial risk estimates made in Step 1b can be updated to reflect the impact of design options. The purpose of this stage of the risk assessment is threefold:

1. To aid in the selection of design options;
2. To estimate the residual risk for the selected design; and
3. To show the extent to which hazards have been addressed in the design.

This section addresses these goals by developing an analytical framework to estimate the probabilities of hazards and accidents in the presence of one or more design options. Assumptions and caveats that govern the use of the equations are presented. Then, the determination of the individual probabilities used in the equations is discussed. Next, a method for incorporating the effects of organizational factors into the assessment is presented. Finally, the gas leak example is continued to illustrate the application of the concepts presented here. Section 5.8 presents a more complex example, based on the Space Shuttle.

5.7.1 An Analytical Framework for Residual Risk Assessment

This section develops an analytical framework for residual risk assessment that allows risk analysts to form quantitative (when quantitative data are available) or qualitative (when quantitative data are not available) assessments of the probabilities of hazards and accidents. These assessments can be used to select design options to address risks, and to determine whether risks have been adequately addressed.

In the STAMP model of accidents, hazards occur when constraints are violated as a result of a control flaw [Leveson, 2004a]. Accidents occur when the hazard persists, the appropriate environmental conditions exist, and the constraints intended to prevent hazards from leading to accidents are not effective[1]. Figure 5.11 illustrates the concept.

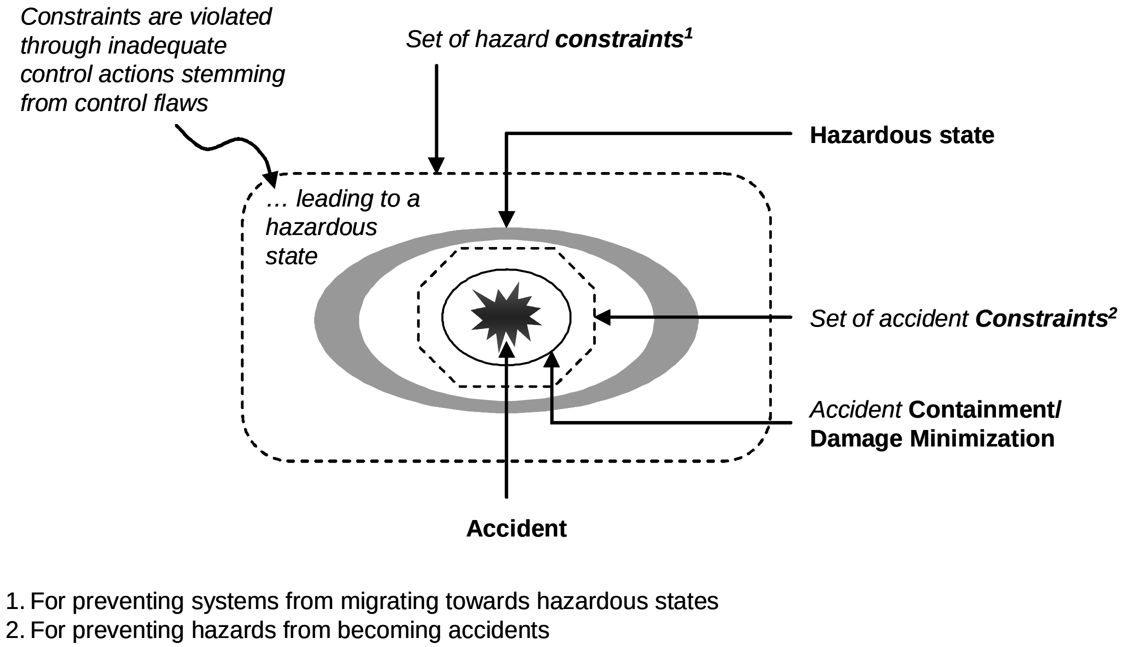

Constraints are violated through inadequate control actions stemming from control flaws
Set of hazard constraints[1]
Hazardous state
… leading to a hazardous state
Set of accident Constraints[2]
Accident Containment/ Damage Minimization
Accident
1. For preventing systems from migrating towards hazardous states
2. For preventing hazards from becoming accidents

Figure 5.11 Preventing Hazards and Accidents

The probability that a constraint is violated is a function of the probabilities that the associated control flaws occur. The probabilities of hazards and accidents can therefore be determined by first determining the probabilities that the relevant constraints are violated,

[1] For convenience and readability, constraints, inadequate control actions, and control flaws are here referred to by subscripts and not the hierarchical numbering system used elsewhere in the risk analysis.

and then determining the probabilities of hazards and accidents given the violation of constraints, as discussed next.

The remainder of this section is organized as follows. First, the probability that a constraint is violated as a result of control flaws is determined. Second, the impact of design options on the probabilities of constraint violation is determined. Third, the probability that a hazard occurs as a result of constraint violations is determined. Finally, the probability that a hazard develops into an accident is determined. Figure 5.12 illustrates the development of the framework.

1. Violation of Constraints

Constraints may be violated when one or more control flaws occur. For example, in air traffic control, the constraint that two aircraft must maintain minimum separation (in order to avoid mid-air collisions) may be violated as a result of several control flaws, such as incorrect instructions to one or both of the aircraft from air traffic controllers [Leveson, 1995]. However, control flaws do not guarantee that constraints will be violated. In the aircraft example, the pilots may notice that the air traffic control instructions are incorrect and perform evasive manoeuvres. The probability that the aircraft violate minimum separation can be determined by determining the probabilities of that the control flaws occur, as well as the probabilities that these control flaws lead to constraint violations.

In general, the overall probability of a constraint violation is a function of (1) the probabilities of the associated control flaws and (2) the probabilities that individual control flaws will result in constraint violations. Consider now how the individual control flaw probabilities can be combined to obtain the overall probability of constraint violation. Begin with the simplest case where there is only one control flaw under the constraint. C is defined as the event where the constraint is effective, and $\overline{C}$ is the event where the constraint is violated.

The probability of the constraint being violated, $p(\overline{C})$ is given by:

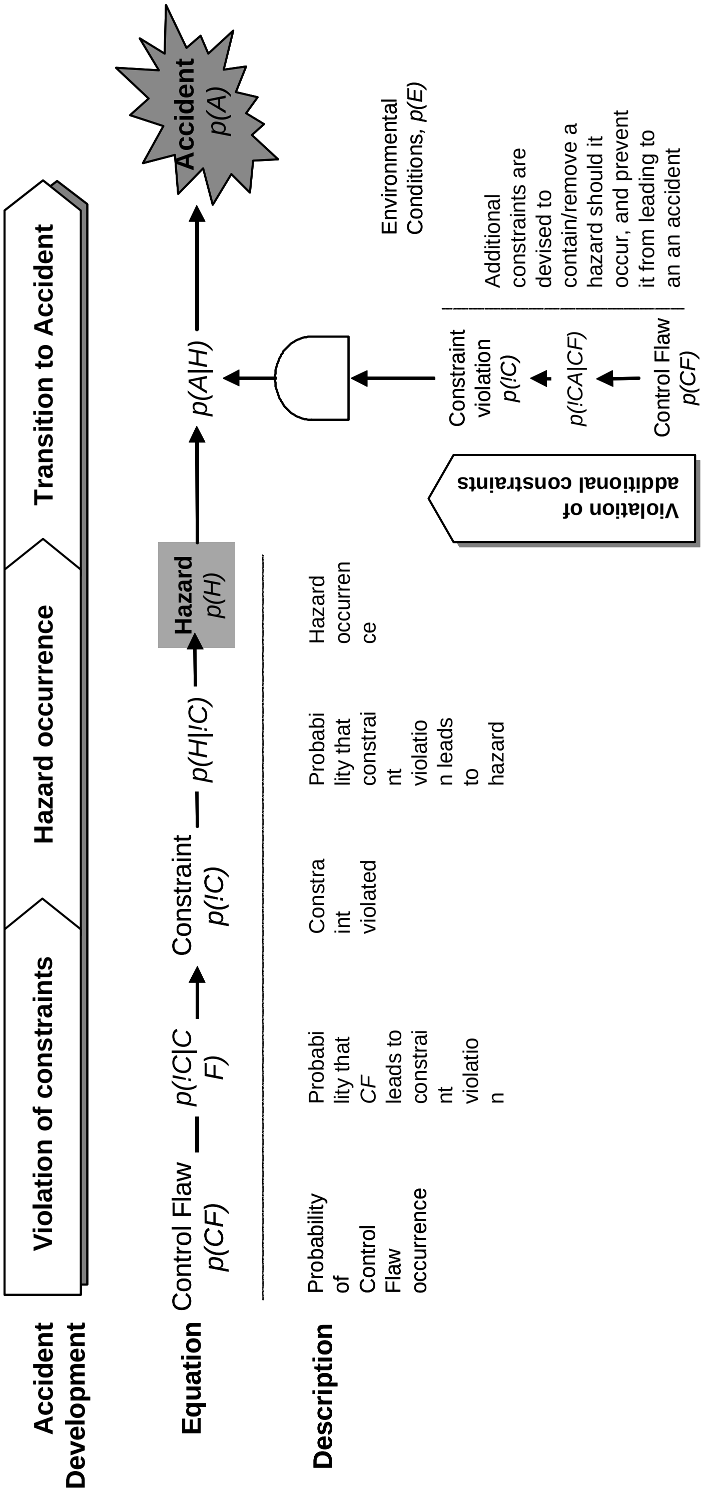

Accident Development
Violation of constraints
Hazard occurrence
Transition to Accident
Equation
Control Flaw p(CF)
p(!C|CF)
Constraint p(!C)
p(H|!C)
Hazard p(H)
p(A|H)
Accident p(A)
Description
Probability of Control Flaw occurrence
Probability that CF leads to constraint violation
Constraint violated
Probability that constraint violation leads to hazard
Hazard occurrence
Environmental Conditions, p(E)
Violation of additional constraints
Constraint violation p(!C)
p(!CA|CF)
Control Flaw p(CF)
Additional constraints are devised to contain/remove a hazard should it occur, and prevent it from leading to an an accident

Figure 5.12 Guide to Development of Risk Assessment Framework

$$p(\bar{C}) = p(\bar{C}|CF) \cdot p(CF) \tag{5.5}$$

where $p(\bar{C}|CF)$ is the conditional probability of constraint violation given a control flaw, and $p(CF)$ is the probability of the control flaw occurring.

When there are several control flaws, the probability of the constraint being violated is determined by noting that the constraint will be satisfied if none of the control flaws occurs:

$$p(\bar{C}) = 1 - \prod_{i=1}^{n_{CF}} (1 - p(\bar{C}|CF_i) \cdot p(CF_i)) \tag{5.6}$$

where n_{CF} is the number of control flaws associated with the constraint and it is assumed that control flaws are independent. The development of techniques to address dependencies between control flaws is left as a subject for future work.

The probabilities of individual control flaws in Eq. (5.6) can be estimated by domain experts together with risk analysts. Quantitative estimates of control flaw probabilities should be used where possible. Qualitative estimates can be used where quantitative estimates are not possible. Some control flaws, such as those due to technical design flaws, are guaranteed to occur, and therefore in such cases:

$$p(CF_{\text{technical design flaw}}) = 1 \tag{5.7}$$

Similarly, determining the probability of constraint violation given a control flaw is also domain specific. Some control flaws will guarantee that the constraint is violated, that is, $p(\bar{C}|CF) = 1$. For example, in the gas leak case, CF1.2.1.3 (the designers erroneously select the wrong type of gas lines and joints) guarantees that the incorrect equipment will be installed.

When the occurrence of any of the control flaws guarantees that the constraint is violated:

$$p(\bar{C}|CF_i) = 1 \qquad \forall i \tag{5.8}$$

In this case, Eq. (5.6) therefore simplifies to:

$$p(\bar{C}) = 1 - \prod_{i=1}^{n_{CF}} (1 - p(CF_i)) \tag{5.9}$$

Other control flaws may not guarantee that the constraint is violated, but do make it more likely that it is violated. For example, in the gas leak case, CF1.1.1.1 (The requirement that gas lines not be placed in offices is not communicated to the designers and the placement of gas lines is not monitored by safety personnel) does not guarantee that designers will place gas lines in offices.

Eq. (5.6) shows how the probabilities of individual control flaws relate to the probability that the associated constraint is violated. This expression can be used both to determine the overall probability of constraint violation, and also to determine the relative importance of control flaws, as discussed below.

Relative Significance of Control Flaws. The contributions of control flaws to Eq. (5.6) defines their relative significance to the probability of constraint violation and therefore indicates to designers which control flaws should be addressed to minimize the probability of constraint violation. Control flaws that are highly probable and whose occurrence makes it highly probable that the constraint will be violated have the largest effect on the overall probability of constraint violation. Thus, where quantitative probabilities are available, the terms $p(\bar{C}|CF_i) \cdot p(CF_i)$ can be used to rank the control flaws and determine where to focus the risk mitigation effort in order to maximize the probability that a particular constraint is not violated.

Where only qualitative probabilities are available, a graphical approach such as that shown in Figure 5.13 can be used. When the probability of a control flaw is low, and the probability of a constraint violation given the control flaw is also low, the control flaw has relatively low impact on the overall probability of constraint violation (e.g., CF_1 in the figure). Conversely, when the probability of a control flaw is high, and the probability of a constraint violation given the control flaw is also high, the control flaw has relatively high impact on the overall probability of constraint violation (e.g., CF_9 in the figure). Identifying the control flaws that have the largest impact on the probability of constraint violation allows designers to focus their risk mitigation efforts on those flaws and therefore develop the most effective risk mitigation strategies.

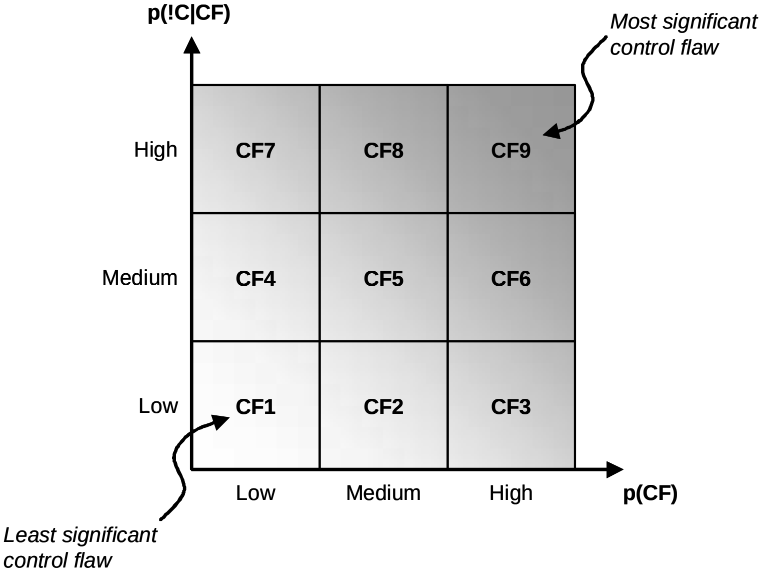


Figure 5.13 Relative Impact of Control Flaws

2. Impact of Design Options

Design options can be used to reduce (or eliminate) the probability that constraints will be violated by reducing (or eliminating) the probability of control flaws or constraint violations. For example, the probability that air traffic controllers give incorrect instructions that cause pilots to violate minimum separation might be reduced by setting appropriate operator workloads so that operator fatigue is avoided. The impact on risk of design options can be assessed by determining the effect of the design option on the control flaw probability. The updated control flaw probability is then used to update the probability of constraint violation, as discussed below, and hence the probability of the hazard and accident, as discussed later. Design options may also directly address the hazard. In this case the impact on risk is determined directly, as shown in the example at the end of this section.

Consider now how the effect of design options on the probability of constraint violation can be assessed. In the simple case where there is only one control flaw, CF_k, under the constraint, and this control flaw is addressed by a single design option, D_k, the probability of constraint violation is obtained by modifying Eq. (5.5):

$$p(\overline{C}|D_k) = p(\overline{C}|CF_k) \cdot p(CF_k|D_k) \tag{5.10}$$

where $p(CF_k|D_k)$ is the probability of the control flaw given the design option. When a design option eliminates a control flaw, $p(CF_k|D_k) = 0$.

In the case where one control flaw is addressed by a single design option, while the probabilities of the other control flaws are unchanged and no new control flaws are introduced, the updated probability of the constraint being violated is obtained by similarly modifying Eq (5.6):

$$p(\overline{C}|D_k) = 1 - (1 - p(\overline{C}|CF_k) \cdot p(CF_k|D_k)) \cdot \prod_{\substack{i = 1 \\ i \neq k}}^{n_{CF}} (1 - p(\overline{C}|CF_i) \cdot p(CF_i)) \tag{5.11}$$

In the general case where each control flaw is addressed by one or more design options, the updated probability for the constraint being violated is:

$$p(\overline{C}|\{D_i\}) = 1 - \prod_{i = 1}^{n_{CF}} (1 - p(\overline{C}|CF_i) \cdot p(CF_i|\{D_i\})) \tag{5.12}$$

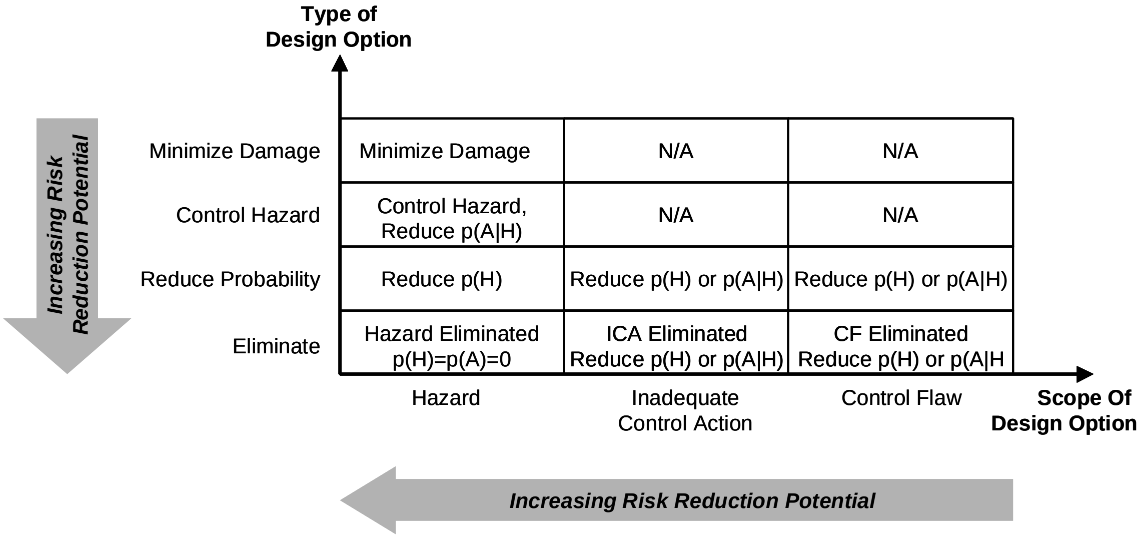
where $\{D_i\}$ is the set of design options for CF_i. For control flaws that are not addressed by design options the term $p(CF_i|\{D_i\})$ is set to $p(CF_i)$[1].

