

## **INFORMATION TO USERS**

**This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.**

**The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.**

**In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.**

**Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.**

**Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.**

# **UMI**

A Bell & Howell Information Company  
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA  
313:761-4700 800:521-0600



**Order Number 9516884**

**Incorporating human and management factors in probabilistic  
risk analysis**

**Murphy, Dean Michael, Ph.D.**

**Stanford University, 1995**

**U·M·I**

300 N. Zeeb Rd.  
Ann Arbor, MI 48106



**INCORPORATING HUMAN AND MANAGEMENT FACTORS  
IN PROBABILISTIC RISK ANALYSIS**

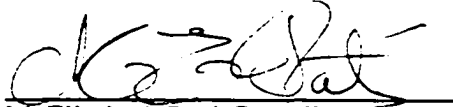
A DISSERTATION  
SUBMITTED TO THE DEPARTMENT OF  
INDUSTRIAL ENGINEERING  
AND ENGINEERING MANAGEMENT  
AND THE COMMITTEE ON GRADUATE STUDIES  
OF STANFORD UNIVERSITY  
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS  
FOR THE DEGREE OF  
DOCTOR OF PHILISOPHY  
IN  
INDUSTRIAL ENGINEERING

Dean Michael Murphy

October 1994

© Copyright by Dean M. Murphy 1994  
All Rights Reserved

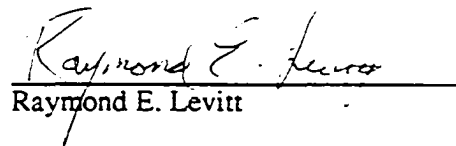
I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

  
M. Elisabeth Paté-Cornell  
(Principal Advisor)

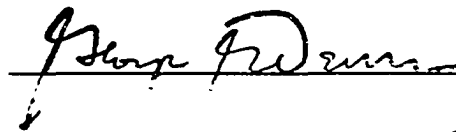
I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

  
James V. Jucker

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

  
Raymond E. Levitt

Approved for the University Committee on Graduate Studies



## ABSTRACT

Complex, engineered systems, such as nuclear power plants, chemical plants, aerospace and marine transport, have the potential for catastrophic failures with disastrous consequences. In recent years, human and management factors have been recognized as a primary cause of major failures in such systems. However, current probabilistic risk analysis (PRA) techniques are unable to handle these effects adequately. This dissertation addresses this problem by extending the PRA methodology with a framework that incorporates human and management effects in a quantitative risk model. The framework provides a structure for incorporating first the actions of individuals that affect the physical system, and then the organizational and management factors that influence those actions. It develops several quantitative models of action that apply to actions in different types of situations, and uses these to make probabilistic predictions of the behavior of individuals in the system. These predictions are made from the perspective of management, and depend on management factors such as incentives, training, policies and procedures, and selection criteria. In this way the framework provides the capability to evaluate how changes in management factors affect the actions of individuals, and thus how they affect system risk. The probabilistic nature of the behavior predictions reflects the limits of information available to management and the inherent uncertainty associated with human behavior. The product of this research is a methodology that can characterize the ways in which management and organizational factors affect system failure risk. This is implemented in a quantitative framework that can evaluate risk management strategies that address management problems. This framework can be used as a tool to "engineer the organization" to increase the safety and reliability of complex technical systems.