The most effective (from a risk reduction viewpoint) design options are those that have the largest effect on the probabilities of hazards and accidents. Therefore the effectiveness of a design option is mediated by the relationship between the control flaw being addressed and the hazard. In order to assess the effectiveness of design options, it is therefore necessary first to characterize this relationship, as described below.

Probabilities of Control Flaws given Design Options. The design option evaluations made in Step 2 (see Section 5.6) of the risk analysis can be used to inform the calculation of the conditional probabilities of the control flaws. The conditional probability is calculated by first noting the type and scope of the design option. As shown in Figure 5.14, the type and the scope of the design option determine the potential risk reduction achievable. Design options that eliminate or reduce the probability of hazards, ICAs or CFs, reduce the probability of the hazard. Design options that control the hazard can reduce the probability of the accident given the hazard and/or the resulting damage. Design options that result in the elimination of hazards have the highest risk reduction potential. In contrast, design options that minimize damage have the lowest risk reduction potential.

In cases where the design option reduces a probability (i.e., reduces $p(H)$ or $p(A|H)$), the effectiveness parameter can be used to determine the conditional probability. In cases where the design option controls a hazard or minimizes damage, the effectiveness parameter can be used to determine the reduced consequences and costs.

[1] Design options may have side-effects that increase or decrease the probabilities of existing control flaws or introduce additional control flaws. The incorporation of these effects is discussed in Addendum 3 at the end of this chapter.



	Hazard	Inadequate Control Action	Control Flaw
Minimize Damage	Minimize Damage	N/A	N/A
Control Hazard	Control Hazard, Reduce p(A|H)	N/A	N/A
Reduce Probability	Reduce p(H)	Reduce p(H) or p(A|H)	Reduce p(H) or p(A|H)
Eliminate	Hazard Eliminated p(H)=p(A)=0	ICA Eliminated Reduce p(H) or p(A|H)	CF Eliminated Reduce p(H) or p(A|H

Figure 5.14 Risk Reduction Potential According to Design Option Type and Scope

3. Hazard Occurrence

Hazards occur when the constraints intended to prevent the hazards are violated. In the simple case where there is only one constraint for the hazard, the probability of the hazard is given by:

$$p(H) = p(H|\overline{C}) \cdot p(\overline{C}) \qquad (5.13)$$

When the violation of the constraint guarantees that the hazard will occur, $p(H|\overline{C}) = 1$.

When there are n_C constraints under the hazard, the probability of the hazard is determined by first noting that the hazard will not occur if all of the constraints are effective:

$$p(\overline{H}) = \prod_{i=1}^{n_C} p(C_i) \qquad (5.14)$$

Therefore, in the special case where $p(H|\overline{C_i}) = 1$ (i.e., the violation of any one of the constraints guarantees that the hazard will occur), the probability of the hazard is:

$$p(H) = 1 - \prod_{i=1}^{n_C} (1 - p(\overline{C_i})) \tag{5.15}$$

In the general case where $p(H|\overline{C_i}) \neq 1$, that is, the violation of constraints does not guarantee that the hazard will occur, combining Eq. (5.13) and Eq. (5.15) gives:

$$p(H) = 1 - \prod_{i=1}^{n_C} (1 - p(H|\overline{C_i}) \cdot p(\overline{C_i})) \tag{5.16}$$

Relative Effectiveness of Constraints. The contributions of constraints to Eq. (5.16) defines their effectiveness in constraining the hazard. Constraints that are highly likely to be violated and whose occurrence makes it highly probable that the hazard will occur are the least effective. The calculation of the probability of a hazard given a constraint violation is domain specific (e.g., see space shuttle example in Section 5.8). Where quantitative probabilities are available, the terms $p(H|\overline{C_i}) \cdot p(\overline{C_i})$ can be used to rank the constraints.

Where only qualitative probabilities are available, a graphical approach as shown in Figure 5.15 can be used to compare constraints. When the probability of a constraint being violated is low, and the probability of a hazard given the constraint violation is also low, the constraint is defined to be highly effective (e.g., C_1 in the figure). Conversely, when the probability of a constraint being violated is high, and the probability of a hazard given the constraint violation is also high, the constraint is defined to be highly effective (e.g., C_9 in the figure). Identifying the constraints that have the largest impact on the probability of a hazard allows designers to focus their risk mitigation efforts on these constraints therefore develop the most effective risk mitigation strategies.

4. Transition to Accident

As discussed previously, the occurrence of a hazard does not guarantee that an accident will occur. Accidents occur when a hazardous state persists and coincides with the appro-

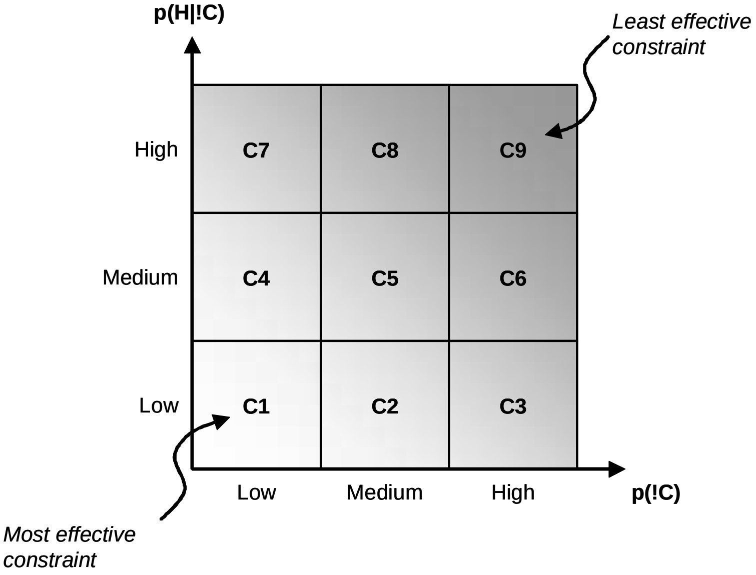


Figure 5.15 Relative Effectiveness of Constraints w.r.t. Hazards

priate environmental conditions. Therefore the overall probability of an accident related to a specific hazard is given by:

$$p(A) = p(A|H) \cdot p(H) \qquad (5.17)$$

where $p(A|H)$ is the probability of an accident occurring given the hazard. $p(A|H)$ is determined by context specific factors and should be determined by domain experts together with risk analysts. For example, the probability that two aircraft will collide if they have violated the minimum separation distance depends on several factors, such as visibility conditions, pilot experience, and whether the on-board collision alert system is functioning [Leveson, 1995].

Design options can be used to implement constraints to reduce $p(A|H)$. For example, in the gas leak case, the more quickly a gas leak is shut off the lower the probability that it will coincide in time or space with an ignition source and lead to a fire. An analysis of the control flaws and evaluation of the design option can be used to identify ways that the design option design could be ineffective, could be implemented incorrectly, or become ineffective over time. Eq. (5.6) can be used to determine the probability that constraints/ design options are violated.

This section has developed an analytical framework to determine the probabilities of hazards and accidents, by using control flaws as the basic element. A similar derivation can be made for consequences and costs of accidents. The next section discusses how this framework can be applied in risk management.

5.7.2 Application of the Analytical Framework

This section began by noting that the purpose of this step of the risk assessment is three-fold:

1. To aid in the selection of design options;
2. To estimate the residual risk for the selected design; and
3. To make the safety case, that is, show the extent to which hazards have been addressed in the selected design.

The analytical framework derived here can be used to aid in achieving these objectives as follows:

Selection of Design Options. The selection of design options is informed by the risk assessment as well as other factors such as cost and performance implications. The overall effectiveness of design options in reducing risk can be determined by propagating the adjusted probabilities from control flaws through constraints to hazards and accidents. Design options that eliminate hazards are preferred from a risk viewpoint. Next in the ranking are design options that reduce the probability of hazards, followed by design options that reduce the probability of accidents given hazards. When none of these design options is available, design options that minimize damage are the last resort. Of course, design options can also be used in combination, as illustrated in the examples.

Estimation of Residual Risk. The residual risk is estimated by applying the expressions developed here to the selected set of design options. The overall risk level for the entire system can be estimated based on the risk assessments for each hazard/accident. A quantitative estimate (probabilities and/or consequences) is only possible when the estimates for all the hazards are also quantitative. In other cases, a general estimate of the overall risk level can still be made. Graphical techniques may be useful in these cases.

While an overall risk estimate may be a useful public relations tool (when the estimate is low), it does not have much value as a decision aid. Accidents are caused by individual hazards and it is therefore more important to decision makers to know whether particular hazards have been sufficiently mitigated.

Making the Safety Case. The safety case is made by using the selected design options and the estimates of residual risk to show how hazards were addressed and that the risks associated with individual hazards have been brought to an acceptable level. In addition, the residual risk of designs that were not selected can be used to show that the approach selected does indeed address risks in the most effective way, given cost, performance, and other constraints.

As noted in the development of the analytical framework, it may not always be possible to assign quantitative values to risk probabilities or consequences. However, as shown in both the gas leak example and space shuttle example (Section 5.8), a detailed presentation of the sources of risk, the strategies to mitigate risk, and potential reasons for an increase in risk is provided by this approach. Such a detailed presentation may provide decision makers with an improved understanding of risk that leads to better decisions than single numbers, especially when the accuracy of these numbers is doubtful. Further work should address the topic of how best to present the results of the risk analysis to decision makers.

5.7.3 Organizational Risk Factors

This section discusses how the impact of organizational factors on risk can be qualitatively incorporated into the analysis. First, two approaches to incorporating organizational factors into probabilistic risk assessment are reviewed. Both approaches claim to allow the quantitative assessment of organizational factors. Next, the incorporation of organizational factors into the suggested risk assessment approach is discussed.

Approaches to Organizational Risk Factors in PRA

A number of other researchers have suggested approaches to accounting for organizational factors in risk assessment, primarily in the area of probabilistic risk assessment (PRA). These approaches include the SAM (System-Action-Management) framework [Murphy and Paté-Cornell, 1996], and WPAM (Work Process Analysis Model) [Davoudian et al., 1994a; Davoudian et al., 1994b], both of which build on the probabilistic risk/safety assessment (PRA/PSA) framework.

In the SAM framework management (or organizational) factors are viewed as affecting component failure probabilities through human decision and actions. The effect of management factors on component failure probabilities is modelled using different models of decision making and execution. The SAM approach claims to be quantitative because quantitative values are assigned to the model parameters. However, because little empirical or analytical data exists for the required parameter values, the example SAM applications rely on expert opinion. Such estimates run the risk of being qualitative guesses masquerading as quantitative values. Nevertheless, the core idea of SAM, that management factors affect the probabilities of component failures, is useful. A similar view is taken in this approach to risk assessment: here organizational factors are seen as affecting the probabilities of control flaws, but this influence is judged in a qualitative way only.

WPAM takes a similar approach to SAM. In this case, organizational factors are seen as introducing dependencies among probabilistic safety assessment parameters. Expert opinion is used to assign relative quantitative weights to the pertinent organizational factors, which are then incorporated into the PSA. WPAM explicitly takes a formal approach to ensuring the consistency of expert opinion, but the problems discussed above with quantitative expert opinion apply here too.

Organizational Risk Factors in the New Approach

While a quantitative assessment of the impact of organizational factors could be useful, the current state of knowledge does not allow believable quantification of these factors. It is preferable to use qualitative estimates that make the lack of quantitative information evident, than to use quantitative estimates that give a false impression of accuracy and precision. This approach to risk assessment focusses on identifying areas where organizational risk factors have a significant impact on risk.

The effect of organizational factors on risk is incorporated into the risk assessment by first determining the residual hazard and accident probabilities assuming ideal organizational conditions and then estimating the increase in probabilities due to organizational risk factors. Under ideal organizational conditions, organizational control flaws do not occur and design option effectiveness is limited only by technical factors.

Organizational risk factors are identified in two ways. First, control flaws may be explicitly organizational. For example, in the gas leak scenario several organizational control flaws were identified (e.g., CF1.3.1.1: Sensors to detect leaks were not installed). Organizational control flaws will increase the probability of hazards and accidents. Second, organizational factors may limit the effectiveness of design option implementation and detract from the continued effectiveness of design options. In the presence of organizational risk factors, the updated probabilities of the control flaws will therefore be larger than the residual values estimated under ideal organizational conditions.

As discussed above, in general it will not be possible to assign a quantitative value to the increase in risk due to organizational factors. The impact of organizational factors on risk should be qualitatively estimated by the relevant domain experts and risk analysts. An analysis of the organization, including its structure and safety culture, can qualitatively identify how likely control flaws are, and how likely it is that design options are ineffective. For example, an organization with a poor safety culture is likely to exhibit poor compliance to maintenance procedures. In such an organization design options that do not involve maintenance may therefore be preferable. In addition, system models such as those proposed by [Murphy and Paté-Cornell, 1996], and in particular system dynamics models, can aid in evaluating the relative impact of organizational factors. For example, Chapter 4 provides a set of organizational safety archetypes based on system dynamics that can be used to estimate the impact of certain specific organizational policies. Chapter 4 also illustrated how the often implicit trade-off between safety and performance can increase risk.

The gas leak and shuttle examples illustrate how organizational factors can be incorporated into risk assessments.

5.7.4 Gas Leak Example Continued: Residual Risk Assessment

Step 2 of the hazard analysis identified the following constraints (see Section 5.11):

- C1.1: Gas lines must not be placed in offices.
- C1.2: Gas lines and joints must be leak-resistant.
- C1.3: The gas supply must be shut off in the event of a gas leak.
- C1.4: The building must be evacuated in the event of a gas leak.
- C1.5: The building sprinkler system must be activated in the event of a fire.

The constraints indicate the possibility of different design architectures. In this case, each constraint can also be restated as a design option, as discussed in the next section. C1.1. suggests a design where the gas lines are not placed in offices (i.e., hazard is eliminated). The remaining constraints assume a design where the gas lines are placed in offices, but steps are taken to prevent leaks and minimize the damage if a leak does occur. If C1.1 is not implemented, or done so unsuccessfully, the remaining constraints are necessary.

The hazard can therefore be directly addressed by the following design options, obtained by rephrasing the high-level constraints identified in Step 2:

- D1.1: Remove gas line from offices/do not install gas lines in offices.
- D1.2: Use leak-resistant tubing and joints.

- D1.3: Periodically inspect gas lines and joints for damage and/or leaks.
- D1.4: Use leak sensors and shut down gas in event of gas leaks.
- D1.5: Detect gas leaks and evacuate building.
- D1.6: Detect fires and activate building sprinkler system.

These design options can be used to develop two main design scenarios. In the first scenario the hazard is eliminated by implementing D1.1. In the second scenario the hazard and accident probabilities are reduced, and the potential damage is minimized, by implementing design options D1.2 through D1.6.

Design Scenario One: Gas lines routed outside offices

This scenario uses D1.1 on its own. D1.1 eliminates the hazard, provided it is implemented correctly. The probability of the hazard given D1.1 and assuming ideal organizational conditions is zero:

$$p_{\text{ideal org}}(H1|D1.1) = 0 \tag{5.18}$$

There are several organizational control flaws that could lead to D1.1 not being implemented, identified under C1.1 (see Section 5.11):

- CF1.1.1.1: Requirement that gas lines not be placed in offices is not communicated to designers and placement of gas lines is not monitored by safety personnel.
- CF1.1.1.2: Designers disregard the requirement that gas lines not be placed in offices and placement of gas lines is not monitored by safety personnel.
- CF1.1.2.1: The building plans are unclear about the gas line restriction and the placement of gas lines is not monitored by safety personnel.
- CF1.1.2.2: The builders disregard the building plans and placement of gas lines is not monitored by safety personnel.

The probability, $p(\overline{D1.1})$, that the design option is not implemented is the probability that constraint C1.1 is violated, $p(\overline{C1.1})$, and is given by Eq. (5.6):

$$p(\overline{D1.1}) = p(\overline{C1.1}) = 1 - \prod_{i=1}^{4} (1 - p(\overline{C1.1} | CF_i) \cdot p(CF_i)) \qquad (5.19)$$

where the CF_i are the control flaws listed above. Because all the identified control flaws operate at the organizational level, it is difficult to assign a quantitative value for the probabilities of the control flaws. An analysis of the organization may indicate qualitatively how likely these control flaws are. In addition, the existence of control flaws indicates that the design option could be implemented incorrectly, and the specific control flaws identified point out the areas that should be monitored during design and development.

Determining whether D1.1 is implemented correctly can be easily done by inspection (easy observability). Once the design option is implemented correctly, it is likely to remain in place because rerouting gas lines is unlikely to occur unless the building is altered for some reason (high stability).

D1.1 eliminates the hazard and therefore provides an ideal solution from a risk viewpoint. However, other considerations may make this option impractical. A second scenario that reduces the probability of the hazard and the accident, and minimizes the damage associated with an accident is therefore considered, as illustrated in Figure 5.11.

Design Scenario Two: Gas lines routed inside offices

In this scenario, gas lines are routed through offices, and additional design options are selected to reduce the hazard and accident probabilities and minimize damage should an accident occur. The following design options are used:

- D0: Gas lines are routed through offices.
- D1.2: Use leak-resistant tubing and joints.
- D1.3: Periodically inspect gas lines and joints for damage and/or leaks.
- D1.4: Use leak sensors and shut down gas in event of gas leaks.
- D1.5: Detect gas leaks and evacuate building.
- D1.6: Detect fires and activate building sprinkler system.

The possible consequences of a gas leak have already been estimated in Step 1b. Here, the probability of a gas leak given that gas lines are installed in offices is updated to reflect the above design options.

- **D1.2: Use leak-resistant tubing and joints.**

D.1.2 directly reduces the probability of the hazard, provided it is implemented correctly. Under ideal organizational conditions the probability of the hazard given D1.2 is a function of the technical properties of the equipment used:

$$p_1(H1|D1.2) = f(\text{technical properties}) \tag{5.20}$$

There are several organizational control flaws that could lead to an ineffective initial implementation of D1.2, or to a decline in effectiveness over time (see Section 5.11):

- C1.2: Gas lines and joints must be leak-resistant.
- CF1.2.1.1: The requirement for leak-resistant gas lines and joints is not communicated to the designers and specification of gas lines and joints is not monitored by safety personnel.
- CF1.2.1.2: The designers disregard the requirement for leak resistant gas lines and joints and specification of gas lines and joints is not monitored by safety personnel.
- CF1.2.1.3: Designers erroneously select wrong type of gas lines and joints and specification of gas lines and joints is not monitored by safety personnel.
- CF1.2.2.1: The specification for the gas lines and joints is not properly communicated to the builders.
- CF1.2.2.2: The builders disregard the specification and installation is not monitored by safety personnel.
- CF1.2.2.3: Installation is not monitored by safety personnel and installation is not monitored by safety personnel.
- CF1.2.3.1: The builders do not have the proper installation skills and installation is not monitored by safety personnel.
- CF1.2.3.2: The builders disregard the installation rules and installation is not monitored by safety personnel.
- CF1.2.4.1: Defects are not detected.
- CF1.2.4.2: Gas lines and joints are installed despite detected defects.

- CF1.2.5.1: Deterioration is not detected or corrected.

The probability that the design option is improperly implemented is the probability that constraint C1.2 is violated, and is determined by applying Eq. (5.6):

$$p(\overline{D1.2}) = p(\overline{C1.2}) = 1 - \prod_{i=1}^{4}(1 - p(\overline{C1.2}|CF_i) \cdot p(CF_i)) \tag{5.21}$$

where the CF_i are the control flaws listed above. As with the first design scenario, all the identified control flaws are ‘soft’ in nature and operate at the organizational level. It is difficult to determine a quantitative value for the probability. The specific control flaws identified point out the areas that should be monitored during design (e.g., CF1.2.1.1), installation (e.g., CF1.2.2.1), and operation (e.g., CF1.2.5.1).

Determining whether D1.2 is implemented correctly can be done by simple inspections (observability). Continued vigilance is necessary to ensure that gas lines and joints are not replaced with lower quality items during periodic maintenance (stability).

D1.2 provides an initial line of defense against gas leaks. High-quality components and installation can significantly reduce the probability of a gas leak. Continued vigilance is necessary to ensure that deteriorating equipment is detected and replaced with the correct equipment. The next design option addresses the issue of detecting and correcting damage or leaks.

- **D1.3: Periodically inspect gas lines and joints for damage and/or leaks.**

D1.3 is the first design option in this example that is mainly organizational in nature. D1.3 reduces the probability of the hazard, provided inspections are performed correctly and over the lifetime of the building. In this case the probability of the hazard depends on the technical properties of the equipment, as well as the effectiveness of inspections in promptly detecting and repairing damage or leaks:

$$p(H1|D1.3) = f(\text{inspection quality and frequency, technical properties}) \tag{5.22}$$

The more frequent inspections are, the more likely they are to detect damage before it becomes serious. The required frequency depends on the technical properties of the equipment—some types of equipment may deteriorate more rapidly. Combining D1.2 and D1.3 therefore reduces the risk more than using either option on its own.

Eq. (5.22) can be used by designers to motivate either installing or not installing gas lines in offices. If the probability of a leak is deemed to be sufficiently low (as defined by stakeholders), and there are other reasons why gas lines should be routed through offices, a decision may be made to install gas lines in offices. On the other hand, if the probability is high, decision-makers have a rational motivation for routing gas lines differently (and possibly incurring cost and other penalties).

Eq. (5.22) assumes that the inspection quality is limited only by technical factors (e.g., what level of deterioration is detectable) and that the frequency of inspections is determined solely by the technical properties of the equipment (e.g, how rapidly is deterioration expected to occur). However, in the long term the quality and frequency of inspections may be limited by organizational factors, as indicated by the evaluation of D1.3 (see Section 5.6.3):

- Effectiveness: Inspections performed improperly.
- Stability: Inspections deteriorate in frequency and thoroughness.
- Observability: May be difficult to determine whether inspectors are shirking their duties

D1.3 may be effective in reducing the probability of leaks, but it should not be relied upon as the only hazard mitigation measure. D1.2 should therefore be used together with D1.3, as well as additional design options that address the consequences of a gas leak. The next design option controls the hazard if it does occur.

- **D1.4: Use leak sensors and shut down gas in event of gas leaks.**

D1.4 controls the hazard by limiting the extent of any gas leaks. This option reduces the probability of an accident given the hazard, and may also reduce the damage resulting from an accident. The accident probability is a function of how rapidly a gas leak is detected and shut down (i.e., limiting hazard exposure) and the probability of an ignition source being present:

$$p((A|H)|D1.4) = f\begin{pmatrix}\text{time to gas leak shut down,}\\ \text{probability of ignition source}\end{pmatrix} \tag{5.23}$$

The time to detection and gas shut-down depend on the technical properties of the system installed.

There are several control flaws that could lead to an ineffective initial implementation of D1.4, or to a decline in effectiveness over time. The control flaws indicated in bold result point to possible flaws in the design.

- C1.3: The gas supply must be shut off immediately in the event of a gas leak.
  CF1.3.1.1: Sensors to detect leaks were not installed.
  CF1.3.1.2: Defective sensors were installed.
  CF1.3.1.3: Sensors were incorrectly installed.
  CF1.3.1.4: Sensors stopped working and were not replaced.
  CF1.3.2.1: Sensors not connected to gas shut-off system.
  CF1.3.2.2: Sensor signal corrupted.
  CF1.3.2.3: Gas shut-off system incorrectly interpret sensor signals.
  CF1.3.2.4: Gas shut-off system does not issue shut-off signal.
  CF1.3.2.5: Shut-off signal corrupted.
  CF1.3.2.6: Shut-off signal incorrectly interpreted.
  CF1.3.2.7: Shut-off actuator fails.
  CF1.3.2.8: Shut-off signal sent too late.

The design should be inspected and analyzed to ensure that these control flaws are eliminated. For example, CF1.3.2.2 indicates the possibility of signal corruption. This control flaw can be eliminated by shielding wires against interference. Determining whether D1.4 is implemented correctly can be done by analyzing the design for the technical control flaws, and by inspecting the physical system (observability).

The remaining control flaws are organizational in nature. These control flaws indicate the ways in which D1.4 can be poorly implemented despite a good design, and how the effectiveness of D1.4 can decline over time (stability). The effectiveness of this option depends primarily on the technical design, but organizational control flaws may affect the quality of the design, installation, or maintenance, thereby increasing the probability of a leak above that given by Eq. (5.23).

The next two design options address the consequences of a gas leak and/or accident.

- **D1.5: Detect gas leaks and evacuate building.**
- **D1.6: Detect fires and activate building sprinkler system.**

D1.5 and D1.6 reduce the damage resulting from an accident. The reduction in damage depends on (1) how rapidly gas leaks and/or fires are detected, (2) how rapidly the appropriate alarms are sounded, (3) how effectively the building can be evacuated, and (4) how effective the sprinkler system is. As in the previous cases, the control flaws identified in Step 4 can be used to eliminate design flaws, inform the calculation of expected damage, and identify organizational factors that may limit the effectiveness of these options.

5.8 Partial Risk Analysis of the Space Shuttle

This section applies the proposed risk analysis approach to one aspect of a more complex system, the space shuttle. The focus is on the factors that led to the Space Shuttle Columbia accident. Note that this analysis is occurring in hindsight and may therefore be based on information that was not necessarily available at the time of the Columbia mission. The purpose of this example is to show the application of the proposed risk analysis technique to a more complex system. There is no intent to judge or criticize any of the people or organizations involved in the accident. For an extensive report on the accident, see [Gehman, 2003].

The hazard analysis underlying the risk analysis is presented in Addendum 2 at the end of this chapter.

5.8.1 Initial Risk Assessment

The Space Shuttle Columbia accident occurred when a hole in the leading edge of the wing allowed hot air to enter the internal wing structure during re-entry [Gehman, 2003]. The overheating of the internal wing structure led to structural failure of the wing and subsequently the disintegration of the shuttle. Consider now one approach to assessing the risk of a re-entry accident resulting from overheating of the internal wing structure and to developing strategies to reduce this risk.

In Step 1 of the risk analysis, initial estimates of the risks are formed to facilitate ranking of these risks and to determine whether the initial estimate of risk falls within an acceptable range to continue system development.

The hazard of interest is “overheating of the wing internal structure”. This hazard will occur if there is a hole in the wing exterior during re-entry. The associated risk is an accident involving structural failure during re-entry.

- H: Overheating of the wing internal structure.
- R: Disintegration of the shuttle during re-entry.

The probability of the hazard is a function of the ability of the wing structure to prevent hot gases from coming into contact with the internal structure of the wing during all phases of the shuttle flight:

$$p(H) = f(\text{ability of wing structure to prevent hot gases from entering}) \qquad (5.24)$$

Note that before the Columbia accident this probability was estimated to be low, because proper analyses and tests were not performed to determine the actual vulnerability of the structure, and because there was a tacit assumption that a foam impact could not damage the reinforced carbon-carbon of which the wing leading edge is comprised. The damage associated with the hazard is the loss of the space shuttle and crew. The shuttle is exposed to hot air during re-entry into the earth’s atmosphere. Thus, the duration of exposure to the hazard is the time taken for re-entry:

$$Exposure = \text{Re-entry duration} \tag{5.25}$$

Assuming that the structural problem (i.e., a hole in the wing) is not corrected, the probability of the hazard leading to an accident is near unity in this case, as the (current design of the) wing internal structure is not resistant to hot gases and structural failure of the wing is therefore inevitable:

$$p(A|H) \approx 1 \tag{5.26}$$

Finally, the immediate consequences of the accident are the loss of the shuttle and crew. Other consequences include the impact of grounding of the shuttle fleet on scientific research and construction of the International Space Station, the loss of reputation for NASA, and the loss of public confidence in the space program:

$$\mathbf{C}(\text{re-entry accident}) = \begin{bmatrix} \text{loss of shuttle and crew} \\ \text{political and research impact} \\ \text{shuttle fleet grounding} \end{bmatrix} \tag{5.27}$$

Based on this hazard, it is possible to identify a number of initial design options. For example, the hazard can be controlled by redesigning the internal wing structure to make it resistant to the hot gases to which the shuttle is exposed during re-entry (D1.1). The hazard damage can be minimized by equipping the shuttle with a crew escape pod to protect and evacuate the crew in the event of a shuttle disintegration (D1.2). Additional design options are developed in step 4 of the hazard analysis based on the inadequate control actions and control flaws.

5.8.2 Design Option Evaluation

In Step 2 of the risk analysis, different design options are evaluated to determine their impact on risk.

The following design options were identified to directly address the hazard and reduce the consequences of an accident, respectively (see Section 5.12):

- D1.1: Redesign the internal wing structure to make it resistant to the hot gases to which the shuttle is exposed during re-entry.
- D1.2: Equip the shuttle with a crew escape pod to protect and evacuate the crew in the event of a shuttle disintegration.

In addition, the following design options can be used to address the identified control flaws under ICA1.2.1:

- C1.2: The fuselage and wing structure must be examined for possible damage before re-entry.
- ICA1.2.1: The fuselage and wing structure is not examined before re-entry.
- CF1.2.1.1: The possibility of damage to the fuselage or wing structure is not considered.

  D1.2.1.1.1: Change assumptions about the risk associated with foam impacts (e.g., by showing employees a video of test demonstrating destructive impact).

  D1.2.1.1.2: Make a pre re-entry inspection of the external structure part of the formal re-entry preparation process.

  These two options are complementary. The first option addresses the underlying incorrect assumptions, while the second option forces consideration of possible damage.
- CF1.2.1.2: Engineering analyses incorrectly indicate that damage is negligible.

  D1.2.1.2.1: Encourage and enforce application of sound engineering practice, such as using models only within the range for which they are validated and calibrated.
- CF1.2.1.3: It is not possible to perform a pre-entry examination of the fuselage and wing structure.

  D1.2.1.3.1: Ensure that NASA has access to satellite resources to view the shuttle on orbit.

  D1.2.1.3.2: Ensure that astronauts are equipped and trained to perform a space walk to examine and repair the wing structure.

  D1.2.1.3.3: Ensure that on-board imaging devices are available to the astronauts and that they are trained to use these devices.

In a complete hazard and risk analysis design options would be developed to address all the identified control flaws.

The design option evaluations are presented in the next section on step 3 (residual risk assessment) for the sake of readability.

5.8.3 Residual Risk Assessment

In step 3 of the risk analysis, the residual risks that remain once particular design options are selected are assessed. Design options modify the risk profile formed in Step 1, by eliminating or reducing the probabilities of hazards and/or accidents, or by limiting the consequences of accidents. The residual risk assessment can be used to (1) estimate the residual risks for individual hazards given particular sets of design options; (2) estimate the overall risk of the system; and (3) make the safety case by presenting the approaches taken/not taken to mitigate hazards.

In a complete risk analysis, one or more design scenarios would be developed and analyzed. For the purpose of illustrating the approach, only the individual impacts on risk of selected design options are considered here.

D1.1: Redesign the internal wing structure to make it resistant to the hot gases to which the shuttle is exposed during re-entry.

TABLE 5.1 Design Option Characterization for D1.1

Design Option	D1.1: Redesign the internal wing structure to make it resistant to the hot gases to which the shuttle is exposed during re-entry.
Scope	Hazard.
Type	Control hazard.
Effectiveness	Depends on redesign. Consult structural engineers and others involved with wing structure (e.g., electrical engineers w.r.t. wiring in wings).
Stability	This design option should remain effective, assuming material comprising wing does not deteriorate (e.g., metal fatigue) over time. Update this parameter when more information on possible material deterioration available.
Observability	New structure can be analyzed and tested to ascertain heat resistance.

This design option controls the hazard and therefore reduces the probability of an accident from near certainty with the present structure. The probability of an accident given the hazard depends on the structural characteristics of the redesigned wing:

$$p(A|H) = f(\text{structural characteristics}) < 1 \quad (5.28)$$

The evaluation of this design option indicated that it can remain effective over time, barring structural deterioration (stability). The actual effectiveness of this option can be ascertained by means of analyses and testing (observability). While this option may significantly reduce the probability of an accident, it may have impacts on performance that make it infeasible, be exorbitantly expensive, or be impossible to manufacture and assemble.

The next two options are complementary. The first option addresses the underlying incorrect assumptions, while the second option forces consideration of possible damage.

D1.2.1.1.1: Change assumptions about the risk associated with foam impacts (e.g., by showing employees a video of test demonstrating destructive impact).

This design option was identified in the hazard analysis to address control flaw CF1.2.1.1 (the possibility of damage to the fuselage or wing structure is not considered), as shown below:

- C1.2: The fuselage and wing structure must be examined for possible damage before re-entry.
- ICA1.2.1: The fuselage and wing structure is not examined before re-entry.
- CF1.2.1.1: The possibility of damage to the fuselage or wing structure is not considered.
- D1.2.1.1.1: Change assumptions about the risk associated with foam impacts (e.g., by showing employees a video of test demonstrating destructive impact).

The design option evaluation is shown below.

This option addresses two aspects of the hazard. First, the probability of damage to the wing exterior leading to an accident is reduced because engineers and technicians are more likely to be on the look-out for foam strikes, and they are more likely to treat any foam strikes that do occur with extreme seriousness:

TABLE 5.2 Design Option Characterization for D1.2.1.1.1

Design Option	D1.2.1.1.1: Change assumptions about the risk associated with foam impacts.
Scope	Control Flaw. CF1.2.1.1
Type	Reduce probability.
Effectiveness	Physically demonstrating the possible foam impact damage to the wing structure to employees can help change employee beliefs about foam impact. If employees at all levels, especially decision makers, believe that foam impacts are a serious problem, they are more likely to consider such impacts and ensure that the necessary damage is carried out. This option therefore has the potential to be highly effective.
Stability	Effectiveness can be expected to decline over time as employees forget about the possibility of impact damage and become complacent. It may therefore be necessary to periodically repeat training in order to maintain awareness of the problem.
Observability	Observing underlying beliefs and assumptions is difficult. Therefore D1.2.1.1.1 has low observability and should not be solely relied upon to ensure that damage inspections occur.

$$p(CF1.2.1.1|D1.2.1.1.1) \ = \ \text{very low} \tag{5.29}$$

where the probability that the possibility of damage is considered depends on how well assumptions are changed.

If foam strikes are identified and properly analyzed before re-entry, strategies to avoid a re-entry accident can be developed. For example, the crew might be transferred to the International Space Station, or another shuttle could be launched to rescue the crew or repair the foam impact damage [Gehman, 2003]. Therefore the probability of damage to the wing exterior leading to a re-entry accident is immediately reduced from the near certainty in the original design:

$$p((A|H)|D1.2.1.1.1) < p(A|H) \approx 1 \tag{5.30}$$

Second, this option also has a longer term effect: Greater awareness about the potential damage caused by foam strikes may spur advances in foam and its application or encourage designers to consider strengthening the wing exterior and interior structures (see D1.1). Therefore the probability of damage to the wing exterior may also be reduced in the long term:

$$p(H|D1.2.1.1.1) = p(\text{damage to NEW wing covering}) < p(H) \tag{5.31}$$

While this design option can significantly reduce the probability that damage is not detected, there are two factors that can limit its immediate and long-term effectiveness. The evaluation of this design option indicated that the effectiveness may decline over time unless active efforts are made to continue reinforcing the possibility and severe consequences of foam impact damage (stability). In addition, it is difficult to determine whether underlying beliefs and assumptions have been changed in the desired way (observability). Note also that even when assumptions are corrected, schedule and other pressures may make it difficult for employees to properly perform the tasks necessary to prove that damage did not occur.

D1.2.1.1.2: Make a pre re-entry inspection of the external structure part of the formal re-entry preparation process.

This design option was identified in the hazard analysis to address control flaw CF1.2.1.1 (the possibility of damage to the fuselage or wing structure is not considered.), as shown below:

- C1.2: The fuselage and wing structure must be examined for possible damage before re-entry.
- ICA1.2.1: The fuselage and wing structures are not examined before re-entry.
- CF1.2.1.1: The possibility of damage to the fuselage or wing structure is not considered.
- D1.2.1.1.2: Make a pre re-entry inspection of the external structure part of the formal re-entry preparation process.

The design option evaluation is shown below.

The probability that damage to the fuselage or wing structure is not considered can be reduced to near zero by making the inspection part of the formal process:

$$p_{ideal}(CF1.2.1.1|D1.2.1.1.2) \approx 0 \tag{5.32}$$

Like D1.2.1.1.1, this option reduces the probability of damage to the wing exterior leading to an accident because damage is more likely to be identified and can therefore be

TABLE 5.3 Design Option Characterization for D1.2.1.1.2

Design Option	**D1.2.1.1.2: Make a pre re-entry inspection of the external structure part of the formal re-entry preparation process.**
Scope	Control Flaw. CF1.2.1.1
Type	Reduce probability.
Effectiveness	Formal process makes it difficult not to perform the inspection. But effectiveness depends on availability of inspection resources (satellites, exterior cameras, etc.) and skill of inspectors (ground crew or astronauts).
Stability	Dependent on the ability of the supporting organizational structure to enforce and monitor inspections.
Observability	Medium. May be difficult to determine if inspections are not properly performed.

addressed. While identifying damage does not mean that it will be addressed, damage that is not identified is guaranteed not to be addressed. Therefore the probability of damage to the wing exterior leading to an accident is reduced from the near certainty in the original design:

$$p_{ideal}((A|H)|D1.2.1.1.2) < p(A|H) \approx 1 \tag{5.33}$$

$p((A|H)|D1.2.1.1.2)$ can be estimated more precisely based on analyses of the ability of the chosen inspection procedure to identify damage to the wings. For example, the next design option uses satellites to image the shuttle exterior.

The evaluation of this design option indicated that the effectiveness depends on the ability of the supporting organizational structure to enforce and monitor inspections over the lifetime of the shuttle (stability). It may become tempting to omit this inspection, either formally or informally, if a long period of time passes without any damage occurring. Even formal steps may be performed improperly and it may be difficult to determine whether inspections are properly performed (observability). These risk factors increase the probability of an accident given the hazard above the ideal probability suggested above.

D1.2.1.3.1: Ensure that NASA has pre-planned access to satellite resources to view the shuttle on orbit.

This design option was suggested to address a control flaw, as indicated below:

- C1.2: The fuselage and wing structure must be examined for possible damage before re-entry.
- ICA1.2.1: The fuselage and wing structures are not examined before re-entry.
- CF1.2.1.3: Resources are not available to perform a pre-entry examination of the fuselage and wing structure.
- D1.2.1.3.1: Ensure that NASA has pre-planned access to satellite resources to view the shuttle on orbit.

The design option evaluation is shown below.

TABLE 5.4 Design Option Evaluation for D1.2.3.1

Design Option	**D1.2.1.3.1: Ensure that NASA has access to satellite resources to view the shuttle on orbit.**
Scope	Control Flaw. CF1.2.1.3
Type	Reduce probability. This option reduces the probability of damage to the wing exterior leading to an accident because damage is more likely to be identified and can therefore be corrected.
Effectiveness	Depends on satellite imaging resolution and coverage of shuttle exterior surface.
Stability	High, as long as these resources are used on every mission. Not using the resources makes it more likely that they will be difficult to obtain in the future.
Observability	Absence of satellite resources is highly visible.

The probability that pre-entry examination cannot be performed can be reduced dramatically by ensuring that satellite resources are reserved for imaging purposes:

$$p(CF1.2.1.1|D1.2.1.3.1) = p(\text{reserved satellite resources available}) \qquad (5.34)$$

The requirement that satellite resources be available may not always be met. For example, when there is strong pressure to launch the shuttle, but satellite resources are not available for some reason, the temptation to ignore the requirement and launch the shuttle may be strong.

Like the previous two design options, this option reduces the probability of damage to the wing exterior leading to an accident because damage is more likely to be identified and can therefore be addressed. Therefore the probability of damage to the wing exterior leading to an accident is reduced from the near certainty in the original design:

$$p_{ideal}((A|H)|D1.2.1.1.2) < p(A|H) \approx 1 \tag{5.35}$$

The probability that damage is detected by on-orbit imaging can be estimated based on analyses of satellite imaging resolution and coverage (i.e., how much of the shuttle exterior can be imaged).

The availability of satellite resources does not guarantee that they will be used. Therefore this design option should be used together with the previous option that mandates pre reentry inspections of the shuttle exterior.

5.9 Current Limitations of the Risk Analysis Approach

Chapter 2 presented a set of criteria that a risk analysis methodology should meet. Consider now whether the proposed technique has the potential to address these criteria.

The approach has the potential to meet all the risk analysis criteria, but more extensive examples based on real systems are needed to evaluate the approach properly. In its present form, it has the following limitations:

- It does not explicitly address the third criterion, “take psychological and social attributes of risk into account”. This limitation is not an inherent feature of the approach. Chapter 1 discussed several possible additional risk attributes (see Table 1.1). Considering these additional risk attributes is important when setting acceptable risk limits and when discussing risks with various stakeholders. Decision makers and other stakeholders should determine which additional measures should be used in the pre-analysis phase of risk management. The determination of these additional factors is left as a topic for future work. Future work should consider how representations of risk can incorporate these attributes to foster an improved understanding of risk
- Like other approaches to risk assessment, the accuracy of the assessment depends on the accurate assessment of the probabilities of basic elements, in this case, control flaws. The approach shows how individual control flaw

probabilities can be combined, but does not indicate how individual control flaw probabilities should be determined.

- The approach provides no guidance on the identification of hazards.
- Dependencies between control flaws have not been addressed.

These limitations can be addressed in future research. This approach to risk analysis has the potential to overcome many of the limitations of existing approaches.

5.10 Summary

This chapter has developed and advocated the concept of continuous participative risk management (CPRM). CPRM is continuous: it addresses risk throughout the lifetime of a system, thus ensuring that risk is maintained at an acceptable level throughout the system’s lifetime. In addition, CPRM is participative: it draws decision makers, members from all levels of the organization, as well as other stakeholders into the risk analysis process. Including all stakeholders in the process increases the quality of the risk analysis, nurtures understanding of risk across the organization, and encourages commitment to risk management strategies.

One crucial aspect of risk management is risk analysis. This chapter also introduced a new approach to risk analysis, which builds on hazard analysis and uses concepts from the STAMP model of accidents developed by [Leveson, 2004a]. The proposed approach includes a set of parameters by which different options to address risks can be evaluated. The proposed approach addresses some of the limitations of existing risk analysis techniques, and both encourages and enables the continuous risk management approach advocated here.

Finally, the organizational aspects of risk in the proposed approach were discussed in detail. Guidelines were presented for (1) identifying organizational contributors to risk and (2) evaluating options to address risks from an organizational perspective. An example based on the Space Shuttle was presented to illustrate the risk analysis approach and in particular the incorporation of organizational factors.

5.11 Addendum 1: Gas Leak Hazard Analysis

This section describes a hazard analysis for a hypothetical gas leak in an office building. The results of these steps are used to help illustrate the new risk analysis approach. They are presented here for completeness.

1: Identify High-Level Hazards

An office building is being designed, and one concern is that gas leaks in offices could be set off by ignition sources such as matches or cigarette lighters. For the sake of this example, it is assumed that ignition sources are beyond the control of the building developers. In this case, the hazard is:

- H1: Gas leak in office.

2: Identify Safety-Related Requirements and Constraints

The following safety constraints can immediately be identified based on the basic system design:

- C1.1: Gas lines may not be placed in offices.
- C1.2: Gas lines and joints must be leak-resistant.
- C1.3: The gas supply must be shut off in the event of a gas leak.
- C1.4: The building must be evacuated in the event of a gas leak.
- C1.5: The building sprinkler system must be activated in the event of a fire.
- C1.6: Other [*Placeholder*]

C1.1 (most preferred) eliminates the hazard, C1.2 reduces the probability of the hazard, C1.3 controls the hazard, and C1.4 and C1.5 (least preferred) minimize the damaged associated with the hazard. Step 5 discusses the different approaches to hazard elimination and mitigation in more detail.

The final constraint, ‘Other’, is used as a placeholder to indicate that other constraints may be possible and/or necessary. Explicitly noting this possibility encourages analysts to consider it during initial iterations of the risk analysis during system development and during updates of the analysis in later phases of the system lifecycle.

3: Identify Possible Inadequate Control Actions

Continuing the gas leak example, the following inadequate control actions can be identified for the constraints associated with the hazard, “gas leak in office”.

- C1.1: Gas lines may not be placed in offices.

  ICA1.1.1: The building design erroneously places gas lines in one or more offices.

  ICA1.1.2: The builders erroneously install gas lines in one or more offices.

- C1.2: Gas lines and joints must be leak-resistant.

  ICA1.2.1: The building design specifies the wrong kind of gas lines and joints.

  ICA1.2.2: The builders install the wrong gas lines and joints.

  ICA1.2.3: The builders incorrectly install the gas lines and joints.

  ICA1.2.4: Defective gas lines and joints are installed.

  ICA1.2.5: Deterioration of the gas lines and joints over time is not noted or corrected.

- C1.3: The gas supply must be shut off in the event of a gas leak.

  ICA1.3.1: The gas leak sensors are defective and the leak is not detected.

  ICA1.3.2: The gas leak is detected, but gas is not shut off.

  C1.4: The building must be evacuated in the event of a gas leak.

  ICA1.4.1: The gas leak is not detected.

  ICA1.4.2: The gas leak is detected, but the evacuation alarm is not sounded.

  ICA1.4.3: Evacuation is impeded by locked doors.

- C1.5: The building sprinkler system must be activated in the event of a fire.

  ICA1.5.1: The fire is not detected.

  ICA1.5.2: The fire is detected, but the sprinkler system is not activated.

  ICA1.5.3: The sprinkler system is activated, but is not effective.

- C1.6: Other [Placeholder]

The next step identifies control flaws and design options.

4: Identify Control Flaws and Design Options

For example, the following control flaws can be identified for the gas leak example:

- C1.1: Gas lines may not be placed in offices.

  ICA1.1.1: The building design erroneously places gas lines in one or more offices and this error is not detected or corrected.

  CF1.1.1.1: The requirement that gas lines not be placed in offices is not communicated to the designers and the placement of gas lines is not monitored by safety personnel.

  CF1.1.1.2: The designers disregard the requirement that gas lines not be placed in offices and the placement of gas lines is not monitored by safety personnel.

  CF1.1.1.3: Other

  ICA1.1.2: The builders erroneously install gas lines in one or more offices and this error is not detected or corrected.

  CF1.1.2.1: The building plans are unclear about the gas line restriction and the placement of gas lines is not monitored by safety personnel.

  CF1.1.2.2: The builders disregard the building plans and placement of gas lines is not monitored by safety personnel.

  CF1.1.2.3: Other.

- C1.2: Gas lines and joints must be leak-resistant.

  ICA1.2.1: The building design specifies the wrong kind of gas lines and joints.

  CF1.2.1.1: The requirement for leak-resistant gas lines and joints is not communicated to the designers.

  CF1.2.1.2: The designers disregard the requirement for leak resistant gas lines and joints.

  CF1.2.1.3: The designers erroneously select the wrong type of gas lines and joints.

  CF1.2.1.4: The specification of gas lines and joints is not monitored by safety personnel.

  CF1.2.1.5: Other.

  ICA1.2.2: The builders install the wrong gas lines and joints.

  CF1.2.2.1: The specification for the gas lines and joints is not properly communicated to the builders and installation of gas lines and joints is not monitored by safety personnel.

  CF1.2.2.2: The builders disregard the specification and installation of gas lines and joints is not monitored by safety personnel.

CF1.2.2.3: Other.

ICA1.2.3: The builders incorrectly install the gas lines and joints.

CF1.2.3.1: The builders do not have the proper installation skills and installation of gas lines and joints is not monitored by safety personnel.

CF1.2.3.2: The builders disregard the installation rules and installation of gas lines and joints is not monitored by safety personnel.

CF1.2.3.3: Other.

ICA1.2.4: Defective gas lines and joints are installed and installation of gas lines and joints is not monitored by safety personnel.

CF1.2.4.1: Defects are not detected.

CF1.2.4.2: Gas lines and joints are installed despite detected defects.

CF1.2.4.3: Other

ICA1.2.5: Deterioration of the gas lines and joints deteriorate over time is not noted or corrected.

CF1.2.5.1: Deterioration is not detected.

CF1.2.5.2: Deterioration is detected but is not corrected.

CF1.2.5.3: Other.

- C1.3: The gas supply must be shut off in the event of a gas leak.

  ICA1.3.1: The gas leak is not detected.

  CF1.3.1.1: Sensors to detect leaks were not installed.

  CF1.3.1.2: Defective sensors were installed.

  CF1.3.1.3: Sensors were incorrectly installed.

  CF1.3.1.4: Sensors stopped working and were not replaced.

  CF1.3.1.5: Other.

  ICA1.3.2: The gas leak is detected, but gas is not shut off.

  CF1.3.2.1: Sensors not connected to gas shut-off system.

  CF1.3.2.2: Sensor signal corrupted.

  CF1.3.2.3: Gas shut-off system incorrectly interpret sensor signals.

  CF1.3.2.4: Gas shut-off system does not issue shut-off signal.

  CF1.3.2.5: Shut-off signal corrupted.

  CF1.3.2.6: Shut-off signal incorrectly interpreted.

  CF1.3.2.7: Shut-off actuator fails.

  CF1.3.2.8: Shut-off signal sent too late.

  CF1.3.2.9: Other.

- C1.4: The building must be evacuated in the event of a gas leak.

ICA1.4.1: The gas leak is not detected.

CF1.1.1.1: See CF1.3.1.1-4

CF1.1.1.2: Other.

ICA1.4.2: The gas leak is detected, but the evacuation alarm is not sounded.

CF1.4.2.1: Sensors not connected to evacuation alarm system.

CF1.4.2.2: Sensor signal corrupted.

CF1.4.2.3: Evacuation alarm system incorrectly interpret sensor signals.

CF1.4.2.4: Evacuation alarm system does not issue shut-off signal.

CF1.4.2.5: Signal to sound alarm corrupted.

CF1.4.2.6: Alarm signal incorrectly interpreted.

CF1.4.2.7: Alarm actuator fails.

CF1.4.2.8: Alarm sounded too late.

CF1.4.2.9: Other.

ICA1.4.3: Evacuation is impeded by locked doors etc.

CF1.4.3.1: Building violates fire safety codes.

CF1.4.3.2: Other.

- C1.5: The building sprinkler system must be activated in the event of a fire.

ICA1.5.1: The fire is not detected.

CF1.5.1.1: Sensors to detect fires were not installed.

CF1.5.1.2: Defective sensors were installed.

CF1.5.1.3: Sensors were incorrectly installed.

CF1.5.1.4: Sensors stopped working and were not replaced.

CF1.5.1.5: Other.

ICA1.5.2: The fire is detected, but the sprinkler system is not activated.

CF1.5.2.1: Sensors not connected to sprinkler system.

CF1.5.2.2: Sensor signal corrupted.

CF1.5.2.3: Sprinkler system incorrectly interpret sensor signals.

CF1.5.2.4: Sprinkler system does not issue shut-off signal.

CF1.5.2.5: Sprinkler signal corrupted.

CF1.5.2.6: Sprinkler signal incorrectly interpreted.

CF1.5.2.7: Sprinkler actuator fails.

CF1.5.2.8: Sprinklers activated too late.

CF1.5.2.9: Other.

ICA1.5.3: The sprinkler system is activated, but is not effective.

CF1.5.3.1: No water available.
CF1.5.3.2: Sprinklers blocked.
CF1.5.3.3: Other.

- C1.6: Other [*Placeholder*]

For the gas leak example, the hazard can be directly addressed by the following design options, which implement the high-level constraints identified in Step 2:

- D1.1: Remove gas line from offices/do not install gas lines in offices.
- D1.2: Use leak-resistant tubing and joints.
- D1.3: Periodically inspect gas lines and joints for damage and/or leaks.
- D1.4: Use leak sensors and shut down gas in event of gas leaks.
- D1.5: Detect gas leaks and evacuate building.
- D1.6: Detect fires and activate building sprinkler system.
- D1.7: Other.

In a complete hazard or risk analysis, design options should also be developed to address the inadequate control actions and control flaws identified in steps three and four.

The inadequate control actions and control flaws developed above are used to (1) update the initial risk estimates as shown in Step 4b; (2) aid in evaluating the design options as shown in Step 5; and (3) drive the development of further design options in subsequent iterations. For example, if D1.1 is selected, the inadequate control actions and control flaws associated with C1.1 should be considered when developing further design options. Thus CF1.1.1.3 (the designers disregard the requirement that gas lines not be placed in offices) gives rise to the following design option:

- D1.1.1.3.1: Installation of gas lines must be monitored and inspected by safety personnel.

5.12 Addendum 2: Partial Hazard Analysis of Space Shuttle

This section presents a partial hazard analysis of the space shuttle, which is used as a basis for the risk analysis presented earlier in this chapter.

Step 1: Identify Hazards and Risks

The Space Shuttle Columbia accident occurred when a hole in the leading edge of the wing allowed hot air to enter the internal wing structure during re-entry [Gehman, 2003]. The overheating of the internal wing structure led to structural failure of the wing and subsequently the disintegration of the shuttle. Here the hazard of interest is "overheating of the wing internal structure". This hazard can occur if there is a hole in the wing exterior during re-entry. The associated risk is an accident involving structural failure during re-entry.

- H1: Overheating of the wing internal structure.
- R1: Disintegration of the shuttle during re-entry.

The following design options can be used to directly address the hazard:

- D1.1: Redesign the internal wing structure to make it resistant to the hot gases to which the shuttle is exposed during re-entry.
- D1.2: Equip the shuttle with a crew escape pod to protect and evacuate the crew in the event of a shuttle disintegration.

Step 2: Identify Constraints

The following safety constraints can immediately be identified based on a high-level understanding of the space shuttle design:

- H1: Overheating of the wing internal structure.
- C1.1: The wing structure must not be damaged by pressures of up to TBD MPa.
- C1.2: The wing structure must be examined for possible damage before re-entry.
- C1.3: Re-entry with a damaged wing structure is not permitted.
- C1.4: Damage to the wing structure must be repaired before re-entry.
- C1.5: Other [Placeholder]

Step 3: Identify Possible Inadequate Control Actions

The following inadequate control actions can be identified for the constraints associated with the hazard, “overheating of the wing internal structure”.

- C1.1: The wing structure must not be damaged by pressures of up to TBD MPa.

  ICA1.1.1: The wing structure is exposed to larger than expected impacts.

  ICA1.1.2: The wing structure does not withstand within-specification impacts.

  ICA1.1.3: The wing structure fails for some other reason. [This is a ‘place-holder’ inadequate control action used to ensure that additional failure modes are considered as the design progresses.]

- C1.2: The wing structure must be examined for possible damage before re-entry.

  ICA1.2.1: The wing structure is not examined before re-entry.

  ICA1.2.2: The examination misses damage to the wing structure.

  ICA1.2.3: Other.

- C1.3: Re-entry with a damaged wing structure is not permitted.

  ICA1.3.1: Damage to the wing structure is not noted. Addressed by C1.2.

  ICA1.3.2: Wing structure damage is noted but not communicated to the relevant organizational members.

  ICA1.3.3: The damage is communicated to the appropriate people but a decision is made not to respond.

  ICA1.3.4: Other.

- C1.4: Damage to the wing structure must be repaired before re-entry.

  ICA1.4.1: Damage to the wing structure is not noted. Addressed by C1.2.

  ICA1.4.2: The wing structure cannot be repaired for some reason. For example, material to repair the damage is not available on the shuttle.

  ICA1.4.3: The repair is not done properly.

  ICA1.4.4: Other.

In some cases inadequate control actions may already be addressed by existing constraints, as shown above for ICA1.3.1 and ICA1.4.1.

Step 4: Identify Possible Control Flaws and Design Options

This step identifies the control flaws that could lead to the inadequate control actions. For example, consider the control flaws underlying ICA1.2.1:

- C1.2: The fuselage and wing structure must be examined for possible damage before re-entry.
- ICA1.2.1: The fuselage and wing structure is not examined before re-entry.

  CF1.2.1.1: The possibility of damage to the fuselage or wing structure is not considered.

  CF1.2.1.2: Engineering analyses incorrectly indicate that damage is negligible.

  CF1.2.1.3: It is not possible to perform a pre-entry examination of the fuselage or wing structure.

  CF1.2.1.4: Decision makers incorrectly interpret ambiguous data.

  CF1.2.1.5: Other.

The following design options can be used to address the identified control flaws:

- CF1.2.1.1: The possibility of damage to the fuselage or wing structure is not considered.

  D1.2.1.1.1: Change assumptions about the risk associated with foam impacts (e.g., by showing employees a video of test demonstrating destructive impact).

  D1.2.1.1.2: Make a pre re-entry inspection of the external structure part of the formal re-entry preparation process.

  These two options are complementary. The first option addresses the underlying incorrect assumptions, while the second option forces consideration of possible damage.

  D1.2.1.1.3: Other.
- CF1.2.1.2: Engineering analyses incorrectly indicate that damage is negligible.

  D1.2.1.2.1: Encourage and enforce application of sound engineering practice, such as using models only within the range for which they are validated and calibrated.

  D1.2.1.2.2: Other.
- CF1.2.1.3: It is not possible to perform a pre-entry examination of the fuselage and wing structure.

D1.2.1.3.1: Ensure that NASA has access to satellite resources to view the shuttle on orbit.

D1.2.1.3.2: Ensure that astronauts are equipped and trained to perform a space walk to examine and repair the wing structure.

D1.2.1.3.3: Ensure that on-board imaging devices are available to the astronauts and that they are trained to use these devices.

D1.2.1.3.4: Other.

5.13 Addendum 3: Incorporation of Design Option Side-Effects in Residual Risk Assessment

Design options may have side-effects that increase or decrease the probabilities of existing control flaws or introduce additional control flaws. These effects can be incorporated as shown below for the single design option case:

$$p(\bar{C}|D_k) = 1 - \left[\begin{array}{r} (1 - p(\bar{C}|CF_k) \cdot p(CF_k|D_k)) \times \prod\limits_{\substack{i=1 \\ i \neq k}}^{n_{CF}} (1 - p(\bar{C}|CF_i) \cdot p(CF_i|D_k)) \times \\ \times \prod\limits_{\substack{i = n_{CF}+1 \\ i \neq k}}^{n_{CF} + n_{\text{new CFs}}} (1 - p(\bar{C}|CF_i) \cdot p(CF_i|D_k)) \end{array}\right] \tag{5.36}$$

where $n_{\text{new CFs}}$ is the number of new control flaws introduced by the design option. Similar calculations can be used to account for side-effects in the other probabilities.

Chapter 6

CONCLUSIONS AND THOUGHTS ON FUTURE WORK

Research should have not only results, but also pointers toward the incomplete; who should know better than the author the limits of the work?

Kenneth J. Arrow, I Know a Hawk from a Handsaw

This is our true state; this is what makes us incapable of certain knowledge and of absolute ignorance. We sail within a vast sphere, ever drifting in uncertainty, driven from end to end. When we think to attach ourselves to any point and to fasten to it, it wavers and leaves us; and if we follow it, it eludes our grasp, slips past us, and vanishes for ever. Nothing stays for us. This is our natural condition and yet most contrary to our inclination; we burn with desire to find solid ground and an ultimate sure foundation whereon to build a tower reaching to the Infinite. But our whole groundwork cracks, and the earth opens to abysses.

Blaise Pascal, Pensées, 1660

Complex socio-technical systems pose a challenge to risk managers. The complexity of these systems makes it difficult to identify and assess risk. When risks are not identified or are underestimated disaster can ensue. In particular, recent accidents have demonstrated that an understanding of the way organizational behaviour contributes to risk is critical to effectively managing risk.

Existing risk analysis methodologies do not provide a good understanding of the risks associated with complex socio-technical systems. These methodologies are appropriate for simple systems where ‘mechanical’ failures prevail. But they are fundamentally limited when it comes to complex socio-technical systems because they are all event-based and do not adequately capture emergent behaviour.

This thesis has developed an alternative approach to risk analysis that accounts for the characteristics of modern socio-technical systems. The proposed approach moves beyond event-based models to include risks that do not depend only on component or subsystem failures, and incorporate human, organizational, and societal factors. By taking an explicit lifecycle view of systems, the approach enables (1) the early identification of risks and risk mitigation strategies, (2) aids in the allocation of resources to best manage risk, and (3) provides for the continuous monitoring of risk throughout the system lifecycle. In addition, the approach emphasizes and enables the participation of members at all levels of the organization as well as other stakeholders in order to best identify, assess, and manage risks. The proposed approach addresses technical, human, organizational, and other factors.

6.1 Thesis Summary

Chapter 1 introduced the concept of risk and the nature of accidents in modern complex socio-technical systems.

Chapter 2 looked at risk management as a decision-making process and developed a set of criteria for what makes a good risk analysis. Based on these criteria, current risk analysis techniques were evaluated. In particular, an extensive evaluation of probabilistic risk assessment revealed significant limitations.

The importance of organizational factors in accidents has gained increasing acceptance. Numerous authors have written on the topic. In particular, sociologists have developed several theories to explain why accidents occur and how they can be prevented. Chapter 3 reviewed three of the most popular sociological and organizational approaches to safety: normal accidents theory (NAT), high reliability organizations (HRO), and normalization of deviance. While these approaches do provide some useful insights, the chapter showed that they all have significant limitations.

Chapter 4 discussed the dynamic aspects of risk in organizations. First, a framework was developed to analyze the strategic trade-off between short and long term goals and understand why organizations tend to migrate to states of increasing risk. The apparent conflict between performance and safety was shown to result from the different time horizons applying to performance and safety. Performance is measured in the short term, while safety is indirectly observed over the long term. A short-term view creates the impression that safety and performance necessarily conflict. Expanding the time horizon attenuates the tension. By increasing awareness of the often implicit trade-offs between safety and performance organizations can avoid decisions that increase risk.

Next, a set of archetypes of organizational safety was presented. Accidents in diverse industries, while unique in their technical aspects, often exhibit common patterns of organizational behaviour. Chapter 4 identified several such patterns, or archetypes, and demonstrated their application in diverse industries. In addition, NASA specific archetypes were developed based on investigations into the Challenger and Columbia accidents.

Chapter 5 developed a new approach to risk analysis that is applicable to modern, complex socio-technical systems. First, the concept of continuous participative risk management was introduced. In this approach to risk management, risks are addressed throughout the lifetime of a system, and members from all levels of the organization are involved both in risk analysis and in risk mitigation.

Next, a new approach to risk analysis, which builds on hazard analysis and uses concepts from the STAMP model of accidents developed by [Leveson, 2004a] was developed. The proposed approach moves beyond event-based models to include risks that do not depend only on component or subsystem failures, and incorporate human, organizational, and societal factors. By taking an explicit lifecycle view of systems, the approach enables the early identification of risk mitigation strategies, aids in the allocation of resources to best manage risk, and provides for the continuous monitoring of risk throughout the system lifecycle. The proposed approach addresses some of the limitations of existing techniques, and both encourages and enables the continuous risk management approach.

6.2 Future Work

The limitations of event-based approaches to understanding accidents and risk and the importance of organizational factors are gaining increasing importance. However, while the problems are now well-recognized by many scholars, there is still little available on how to address these problems and prevent accidents. The risk analysis approach presented in this thesis is but one small step towards safer systems.

The following areas provide fertile ground for future work.

6.2.1 Representation of Risk Analysis Results

Throughout this thesis the importance of employee input and commitment has been emphasized. The importance of management commitment and understanding should not be underestimated either. Both the organizational safety archetypes and the risk analysis approach suggested here are intended to improve understanding of risk and how it arises. It would be useful to develop graphical techniques, and possibly animations, to illustrate the risk analysis results and archetypes. For example, One way of improving understanding at all levels of an organization is to run simulations of problem behaviours [Senge, 1990; Sterman, 2002a]. It would therefore be useful to develop executable models of the archetypes using the system dynamics modelling language.

6.2.2 Stakeholder Involvement in Risk Analysis

Stakeholder involvement is a crucial aspect of building and operating complex systems. For example, community stakeholders can determine whether or not nuclear power plant operators gain local and federal government approval to site plants in particular areas. By involving and including community stakeholders in the system development process and risk analysis, operators may be able to increase the chances of obtaining community approval and ensure that they properly address community concerns. Similarly, involving other stakeholders in system development and risk analysis may also yield benefits. For example, involving employees at all levels in risk analysis ensures that all available information is obtained and therefore increases the likelihood that an accurate representation of risk is made. In addition, involving employees in the risk analysis process increases the likelihood that they will be committed to the risk management plan.

Determining who to involve in the risk analysis process, and how best to ensure that all the necessary stakeholders are heard and understand the process, provides an interesting topic for future research. While professional risk analysts may feel comfortable discussing probabilities and domain experts may feel comfortable reading technical drawings, adequately representing risk-related issues to people with different qualifications presents more of a challenge.

6.2.3 Addressing Dependencies in the Risk Analysis

The risk assessment equations presented in Chapter 5 assume that control flaws are independent. While the two example analyses showed that this assumption may often be true, and that dependencies between control flaws often do not affect the risk assessment, one can expect cases where dependencies are a concern. Future work should address this issue, and develop ways of addressing dependencies. Much research in probabilistic risk assessment has focussed on dependencies, and may therefore provide a useful starting point.

REFERENCES

[Abramovici, 1998] Abramovici, Marianne, "Beyond the Black Box: Organizational Factors in Probabilistic Risk Assessment Methods," *Proceedings of the Society for Risk Analysis 1998 Annual Conference*, Paris, October 1998.

[Ackoff, 1971] Ackoff, Russell L., "Towards a system of systems concept", *Management Science*, Vol. 17, No. 11, July 1971, pp. 661–671.

[Ackoff, 2001] Ackoff, Russell L., "OR: After the Post Mortem", *System Dynamics Review*, Vol. 17, No. 4, pp. 341-346, Winter 2001.

[Adler, 1995] Adler, Paul S., "Interdepartmental interdependence and coordination: The case of the design/manufacturing interface", *Organization Science*, Vol. 6, No. 2, March-April 1995, pp. 147-167.

[Adler and Borys, 1996] Adler, Paul S. and Borys, Bryan, "Two Types of Bureaucracy: Enabling and Coercive", *Administrative Science Quarterly*, Vol. 41 Issue 1, March 1996, pp. 61-89.

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

[Allen, 1997] Allen, Thomas J., "Architecture and Communication Among Product Development Engineers", Sloan School of Management, MIT, Sloan Working Paper #3983, 1997.

[Allen, 2001] Allen, Thomas J., "Organizing for Product Development", MIT Sloan Working Paper No. 4229-01, December 2001. Available online at http://ssrn.com/abstract=297304

[Amalberti, 1996] Amalberti, R., "The Paradoxes of Almost Totally Safe Transportation Systems," *Safety Science*, Vol. 37, pp. 109-126, 2001.Bernstein, 1996.

[Ancona et al., 1998] Ancona, Deborah G., Kochan, Thomas, Van Maanen, John, Scully, Maureen, and Westney, Eleanor, *Managing For The Future: Organizational Behavior and Procedures,* Second Edition, South Western College Pub, 1998.

[Apostolakis, 2004] Apostolakis, George E., "How Useful is Quantitative Risk Assessment?", *Risk Analysis*, Volume 24, Issue 3, June 2004, pp. 515-520.

[Apostolakis and Wu, 1995] Apostolakis, George E. and Wu, J-S, "A structured approach to the assessment of the quality culture in nuclear installations", *Proceedings of American Nuclear Society International Topical Meeting on Safety Culture in*

Nuclear Installations, Vienna, April 24-28, 1995.

[Argyris and Schön, 1978] Argyris, Chris and Schön, Donald A., *Organizational learning*, Addison-Wesley Pub. Co., Reading, Mass, 1978.

[Armstrong and Cole, 2002] Armstrong, D.J. and Cole, P., "Managing distances and differences in geographically dispersed work groups", In Hinds, Pamela and Kiesler, Sara (Eds.), *Distributed Work*, Cambridge, MA, MIT Press, 2002, pp. 167-212.

[Ashford and Cummings, 1983] Ashford, Susan J., and Cummings, L. L., "Feedback as an individual resource: Personal strategies of creating information", *Organizational Behavior and Human Performance*, Vol. 32, 1983, pp. 370-398. Cited in [Morrison and Milliken, 2000].

[Ashford et al., 1998] Ashford, Susan J., Rothbard, Nancy P., Piderit, Sandy K., and Dutton, Jane E., "Out on a limb: The role of context and impression management in selling gender-equity issues", *Administrative Science Quarterly*, Vol. 43, 1998, pp. 23-57.

[Aven and Kørte, 2003] Aven, Terje and Korte, Jens, "On the use of risk and decision analysis to support decision-making." *Reliability Engineering and System Safety*, Vol. 79, Iss. 3, March 2003, pp. 289-299.

[Backström, 1997] Backström, Tomas, "Risk Assessment as a Part of Continuous Safety Management," *Society for Risk Analysis-Europe Annual Meeting*, 1997.

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

[Bate et al., 2000] Bate, Paul, Khan, Raza, and Pye, Annie, "Towards a culturally sensitive approach to organization structuring: Where organization design meets product development", *Organization Science*, Vol. 11, No. 2, March-April 2000, pp. 197-211.

[Bella, 1987] Bella, D., "Organizations and the systemic distortion of information", *Journal of Professional Issues in Engineering*, Vol. 113, No. 4, 1987, pp. 360-370.

[Bendor, 1985] Bendor, Jonathan, *Parallel Systems: Redundancy in Government*, Berkeley, University of California Press, 1985.

[Bendor and Kumar, 2004] Bendor, Jonathan, Kumar, Sunil, "The perfect is the enemy of the good: Adaptive versus optimal organizational reliability", forthcoming in *Journal of Theoretical Politics*, 2004.

[Bernstein, 1996] Bernstein, Peter L., *Against the Gods: The Remarkable Story of Risk*,

John Wiley & Sons, New York, 1996.

[Bier, 1999] Bier, Vicki M., “Challenges to the Acceptance of Probabilistic Risk Analysis,” *Risk Analysis*, Vol. 19, No. 4, 1999, pp. 703-710.

[Bourrier, 1998] Bourrier, Mathilde, “Beyond the Black Box: Organisational Factors in Probabilistic Risk Assessment Methods”, *Society for Risk Analysis Annual Conference—Risk Analysis: Opening the Process*, Paris, October 1998.

[Bourrier, 2003] Bourrier, Mathilde, “Assessing the Contribution of Organizational Design to Safety: A Long neglected Question”, *NATO-Russia Advanced Research Workshop—Forecasting and Preventing Catastrophes: Understanding Human Factors to Enhance Safety Management*, June 2003.

[Bowman and Kunreuther, 1988] Bowman, Edward H. and Kunreuther, Howard, “Post-Bhopal Behaviour at a Chemical Company,” *Journal of Management Studies*, Vol. 25, No. 4, July 1988, pp. 387-402.

[Braun, 2002] Braun, William, “The System Archetypes,” Available online at: http://www.uni-klu.ac.at/~gossimit/pap/sd/wb_sysarChapterpdf, 2002.

[Bruggeman, 2002] Bruggeman, David, “NASA: A Path Dependent Organization,” *Technology in Society*, Vol. 24, 2002, pp. 415-431.

[Brunk et al., 1991] Brunk, Conrad G., Haworth, Lawrence, Lee, Brenda, *Value Assumptions in Risk Assessment: A Case Study of the Alachlor Controversy*, Waterloo, Ontario, Wilfrid Laurier University Press, 1991.

[Buchbinder, 1989] Buchbinder, Benjamin, *Independent Assessment of Shuttle Accident Scenario Probabilities for the Galileo Mission*, Volume 1, NASA/HQ Code QS, Washington, D.C., April 1989.

[Buss et al, 1986] Buss, D.M., Craik, K.H., and Dake, K. M., “Contemporary Worldviews and Perception of the Technological System,” in *Risk Evaluation and Management*, Covello, V.T., Menkes, J., and Mumpower, J. L. (eds.), New York, Plenum, 1986, pp. 93-130.

[Capers and Lipton, 1993] Capers, Robert S. and Lipton, Eric, “Hubble error: Time, money and millionths of an inch”, *Academy of Management Executive*, Vol. 7, No. 4, 1993, pp. 41-57.

[Carroll, 1998a] Carroll, John S., “Safety culture as an ongoing process: culture surveys as opportunities for enquiry and change”, *Work and Stress*, Vol. 12, Iss. 3, July-September 1998, pp. 272-284.

[Carroll, 1998b] Carroll, John S., “Organizational Learning Activities in High-Hazard Industries: The Logics Underlying Self-Analysis,” *Journal of Management Studies*, Vol. 35, No. 6, November 1998, pp. 699-717.

[Carroll et al., 1998] Carroll, John S., Sterman, John and Marcus, A.A., “Losing the Maintenance Game: How Mental Models Drive Organizational Decisions”, In R. N. Stern and J. J. Halpern (Eds.), *Debating Rationality: Nonrational Aspects of Organizational Decision Making*, Ithaca, NY, Cornell University ILR Press, 1998.

[Carroll et al., 2002] Carroll John S., Rudolph J.W., Hatakenaka S., “Learning from experience in high-hazard organizations”, *Research in Organizational Behavior*, Vol. 24, 2002, pp. 87-137.

[Carver et al., 1985] Carver, C. S., Antoni, M., and Scheier, M. F., “Self-consciousness and self-assessment”, *Journal of Personality and Social Psychology*, Vol. 48, 1985, pp. 117-124. Cited in [Morrison and Milliken, 2000].

[Catton, 1985] Catton, W. R., Jr., “Emile Who and the Division of What?” *Sociological Perspectives*, Vol. 28, 1985, pp. 251-80.

[CETS, 1993] Commission on Engineering and Technical Systems, *An Assessment of Space Shuttle Flight Software Development Processes*, Committee for Review of Oversight Mechanisms for Space Shuttle Flight Software Processes, Aeronautics and Space Engineering Board, National Research Council, National Academies Press, 1993. Available online at: http://www.nap.edu/

[Cha and Edmondson, 2003] Cha, Sandra E. and Edmondson, Amy C., “How Values Backfire: Leadership, Attribution, and Disenchantment in a Values-Driven Organization”, Working Paper 03-013, Harvard Business School, June 2003.

[Chatman and Cha, 2003] Chatman, Jennifer A. and Cha, Sandra Eunyoung, “Leading by leveraging culture”, *California Management Review,* Vol. 45, No. 4, Summer 2003, pp. 20-34.

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

[Chess, 2001] Chess, Caron, “Organizational Theory and the Stages of Risk Communication,” *Risk Analysis*, Vol. 21, No. 1, 2001, pp. 179-187.

[Cheyne et al., 1999] Cheyne, Alistair, Tomás, José Manuel, Cox, Sue and Oliver, Amparo, “Modelling Employee Attitudes to Safety: A Comparison Across Sectors”, European Psychologist, Vol. 4, No. 1, March 1999, pp. 1-10.

[Chiles, 2001] Chiles, James R., *Inviting Disaster: Lessons from the Edge of Technology,* Harper Business, New York, 2001.

[Clarke, 1998] Clarke Sharon, “Organizational factors affecting the incident reporting of train drivers”, *Work and Stress*, Vol. 12, No. 1, January-March 1998, pp. 6-16.

[Clarke, 1999] Clarke, Sharon, “Perceptions of Organizational Safety: Implications for the Development of Safety Culture,” *Journal of Organizational Behaviour*, Vol. 20, 1999, pp. 185-198.

[Clarke and Short, 1993] Clarke, Lee and Short, James F. Jr., “Social Organization and Risk: Some Current Controversies,” *Annual Review of Sociology*, Volume 19, 1993, pp. 375-399.

[Clarkson, 1995] Clarkson, Max B.E., “A Stakeholder Framework for Analyzing and Evaluating Corporate Social Performance,” The Academy of Management Review, Vol. 20, Iss. 1, January 1995, pp. 92-117.

[Collinson, 1999] Collinson, David L., “Surviving the rigs: Safety and surveillance on North Sea oil installations,” *Organization Studies*, Vol. 20, Iss. 4, 1999, pp. 579-600.

[Cooke, 2003] Cooke, David L., “A System Dynamics Analysis of the Westray Mine Disaster,” *System Dynamics Review*, Vol. 19, No. 2, Summer 2003, pp. 139-166.

[Cooper, 2000] Cooper, M.D., “Towards a Model of Safety Culture,” *Safety Science*, Vol. 36, Iss. 2, November 2000, pp. 111-136.

[Cotgrove, 1981] Cotgrove, Stephen, “Risk, Value Conflict, and Political Legitimacy,” in Griffiths, Richard F., editor, *Dealing with Risk: The Planning, Management, and Acceptability of Technological Risk*, John Wiley and Sons, New York, 1981.

[Cox et al., 1998] Cox, Sue, Tomas, Jose M., Cheyne, Alistair and Oliver, Amparo, “Safety Culture: The Prediction of Commitment to Safety in the Manufacturing Industry”, *British Journal of Management*, September 1998, Special Issue 1997 Conference, Vol. 9 Iss. 3, 1998.

[Crawford, 2001] Crawford, Jack, “What‘s Wrong With the Numbers? A Questioning Look at Probabilistic Risk Assessment”, *Journal of System Safety*, 3rd Quarter 2001.

[Cyert and March, 1992] Cyert, R.M. and March, James G., *A Behavioral Theory of the Firm*, 2nd ed., Blackwell, Oxford, 1992.

[Daouk et al., 2004] Daouk, M., Dulac, N., Weiss, K., Zipkin, D., Leveson, N., “A Practi-

cal Guide to STAMP-Based Hazard Analysis", International System Safety Conference, 2004.

[Davoudian et al., 1994a] Davoudian, Kvyan, Wu, Jya-Syin., Apostolakis, George, "Incorporating Organizational Factors into Risk Assessment through the Analysis of Work Processes," *Reliability Engineering and System Safety*, Vol. 45, 1994, pp. 85-105.

[Davoudian et al., 1994b] Davoudian, Kvyan, Wu, Jya-Syin., Apostolakis, George, "The work process analysis model (WPAM)", *Reliability Engineering and System Safety*, Vol. 45, 1994, pp. 107-125.

[Deal and Kennedy, 1982] Deal, Terrence E. and Kennedy, Allan A., *Corporate cultures: The rites and rituals of corporate life*, Reading, MA, Addison-Wesley Pub. Co., 1982.

[De Jong and Koster, 1974] De Jong, J.J. and Koster E. P., "The human operator in the computer-controlled refinery". In Edwards, Elwyn and Lees, Frank P. (Eds.), *The Human Operator in Process Control*, London, Taylor & Francis, 1974, pp. 196-205. Cited in [Perrow, 1999b].

[Dekker, 2003] Dekker, Sydney, "Failure to adapt or adaptations that fail: contrasting models on procedures and safety", Applied Ergonomics, Vol. 34, No. 3, May 2003, pp. 233-238.

[Denison, 1996] Denison, Daniel R., "What is the difference between organizational culture and organizational climate? A native's point of view on a decade of paradigm wars", *The Academy of Management Review*, Vol. 21, Iss. 3, July 1996, pp. 619-654.

[Dietz and Rycroft, 1987] Dietz, Thomas M., Rycroft, Robert W., *The Risk Professionals*, Sage, New York, 1987.

[Dulac and Leveson, 2004] Dulac, Nicolas and Leveson, Nancy, G., "An Approach to Design for Safety in Complex Systems", INCOSE International System Engineering Conference, June 2004.

[Dutton et al., 1997] Dutton, Jane E., Ashford, Susan J., O'Neill, Regina M., Hayes, Erika, and Wierba, Elizabeth E., "Reading the wind: How middle managers assess the context for selling issues to top managers", *Strategic Management Journal*, Vol. 18, 1997, pp. 407-425.

[Dyer, 2000] Dyer, C., "The lessons from Sellafield", *Health and Safety Bulletin*, No. 287, 2000, pp. 7-14.

[Edmondson, 1999] Edmondson, Amy, “Psychological safety and learning behavior in work teams”, *Administrative Science Quarterly*, Vol. 44, 1999, pp. 350-383.

[Embrey, 1986] Embrey, D.E., “SHERPA: A Systematic Human Error Reduction and Prediction Approach,” *Proceedings of Advances in Human Factors in Nuclear Power Systems Meeting*, Knoxville, Tennessee, 1986.

[Embrey, 1992] Embrey, D.E., “Incorporating Management and Organizational Factors into Probabilistic Safety Assessment,” *Reliability Engineering and System Safety,* Vol. 38, 1992, pp. 199-208.

[Erev et al., 1993] Erev, Ido, Bornstein, Gary, and Wallsten, Thomas S., “The Negative Effect of Probability Assessments on Decision Quality,” *Organizational Behavior and Human Decision Processes*, Vol. 55, 1993, pp. 78-94.

[Feynman, 1986] Feynman, Richard, “Personal observations on the reliability of the Shuttle”, Appendix F of Rogers, William P., Chairman, *Report of the Presidential Commission on the Space Shuttle Challenger Accident*, Government Printing Office, Washington DC, 1986.

[Fischhoff, 1995] Fischhoff, Baruch, “Risk perception and communication unplugged: Twenty years of process,” *Risk Analysis*, Vol. 15, No. 3, 1995, pp. 137-145.

[Fischhoff et al., 1977] Fischhoff, Baruch, Slovic, Paul, Lichtenstein, Sarah, “Knowing with Certainty: The Appropriateness of Extreme Confidence,” *Journal of Experimental Psychology: Human Perception and Performance 3*, 1977, pp. 552-564.

[Fischhoff et al., 1983] Fischhoff, Baruch, Lichtenstein, Sarah, Slovic, Paul, Derby, Steven L., Keeney, Ralph L., *Acceptable Risk*, Cambridge University Press, Cambridge, 1983.

[Fisher, 2000] Fisher, Kimball, *Leading Self-Directed Work Teams: A Guide to Developing New Team Leadership Skills*, New York, McGraw-Hill, 2000.

[Fletcher 1996] Fletcher, Sharon K., “Risk Management: What About Software?”, Proceedings of the 14th National System Safety Conference, Albuquerque, New Mexico, 12-17 August 1996.

[Forrester, 1961] Forrester, Jay W., *Industrial Dynamics*, Pegasus Communications, 1961.

[Foucault, 1970] Foucault, Michel, *The order of things: an archaeology of the human sciences, A translation of Les mots et les choses,* New York, Pantheon Books, 1970.

[Freudenburg, 1992] Freudenburg, William R., “Nothing Recedes Like Success? Risk Analysis and the Organizational Amplification of Risks”, *Risk: Issues in Health and Safety,* Vol. 3, No. 1, 1992, pp. 1-35.

[Futron, 2000] Futron Corporation, “NASA PRA Practices and Needs for the New Millennium: International Space Station Probabilistic Risk Assessment Stage 7A,” 2000. Available at: www.hq.nasa.gov/office/codeq/risk/workshop/smith.ppt.

[Gaertner et al., 1987] Gaertner, G., Newman, P., Fisher, G., and Whitehead, K., “Determining the effects of management practices on coal miners’ safety”, Proceedings of Human Engineering and Human Resource Management in Mining, 1987. Cited in [O’Dea and Flin, 2003].

[Garret and Apostolakis, 1999] Garret, Chris J., Apostolakis, George E., “Context in the Risk Assessment of Digital Systems”, *Risk Analysis,* Vol. 19, No. 1, 1999, pp. 23-32.

[Garret and Apostolakis, 2002] Garret, C. J., Apostolakis, G. E., “Automated Hazard Analysis of Digital Control Systems”, *Reliability Engineering and System Safety,* Vol. 77, Iss. 1, July 2002, pp. 1-17.

[Garret et al, 1995] Garret, C. J., Guarro, S. G., Apostolakis, G. E., “The Dynamic Flowgraph Methodology for Assessing the Dependability of Embedded Software Systems”, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 25, No. 5, May 1995.

[Garrick, 1987] Garrick, B.J., “Examining the Realities of Risk Assessment,” *Society for Risk Analysis International Workshop on Uncertainty in Risk Assessment, Risk Management, and Decision Making (1984: Knoxville, Tenn.),* Plenum Press, New York, 1987.

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

[Giddens, 1984] Giddens, Anthony, *The constitution of society: Outline of the theory of structuration*, Berkeley, University of California Press, 1984.

[Gould, 1985] Gould, Stephen Jay, *The flamingo's smile: reflections in natural history,* New York, Norton, 1985.

[Graham, 1971] Graham, John, *Fast Reactor Safety,* Academic Press, New York, 1971.

[Graicunas, 1933] Graicunas, A.V., “Relationship in Organization”, 1933. Reprinted in Gulick, Luther and Urwick, Lyndall F. (Eds.), *Papers on the Science of Administration*, Institute of Public Administration, New York, NY, 1937.

[Griffiths, 1981] Griffiths, Richard F., “The Nature of Risk Assessment,” in Griffiths, Richard F., editor, *Dealing with Risk: The Planning, Management, and Acceptability of Technological Risk*, John Wiley and Sons, New York, 1981.

[Haimes, 1998] Haimes, Yacov Y., *Risk Modeling, Assessment, and Management*, Wiley-Interscience, New York, 1998.

[Haimes, 1999] Haimes, Yacov Y., “Risk Management,” in Sage, Andrew P. and Rouse, William B., (Eds.), *Handbook of Systems Engineering and Management*, Wiley-Interscience, New York, 1999.

[Hall et al., 1967] Hall, Richard H., Johnson, Norman J., Haas, J. Eugene, “Organizational Size, Complexity, and Formalization”, *American Sociological Review*, Vol. 32, No. 6. December, 1967, pp. 903-912.

[Hatch, 1993] Hatch, Mary Jo, “The Dynamics of Organizational Culture”, *The Academy of Management Review*, Vol. 18, No. 4, October 1993, pp. 657-693.

[Hatfield and Hipel, 2002] Hatfield, Adam J., Hipel, Keith W., “Risk and Systems Theory,” *Risk Analysis*, Vol. 22, No. 6, 2002, pp. 1043-1057.

[Harrison, 1972] Harrison, R., “Understanding your Organization‘s Character,” *Harvard Business Review*, May-June, 1972, pp. 119-128.

[Heimann, 1993] Heimann, C. F. Larry, “Understanding the Challenger Disaster: Organizational Structure and the Design of Reliable Systems”, *The American Political Science Review*, Vol. 87, No. 2, June 1993, pp. 421-435.

[Heimann, 1997] Heimann, C. F. Larry, *Acceptable risks: politics, policy, and risky technologies*, Ann Arbor, 1997.

[Helmreich and Merritt, 1998] Helmreich, R. L., and Merritt A.C., “Organizational culture,” In Helmreich, R. L., and Merritt A.C. (Eds.), *Culture at work in aviation and medicine*, Ashgate, Brookfield, Vermont, 1998, pp. 107-174.

[Hendricks, 1991] Hendricks, Hal, W., “Ergonomics in Organizational Design and Management”, *Ergonomics*, Vol. 34, No. 6, 1991, pp. 743-756.

[Hess et al., 2005] Hess, Stephen M. Alfonso M. Albano and John P. Gaertner, “Development of a dynamical systems model of plant programmatic performance on nuclear power plant safety risk “, Reliability Engineering and System Safety, In Press, Available online 5 January 2005.

[Hoegberg, 1998] Lars Hoegberg, “Risk perception, safety goals and regulatory decision-making”, *Reliability Engineering & System Safety*, Vol. 59, Iss. 1, January 1998,

pp. 135-139.

[Hollnagel, 1993] Hollnagel, Erik, *Human Reliability Analysis: Context and Control*, Academic Press, London, 1993.

[Hollnagel, 1998] Hollnagel, Erik, *Cognitive Reliability and Error Analysis Method*, Elsevier, London, 1998.

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

[HSC, 1993] Health and Safety Commission, *Organising for Safety: Third Report of ACSNI Study Group on Human Factors*, HMSO, ISBN 0 11 882104 0, London, 1993.

[Hurst, 1998] Hurst, Nick W., *Risk Assessment: The Human Dimension*, The Royal Society of Chemistry, Cambridge, UK, 1998.

[IAEA ASCOT, 1996] International Atomic Energy Agency, Assessment of Safety Culture in Organizations Team, *ASCOT Guidelines: Guidelines for Organizational Self-Assessment of Safety Culture and for Reviews*, IAEA-TECDOC-860, Vienna, 1996.

[IAEA BSS-115, 1996] International Atomic Energy Agency Basic Safety Series No.115: International Basic Safety Standard for Protection Against Ionizing Radiation and for the Safety of Radioactive Materials (IAEA BSS-115), p.354, 1996.

[IAEA INSAG-7, 1992] The IAEA's International Nuclear Safety Advisory Group (INSAG), *The Chernobyl Accident: Updating of INSAG-1 (INSAG-7)*, p.24, 1992.

[Ilgen et al., 1979] Ilgen, Daniel R., Fisher, C. D. and Taylor, M. S., “Consequences of individual feedback on behavior in organizations”, *Journal of Applied Psychology*, Vol. 64, 1979, pp. 349-371. Cited in [Morrison and Milliken, 2000].

[Jones, 2002] Jones, Bethan, “Theoretical Approaches to Organizational Learning,” *LearnSafe: Learning Organizations for Nuclear Safety*, European Commission: 5th Euratom Framework Programme 1998-2002, Contract No. FIKS-CT-2001-00162, 2002.

[Julius et al., 1995] Julius, J., Jorgenson, E., Parry, G. W. and Mosleh, A. M., “A Procedure for the Analysis of Errors of Commission in a Probabilistic Safety Assessment of a Nuclear Power Plant at Full Power,” *Reliability Engineering and*

System Safety, Vol. 50, 1995, pp. 189–201.

[Kahneman and Tversky, 1979] Kahneman, Daniel and Tversky, Amos, “Prospect Theory: An Analysis of Decision under Risk,” *Econometrica*, Vol. 47 (2), 1979, pp. 263-291.

[Kahneman and Tversky, 1984] Kahneman, D., Tversky, A., “Choices, Values and Frames,” *American Psychologist*, Vol. 39, 1984, pp. 341-350.

[Kahneman et al., 1982] Kahneman, D., Slovic, P., Tversky, A., *Judgment under Uncertainty: Heuristics and Biases*, Cambridge University Press, New York, 1982.

[Keating et al., 1999] Keating, Elizabeth K., Oliva, Rogelio, Repenning, Nelson P., Rockart, Scott, and Sterman, John D., “Overcoming the Improvement Paradox”, *European Management Journal*, Vol. 17, No. 2, 1999, pp. 120-134.

[Kemeny, 1979] Kemeny, John G., Chairman, *Report of The President's Commission on The Accident at Three Mile Island*, October, 1979. Available online at http://stellar-one.com/nuclear/report_to_the_president.htm.

[Kennedy and Kirwan, 1998] Kennedy, R., and Kirwan, B., “Development of a Hazard and Operability-based method for identifying safety management vulnerabilities in high risk systems,” *Safety Science*, Vol, 30, No. 3, 1998, pp. 249-274.

[King et al., 2002] King J., Down J.T., Bella D.A., “Learning to Think in Circles”, *Journal Of Management Inquiry*, Vol. 11, No. 2, June 2002, pp. 161-170.

[Kletz, 1994] Kletz, Trevor A., *Learning from Accidents*, Second Edition, Butterworth-Heinemann Ltd., Oxford, UK, 1994.

[Klinke and Renn, 2002] Klinke, Andreas, and Renn, Ortwin, “A New Approach to Risk Evaluation and Management: Risk-Based, Precaution-Based, and Discourse-Based Strategies,” *Risk Analysis,* Vol. 22, No. 6, 2002, pp. 1071-1094.

[Korsgaard et al., 1998] Korsgaard, M. Audrey, Roberson, L. & Rymph, R.D., “What Motivates Fairness? The role of subordinate assertive behavior on managers’ interactional fairness”, *Journal of Applied Psychology*, Vol. 83, 1998, pp. 731-744. Cited in [Morrison and Milliken, 2000].

[Kraft, 1995] Kraft, Christopher, *Report of the Space Shuttle Management Independent Review Team*, February 1995.

[Kunreuther and Bowman, 1997] Kunreuther, Howard and Bowman Edward H., “A Dynamic Model of Organizational Decision Making: Chemco Revisited Six Years After Bhopal,” *Organization Science*, Vol. 8, No. 4, July-August, 1997, pp.

404-413.

[Kunreuther and Meszaros, 1997] Kunreuther, Howard and Meszaros, Jacqueline, “Organizational Choice under Ambiguity,” in Shapira, Zur, Ed., *Organizational Decision Making*, Cambridge University Press, Cambridge, United Kingdom, 1997.

[Lane, 1996] Lane, David C., “Reinterpreting ‘Generic Structure‘: Evolution, Application and Limitations of a Concept,” *System Dynamics Review*, Vol. 12, pp. 87-120, 1996.

[La Porte, 1996] La Porte, Todd R, “High Reliability Organizations: Unlikely, Demanding, and At Risk”, *Journal of Contingencies and Crisis Management*, Vol. 63, No. 4, 1996.

[La Porte and Consolini, 1991] 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*, Vol. 1, 1991, pp. 19–47.

[La Porte et al., 1994] La Porte, Todd R. and Rochlin, Gene, “A Rejoinder to Perrow”, *Journal of Contingencies and Crisis Management*, Vol. 2, No. 4, 1994.

[Larsson and Hale, 2000] Larsson, Tore J., Hale, Andrew R., “Aspects of Risk Assessment, Control and Prevention,” *Safety Science Monitor*, Vol. 4, Iss. 1, 2000.

[Lawler, 1994] Lawler, Edward E. III, “Total Quality Management and employee involvement: Are they compatible?” *Academy of Management Executive*, Vol. 9, No. 1, 1994, pp. 68-76.

[Lee, 1998] Lee Terence, “Assessment of Safety Culture At A Nuclear Reprocessing Plant”, *Work and Stress*, Vol. 12, No. 3, July-September 1998, pp. 217-237.

[Lee and Harrison, 2000] Lee, Terence and Harrison, K., “Assessing safety culture in nuclear power stations”, *Safety Science*, Vol. 34, Iss. 1, February 2000, pp. 61-97.

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

[Leveson, 1995] Leveson, Nancy G., *Safeware: System Safety and Computers*, Addison Wesley, Reading, Massachusetts, 1998.

[Leveson, 2003] Leveson, Nancy G., “A New Approach to Hazard Analysis for Complex Systems”, International Conference of the System Safety Society, Denver, CO, 2003.

[Leveson, 2004a] Leveson, Nancy G., “A New Accident Model for Engineering Safety Systems”, *Safety Science*, Vol. 42, No. 4, April 2004, pp. 237–270.

[Leveson, 2004b] Leveson, Nancy G., “Role of Software in Spacecraft Accidents”, *Journal of Spacecraft and Rockets*, Vol. 41, No. 4, 2004, pp. 564- 575.

[Leveson, 2005] Leveson, N.G., A new approach to system safety engineering, Unpublished manuscript. Available online at sunnyday.mit.edu.

[Leveson and Stolzy, 1987] Leveson, Nancy G., Stolzy, Janet L., “Safety Analysis using Petri Nets”, *IEEE Transactions on Software Engineering*, SE-13(3), pp. 386-397, March 1987.

[Leveson et al., 2002] Leveson, Nancy G., Allen, Polly, Storey, Margaret-Anne, “The Analysis of a Friendly Fire Accident using a Systems Model of Accidents,” *Proceedings of the 20th International System Safety Conference*, Denver Colorado, 5-9 August 2002.

[Leveson et al., 2004a] Nancy Leveson, Joel Cutcher-Gershenfeld, Betty Barrett, Alexander Brown, John Carroll, Nicolas Dulac, Lydia Fraile, Karen Marais, “Effectively Addressing NASA's Organizational and Safety Culture: Insights from Systems Safety and Engineering Systems”, *ESD External Symposium*, March 2004.

[Leveson et al., 2004b] 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.

[Levitt and Parker, 1976] Levitt, Raymond E., and Henry W. Parker, “Reducing Construction Accidents–Top Management's Role,” *ASCE Journal of the Construction Division*, Vol. 102, No. CO3, September 1976, pp. 465-478.

[Logsdon, 1986] Logsdon, John M., “The Space Shuttle Program: A Policy Failure?” *Science*, Vol. 232, No. 4754, May 30, 1986, pp. 1099-1105.

[Lynch, 1996] Nancy Lynch, R. Segala, F. Vaandrager, and H. B. Weinberg, “Hybrid Input/Output Automata” In Alur, R., Henzinger, T., and Sontag, E., Editors, *Hybrid Systems III: Verification and Contro*l (DIMACS/SYCON Workshop on Verification and Control of Hybrid Systems, New Brunswick, New Jersey, October 1995), Volume 1066 of Lecture Notes in Computer Science, pp. 496-510, Springer-Verlag, 1996.

[Majchrzak et al., 2004] Majchrzak, Ann, Malhotra, Arvind, Stamps, Jeffrey, and Lipnack, Jessica, “Can Absence Make a Team Grow Stronger?” *Harvard Business Review*, May2004, Vol. 82, Iss. 5, pp. 131-137.

[Marais and Leveson, 2003] Marais, Karen and Leveson, Nancy, G., “Archetypes for Organizational Safety”, *Workshop on Investigating and Reporting of Incidents and Accidents*, Williamsburg, VA, September 2003. Available online at: http://shemesh.larc.nasa.gov/iria03/p01-marais.pdf

[Marais et al., 2004] Marais, Karen, Dulac, Nicolas, and Leveson, Nancy, G., “Beyond Normal Accidents and High Reliability Organizations: Lessons from the Space Shuttle”, *ESD External Symposium*, Cambridge, MA, March 2004. Available online at: http://esd.mit.edu/symposium/pdfs/papers/marais-b.pdf

[March, 1978] March, James G., “Bounded Rationality, Ambiguity, and the Engineering of Choice”, *The Bell Journal of Economics*, Vol. 9, No. 2. (Autumn, 1978), pp. 587-608.

[March, 1997] March, James G., “Understanding how Decisions happen in Organizations,” in *Organizational Decision Making*, Shapira, Zur, Ed., Cambridge University Press, Cambridge, United Kingdom, 1997.

[March and Shapira, 1987] March, James G. and Shapira, Zur, “Managerial Perspectives on Risk and Risk Taking,” *Management Science*, Vol. 33, No. 11, November 1987, pp. 1404-1418.

[March and Simon, 1993] March, James G., and Simon, Herbert A., *Organizations*, Second Edition, Blackwell Business, Cambridge, MA, 1993. First edition published 1958 by Wiley.

[March et al., 1991] March, James G., Sproull, Lee S., and Tamuz, Michal, “Learning from Samples of One or Fewer,” *Organization Science*, Vol. 2, No. 1, February 1991, pp. 1-13.

[Martin, 2002] Martin, Joanne, *Organizational Culture*, Thousand Oaks, CA, Sage Publications, 2002.

[Martin, in press] Martin, Joanne, “Organizational Culture”, in Nigel Nicholson, P. Audia, and M. Pillutla (eds.), The Blackwell Encyclopedic Dictionary of Organizational Behavior, Second Ed., Blackwell Publishers, In press.

[Masuch, 1985] Masuch, Michael, “Vicious Circles in Organizations”, *Administrative Science Quarterly*, Vol. 30, No. 1, March 1985, pp. 14-33.

[Maurino et al., 1995] Maurino, Daniel E., Reason, James, Johnston, Neil, and Lee, Rob B., *Beyond Aviation Human Factors: Safety in High Technology Systems*, Avebury Aviation, Aldershot, England, 1995.

[McCurdy, 1993] McCurdy, Howard E., *Inside NASA: High Technology and Organiza-*

tional Change in the U.S. Space Program, Johns Hopkins University Press, Baltimore, 1993.

[McCurdy, 2001] McCurdy, Howard E., *Faster, Better, Cheaper: Low-Cost Innovation in the U.S. Space Program*, Johns Hopkins University Press, Baltimore, 2001.

[Meadows et al., 1972] Meadows, Donella, Meadows, Dennis, Randers, Jorgen, and Behrens, William, *Limits to Growth*, New American Library, New York, 1972.

[Meek, 1988] Meek, V. Lynn, “Organizational Culture: Origins and Weaknesses”, *Organization Studies*, Vol. 9, Iss. 4, 1988, pp. 453-473.

[Merton, 1968] Merton, Robert King, *Social theory and social structure*, Free Press, New York, 1968.

[Mileti et al., 1977] Mileti, Dennis S., Gillespie, David F., Haas, J. Eugene, “Size and Structure in Complex Organizations”, *Social Forces*, Vol. 56, No. 1, September 1977, pp. 208-217.

[Miller and Friesen, 1980] Miller, Danny and Friesen, Peter, “Archetypes of Organizational Transition”, *Administrative Science Quarterly*, Vol. 25, No. 2, June 1980, pp. 268-299.

[Milliken and Morrison, 2003] Milliken, Frances J. and Morrison, Elizabeth Wolfe, “Speaking Up, Remaining Silent: The Dynamics of Voice and Silence in Organizations,” *Journal of Management Studies*, September 2003, pp. 1563-1568.

[Milliken et al., 2003] Milliken, Frances J., Morrison, Elizabeth Wolfe and Hewlin, Patrica F., “An Exploratory Study of Employee Silence: Issues that Employees Don't Communicate Upward and Why,” *Journal of Management Studies*, September 2003, pp. 1453-1476.

[Mitroff and Alpaslan, 2003] Mitroff, Ian I. and Alpaslan, Murat C., “Preparing for Evil,” *Harvard Business Review*, April 2003, pp. 109-115.

[Moray and Huey, 1988] Moray, N., and B. Huey, (Eds.), *Human Factors Research and Nuclear Safety. Committee on Human Factors*, National Research Council. Washington, DC: National Academy Press, 1988

[Morecroft, 1983] Morecroft, John D.W., “System Dynamics: Portraying Bounded Rationality,” International Journal of Management Science, Vol. 11, No. 2, 1983, pp. 131-142.

[Morgan, 1996] Morgan, Gareth, *Images of organization*, Beverly Hills, Sage Publications, 1986.

[Morrison and Milliken, 2000] Morrison, Elizabeth Wolfe and Milliken, Frances J., “Organizational silence: A barrier to change and development in a pluralistic world”, *Academy of Management Review*, Vol. 25, No. 4, 2000, pp. 706-725.

[Morrison and Milliken, 2003] Morrison, Elizabeth Wolfe and Milliken, Frances J., “Speaking up, remaining silent: The dynamics of voice and silence in organizations”, *Journal of Management Studies*, Vol. 40, No. 6, September 2003, pp. 1353-1358.

[Moses, 2002] Moses, Joel, “Complexity and Flexibility,” ESD Working Paper, MIT, Cambridge, MA, 2002.

[Mueller, 1996] Mueller, Frank, “Human resources as strategic assets: An evolutionary resource-based theory”, *Journal of Management Studies*, Vol. 33, No. 6, 1996, 757-785.

[Murphy, 1994] Murphy, Dean Michael, *Incorporating Human and Management Factors in Probabilistic Risk Analysis*, Ph.D. Thesis, Department of Industrial Engineering and Management, Stanford University, 1994.

[Murphy and Paté-Cornell, 1996] Paté-Cornell, Elisabeth M., Murphy, Dean Michael, “Human and management factors in probabilistic risk analysis: the SAM approach and observations from recent applications,” *Reliability Engineering and System Safety*, Vol. 53, No. 2, August 1996, pp. 115-126.

[NASA, 1988] Press Release 88-01-05. Available online at: http://spacelink.nasa.gov/NASA.News/NASA.News.Releases/.index.html

[NASA, 1995] National Aeronautics and Space Administration, *NASA Systems Engineering Handbook*, 1995.

[NASA, 2002] Office of Inspector General, “NASA Oversight of United Space Alliance's Safety Procedures at the John F. Kennedy Space Center”, Report No. IG-02-018. Available online at: http://www.hq.nasa.gov/office/oig/hq/ig-02-018r.pdf

[Negandhi, 1973] Negandhi, Anant R., “A model for analysing organization in cross cultural settings: a conceptual scheme and some research findings”” in Negandhi, Anant R., Ed., *Modern organizational theory; contextual, environmental, and socio-cultural variables*, Kent State University Press, Kent, Ohio, 1973.

[Negandhi, 1983] Negandhi, Anant R., “Cross-Cultural Management Research: Trend and Future Directions”, *Journal of International Business Studies, Vol. 14, No. 2, Special Issue on Cross-Cultural Management*, Autumn, 1983, pp. 17-28.

[Nemeth, 1997] Nemeth, Charlan Jeanne, “Managing Innovation: When Less is More”,

California Management Review, Fall 1997, Vol. 40, Iss. 1, 1997, pp. 59-74.

[Neogi, 2002] Neogi, N. A., *Hazard Elimination using Backwards Reachability Techniques in Discrete and Hybrid Models*, Ph. D. Thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, 2002.

[Nilsen and Aven, 2003] Nilsen, Thomas and Aven, Terje, "Models and model uncertainty in the context of risk analysis," *Reliability Engineering and System Safety*, Vol. 79, Iss. 3, March 2003, pp. 309-317.

[NTSB, 1983] NTSB Accident Report, NTSB Identification MIA83AA136, NTSB microfiche number 23663. Available online at: www.ntsb.gov

[Nuclearfiles, 2004] http://www.nuclearfiles.org/hitimeline/nwa/60/1966.html

[O'Dea and Flin, 2003] O'Dea, Angela and Flin, Rhona, *The Role of Managerial Leadership in Determining Workplace Safety Outcomes*, British Health and Safety Executive, Colegate, Norwich, United Kingdom, 2003.

[Ogata, 1990] Ogata, Katsuhiko, *Modern Control Engineering*, Second Edition, Prentice-Hall International, Englewood Cliffs, New Jersey, 1990.

[O'Hara, 2004] O'Hara, Patrick W., http://nuclearsafetyculture.freeyellow.com/page2.html.

[O'Leary and Cummings, 2002] O'Leary, Michael B. and Cummings, Jonathan N., "The spatial, temporal, and configurational characteristics of geographic dispersion in work teams", Center for eBusiness@MIT, Paper 148, December 2002. Available online at http://ebusiness@mit.edu.

[Oliver and Smith, 1990] Oliver, Robert M., Smith, James Q., eds., *Influence Diagrams, Belief Nets and Decision Analysis*, Wiley, New York, New York, 1990.

[Oreskes, 1994] Oreskes, Naomi, Shrader-Frechette, Kristin, Belitz, Kenneth, "Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences," *Science*, New Series, Volume 263, Issue 5147, pp. 641-646, February 1994.

[Orlikowski, 2000] Orlikowski, Wanda, "Using Technology and Constituting Structures: A Practice Lens for Studying Technology in Organizations", *Organization Science*, Vol. 11, No. 4, July-August 2000, pp. 404-428.

[Paich, 1985] Paich, M., "Generic Structures," *System Dynamics Review,* Vol. 1, pp. 126-132, 1985.

[Pate-Cornell and Dillon, 2001] Pate-Cornell, Elisabeth, and Dillon, Robin, "Probabilis-

tic Risk Analysis for the NASA Space Shuttle: A Brief History and Current Work," Reliability Engineering and System Safety, Vol. 74, 2001, pp. 345-352.

[Pate-Cornell and Murphy, 1996] Pate-Cornell, Elisabeth, and Murphy, Dean M., "Human and management factors in probabilistic risk analysis: the SAM approach and observations from recent applications", *Reliability Engineering and System Safety*, Vol. 53, 1996, pp. 115-126.

[Payne, 2000] Payne, Roy L., "Climate and culture: How close can they get?" in Ashkanasy, Neal M., Wilderom, Celeste P.M. and Peterson, Mark F. (eds.), *Handbook of Organizational Culture and Climate*, Sage Publications Inc., Thousand Oaks, CA, 2000, pp. 163-176.

[Perin, 1998] Perin, Constance, "Operating as Experimenting: Synthesising Engineering and Scientific Values in Nuclear Power Production," *Science, Technology and Human Values*, Vol. 23, No. 1, Winter 1998, pp. 98-128.

[Perrow, 1967] Perrow, Charles, "A Framework for the Comparative Analysis of Organizations", *American Sociological Review*, Vol. 32, No. 2, April 1967, pp. 194-208.

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

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

[Perrow, 1999b] Perrow, Charles, "The Organizational Context of Human Factors Engineering", *Administrative Science Quarterly*, Vol. 28, 1983, pp. 521-541.

[Pidgeon, 1998] Pidgeon, Nick F., "Safety culture: key theoretical issues", *Work and Stress*, Vol. 12, No. 3, pp. 202-216.

[Polmar, 1982] Polmar, Norman, *Rickover*, Simon & Schuster, New York, 1982.

[Pritchard, 2001] Pritchard, Jocelyn, "Overview of Landing Gear Dynamics", *Journal of Aircraft*, Vol. 38, No. 1, January-February 2001, pp. 130-137.

[Probst and Brubaker, 2001] Probst, T. M., and Brubaker, T. L., "The effects of job insecurity on employee safety outcomes: cross-sectional and longitudinal explorations, *Journal of Occupational Health Psychology*, Vol. 6, No. 2, 2001, pp. 139-159.

[Quinn and Walsh, 1994] Quinn, Robert E. and Walsh, James P., "Understanding organizational tragedies: The case of the Hubble Space Telescope", *Academy of Man-*

agement Executive, Vol. 8, No. 1, 1994, pp. 62-67.

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

[Ranson et al., 1980] Ranson, Stewart, Hinings, Bob, and Greenwood, Royston, “The structuring of organizational structures”, *Administrative Science Quarterly,* Vol. 25, No. 1, March 1980, pp. 1-17.

[Rausand and Hoyland, 2004] Rausand, Marvin and Hoyland, Arnljot, *System Reliability Theory: Models, Statistical Methods, and Applications*, Wiley-Interscience, Hoboken, New Jersey, 2004.

[Rasmussen and Whetton, 1997] Rasmussen, Birgitte and Whetton, Cris, “Hazard identification based on plant functional modelling”, *Reliability Engineering & System Safety,* Volume 55, Issue 2, February 1997, pp. 77-84.

[Rasmussen, 1974] Rasmussen, Jens, “On the communication between operators and instrumentation in automatic process plants”. In Edwards, Elwyn and Lees, Frank P. (Eds.), *The Human Operator in Process Control*, London, Taylor & Francis, 1974, pp. 196-205. Cited in [Perrow, 1999b].

[Rasmussen, 1991] Rasmussen, Jens, “The Application of Probabilistic Risk Assessment Techniques to Energy Technologies,” in *Reading in Risk,* Glickman, Theodore S. and Gough, Michael, eds., Resources for the Future, Washington D.C., 1991, pp. 195-206.

[Rasmussen, 1997] Rasmussen, Jens, “Risk Management in a Dynamic Society: A Modelling Problem,” *Safety Science*, Vol. 27, No. 2/3, 1997, pp. 183-213.

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

[Rasmussen, 1975] Rasmussen, Norman, *Reactor Safety Study (WASH-1400),* US Nuclear Regulatory Commission, NUREG-75/014, 1975.

[Reason, 1987] Reason, James, *Bulletin of the British Psychological Society,* Vol. 40, April 1987, p. 201.

[Reason, 1997] Reason, James, *Managing the Risks of Organizational Accidents*, Ashgate, Aldershot, England, 1997.

[Redmill, 2002] Redmill, Felix, “Exploring subjectivity in hazard analysis,” *IEEE Engineering Management Journal*, June 2002, pp. 139-144.

[Repenning and Sterman, 2001] Repenning, Nelson P. and Sterman, John D., "Nobody Ever Gets Credit for Fixing Problems that Never Happened: Creating and Sustaining Process Improvement", *California Management Review*, Vol. 43, No. 4, Summer 2001, pp. 64-88.

[Rice, 1970] Rice, Albert Kenneth, *Productivity and Social Organization: The Ahmedabad Experiment*, Tavistock Publications, London, 1970.

[Richter and Koch, 2004] Richter, Anne and Koch, Christian, "Integration, differentiation and ambiguity in safety cultures", *Safety Science*, Vol. 42, Iss. 8, October 2004, pp. 703-722.

[Robbins, 1983] Robbins, Stephen P., *Organization theory: the structure and design of organizations*, Englewood Cliffs, N.J, Prentice-Hall, 1983.

[Robbins, 1992] Robbins, Stephen P., *Essentials of organizational behavior*, Englewood Cliffs, N.J., Prentice Hall, 1992.

[Roberts, 1990a] Roberts, Karlene H., "Managing high reliability organizations", *California Management Review*, Vol. 32, No. 4, 1990, pp. 101–114.

[Roberts, 1990b] Roberts, Karlene H. "Some characteristics of one type of high reliability organization", *Organization Science*, Vol. 1, No. 2, 1990, pp. 160–176.

[Rochlin, 1991] 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.

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

[Rogers, 1986] Rogers, William P., Chairman, *Report of the Presidential Commission on the Space Shuttle Challenger Accident*, Government Printing Office, Washington DC, 1986.

[Rollenhagen, 2000] Rollenhagen, Carl, "A Framework for Assessment of Organizational Characteristics and their Influences on Safety," Safety Science Monitor, Vol. 4, Iss. 1, 2000.

[Rose, 1999] Rose, J., "Towards a structurational theory of IS: Theory development and case study illustrations", *Proceedings of the 7th European Conference on Information Systems*, Copenhagen, 1999.

[Royal Society, 1992] The Royal Society, *Risk: Analysis, Perception, and Management. Report of a Royal Society Study Group*, Royal Society, London, 1992.

[Rudolph and Repenning, 2002] Rudolph, Jenny W. and Repenning, Nelson P., "Disaster Dynamics: Understanding the Role of Quantity in Organizational Collapse," *Administrative Science Quarterly*, Vol. 47, March 2002, pp. 1-30.

[Rundmo et al., 1998] Rundmo, Torbjørn, Hestad, H. and Ulleberg, P., "Organisational factors, safety attitudes and workload among offshore oil personnel", *Safety Science*, Vol. 29, Iss. 2, July 1998, pp. 75-87.

[Russell, 1968] Russell, Bertrand, *Authority and the Individual*, New York, AMS Press 1968.

[Ryan and Oestreich, 1998] Ryan, Kathleen D., and Oestreich, Daniel K., *Driving fear out of the workplace: Creating the high-trust, high-performance organization*, San Francisco, CA, Jossey-Bass, 1998.

[Sagan, 1993] Sagan, Scott D., *The Limits of Safety: Organizations, Accidents, and Nuclear Weapons*, Princeton University Press, Princeton, New Jersey, 1993.

[Sagan, 2004] Sagan, Scott D., "The Problem of Redundancy Problem: Why More Nuclear Security Forces May Produce Less Nuclear Security", Forthcoming, Risk Analysis, 2004.

[Sage and Rouse, 1999] Sage, Andrew P. and Rouse, William B., editors, *Handbook of Systems Engineering and Management*, Wiley-Interscience, New York, 1999.

*[*Saint-Exupéry, 1992*]* Saint-Exupéry, Antoine de, *Wind, sand, and stars*, New York, Harcourt Brace Jovanovich, 1992.

[Sarter and Woods, 1995] Sarter, Nadine D. and Woods, David, "How in the world did I ever get into that mode?: Mode error and awareness in supervisory control", *Human Factors*, Vol. 37, 1995, pp. 5–19.

[SAIC, 1993] Science Applications International Corporation, *Probabilistic Assessment of the Space Shuttle Phase 1: Space Shuttle Catastrophic Failure Frequency Final Report*, New York, 1993.

[SAIC, 1995] Science Applications International Corporation, *Probabilistic Assessment of the Space Shuttle Phase 3: A Study of the Potential of Losing the Vehicle during Nominal Operation*, New York, 1995.

[Saunders et al., 1992] Saunders, D. M., Sheppard, B. H., Knight, V., and Roth, J., "Employees voice to supervisors", *Employee Responsibilities and Rights Jour-*

nal, Vol. 5, 1992, pp. 241-259. Cited in [Milliken et al., 2003].

[Sawacha et al., 1999] Sawacha, E., Naoum, S., and Fong, D., “Factors affecting safety performance on construction sites,” *International Journal of Project Management*, Vol. 17, No. 5., 1999, pp. 309-315.

[Schulman, 1993] Schulman, Paul R., “The negotiated order of organizational reliability”, *Administration and Society*, Vol. 25, No. 3, November 1993, pp. 353-372.

[Senge, 1990] Senge, P. M., The Fifth Discipline: *The Art and Practice of the Learning Organization*, Doubleday Currency, New York, 1990.

[Schein, 1992] Schein, Edgar H., *Organizational Culture and Leadership*, Jossey-Bass, San Francisco, 1992.

[Scott, 1997] Scott, Richard W., *Organizations: Rational, Natural, and Open Systems*, Prentice Hall, Fourth Edition, 1997.

[Shapira, 1994] Shapira, Zur, *Risk Taking: A Managerial Perspective*, Russel Sage Foundation, New York, 1994.

[Shapira, 1997] Shapira, Zur, ed., *Organizational Decision Making*, Cambridge; New York: Cambridge University Press, 1997.

[Shiba et al., 1993] Shiba, S., Graham, A., and Walden, D., *A New American TQM: Four Practical Revolutions*, Productivity Press and the Center for Quality Management, Cambridge, MA, 1993.

[Shrivastava, 1992] Shrivastava, Paul, *Bhopal: anatomy of a crisis*, 2nd ed., P. Chapman, London, 1992.

[Simons, 1995] Simons, Robert L., “Control in an age of empowerment,” Harvard Business Review, March 1995.

[Sitkin, 1992] Sitkin, Sim B., “Learning through Failure: The Strategy of Small Losses”, *Research in Organizational Behavior*, Vol. 14, 1992, pp. 231-266.

[Slovic, 1999] Slovic, Paul, “Trust, Emotion, Sex, Politics, and Science: Surveying the Risk Assessment Battlefield,” Risk Analysis, Vol. 19, No. 4, 1999, pp. 689-701

[Slovic, 2000] Slovic, Paul, *The perception of risk*, Earthscan Publications, Sterling, VA, 2000.

[SMAD, 1999] Wertz, James R., Larson, Wiley J., editors, *Space Mission Analysis and Design*, Third Edition, Microcosm Press, Torrance, CA, 1999.

[Snook, 2000] Snook, Scott A., Friendly Fire: *The Accidental Shootdown of U.S. Black Hawks over Northern Iraq*, Princeton University Press, Princeton, New Jersey, 2000.

[Sorensen, 2002] Sorensen, J.N., “Safety culture: A survey of the state-of-the-art”, *Reliability Engineering and System Safety*, Vol. 76, 2002, pp. 189-204.

[Stein and Kanter, 1993] Stein, Barry A. and Moss Kanter, Rosabeth, “Why good people do bad things: A retrospective on the Hubble fiasco”, *Academy of Management Executive*, Vol. 7 No. 4, 1993, pp. 58-62.

[Sterman, 2000] Sterman, John D., *Business Dynamics: Systems Thinking and Modeling for a Complex World*, Irwin McGraw-Hill, Boston, MA, 2000.

[Sterman, 2002a] Sterman, John D., “System Dynamics: Systems Thinking and Modelling for a Complex World,” *Proceedings of the ESD Internal Symposium*, MIT, Cambridge, MA, May 2002.

[Sterman, 2002b] Sterman, John D., “All Models are Wrong: Reflections on Becoming a Systems Scientist,” *System Dynamics Review*, Vol. 18, No. 4, Winter 2002, pp. 501-531.

[Stirling, 1999] Stirling, A., *On Science and Precaution in the Management of Technological Risks: Final Report of a Project for the EC Forward Studies Unit under Auspices of the ESTO Network*, Report EUR 19056 EN, European Commission, Brussels.

[Sussman, 2002] Sussman, Joseph M., “Collected Views on Complexity in Systems,” *Proceedings of the ESD Internal Symposium*, MIT, Cambridge, MA, May 2002.

[Swain, 1990] Swain, A.D, “Human Reliability Analysis: Need, Status, Trends and Limitations,” *Reliability Engineering and System Safety*, Vol. 29, 1990, pp. 301-313.

[Szenberg, 1992] Szenberg, Michael, ed., *Eminent Economists: Their Life Philosophies*, Cambridge University Press, Cambridge, UK, 1992.

[Tamuz, 1994] Tamuz, M., “Developing Organizational Safety Information Systems.” In Apostolakis, George E., and Wu J.S. (Eds.), *Proceedings of PSAM II*, Vol. 2, Los Angeles, University of California, pp. 71: 7-12.

[Thompson, 1967] Thompson, J., *Organization in action: The social science basis of administrative theory*, New York, McGraw-Hill, 1967.

[Thompson et al., 1998] Thompson, Richard C.; Hilton, Thomas F.; Witt, L. Alan, “Where the safety rubber meets the shop floor: A confirmatory model of manage-

ment influence on workplace safety," *Journal of Safety Research*, Vol 29(1), Spring 1998, pp. 15-24.

[Toft and Reynolds, 1994] Toft, Brian, and Reynolds, Simon, *Learning from Disasters: A Management Approach*, Butterworth-Heinemann Ltd., Oxford, 1994.

[Trist, 1981] Trist, Eric, "The Evolution of Socio-Technical Systems: A Conceptual Framework and an Action Research Program," in eds. A. Van de Ven and W. Joyce, *Perspectives on Organizational Design and Behavior*, Wiley, New York (NY), 1981.

[Tucker and Edmondson, 2003] Tucker, Anita L. and Edmondson, Amy C., "Why Hospitals Don‘t Learn from Failures: Organizational and Psychological Dynamics that Inhibit System Change", *California Management Review*, Vol. 45, No. 2, Winter 2003, pp. 55-72.

[Tucker et al., 2002] Tucker, Anita L., Edmondson, Amy C. and Spear, Steven, "When problem-solving prevents organization learning", *Journal of Organizational Change Management*, Vol. 12, No. 2, 2002, pp. 122-137.

[Tversky and Kahneman, 1974] Tversky, A. and Kahneman, D., "Judgment under uncertainty: Heuristics and biases", *Science*, 185, 1974, pp. 1124-1131.

[Urwick, 1956] Urwick, Lyndall F., "The manager's span of control", *Harvard Business Review*, Vol. 34, 1956, pp. 39 -47.

[Vanderplaats, 2001] Vanderplaats, Garret N., *Numerical optimization techniques for engineering design*, Colorado Springs, CO, Vanderplaats Research and Development, Inc., 2001.

[Vaughan, 1996] Vaughan, Diane, The Challenger launch decision: risky technology, culture, and deviance at NASA, Chicago, University of Chicago Press, 1996.

[Vesely, 2003] Vesely, Bill, *Current Space Shuttle PRA Results*, NASA Shuttle PRA Presentations, 2003. Available online at http://atc.nasa.gov/hosted_events/rmc4/presentations/ Day%201%209-4-03/6%20am%20Vesely.ppt

[Weick, 1987] Weick, Karl E., "Organizational Culture as a Source of High Reliability," *California Management Review*, Winter, 112-117.

[Weick, 1990] Weick, Karl E., "The vulnerable system: an analysis of the Tenerife air disaster", *Journal of Management*, Vol. 16, Iss. 3, September 1990, pp. 571-593.

[Weick, 2004] Weick, Karl E., "Normal Accident Theory as Frame, Link, and Provocation", *Organization and Environment*, Vol. 17, No. 1, March 2004, pp. 27-31.

[Weick and Roberts, 1993] Weick, Karl E. and Roberts, Karlene H., “Collective Mind in Organizations: Heedful Interrelating on Flight Decks”, *Administrative Science Quarterly*, Vol. 38, No. 3, September 1993, pp. 357–381.

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

[Weinberg, 1975] Weinberg, Gerald M., *An Introduction to General Systems Thinking*, Wiley-Interscience, New York, 1975.

[Wenger and Snyder, 2000] Wenger, Etienne C. and Snyder, William M., “Communities of Practice: The Organizational Frontier”, *Harvard Business Review*, Jan./Feb. 2000, Vol. 78, Iss. 1, pp. 139-145.

[West and Clark, 1974] West, B. and Clark, J. A., “Operator interaction with a computer-controlled distillation column”. In Edwards, Elwyn and Lees, Frank P. (Eds.), *The Human Operator in Process Control*, London, Taylor & Francis, 1974, pp. 196-205. Cited in [Perrow, 1999b].

[Wickens and Hollands, 2000] Wickens, Christopher D., Hollands, Justin G., *Engineering Psychology and Human Performance*, Prentice Hall, Upper Saddle River, New Jersey, 2000.

[Wiegmann et al., 2004] Wiegmann Douglas A., Zhang Hui, von Thaden Terry L., Sharma Gunan J., Gibbons Alyssa M., “Safety culture: An integrative review”, *International Journal of Aviation Psychology*, Vol. 14, No. 2, pp. 117-134, 2004.

[Williams, 2001] Williams, M., “In whom we trust: Social group membership as an affective context for trust development”, *Academy of Management Review*, Vol. 26, 2001, pp. 377-396.

[Wilson and Crouch, 2001] Wilson, Richard, Crouch, Edmund A.C., *Risk-Benefit Analysis*, Second Edition, Center for Risk Analysis, Harvard University, August 2001.

[Winsor, 1988] Winsor, Dorothy A., “Communication failures contributing to the Challenger accident: an example for technical communicators”, *IEEE Transactions on Professional Communication*, Vol. 31, Iss. 3, Sept. 1988, pp. 101 - 107.

[Wolstenholme, 2003]] Wolstenholme, Eric F., “Toward the Definition and Use of a Core Set of Archetypal Structures in System Dynamics,” *System Dynamics Review*, Vol. 19, No. 1, Spring 2003, pp. 7-26.

[Woodward, 1965] Woodward, Joan, *Industrial organization: theory and practice*, London, New York, Oxford University Press, 1965.

[Woods, 2003] Woods, David D., “Creating Foresight: How Resilience Engineering Can Transform NASA‘s Approach to Risky Decision Making,” *Testimony on the Future of NASA for the Committee on Commerce, Science, and Transportation, John McCain, Chair*, October 29 2003.

[Woods and Cook, 1999] Woods, David D. and Cook, Richard I., “Perspectives on Human Error: Hindsight Biases and Local Rationality,” In Durso, Nickerson, et al., eds., *Handbook of Applied Cognition*, New York, Wiley, 1999, pp. 141-171.

[Woods and Cook, 2002] Woods, David D. and Cook, Richard I., “Nine Steps to Move Forward from Error”, *Cognition Technology and Work*, Vol. 4, 2002, pp. 137-144.

[Zohar, 1980] Zohar, Dov, “Safety Climate in Industrial Organization: Theoretical and Applied Implications,” *Journal of Applied Psychology*, Volume 65, 1980, pp. 96-102.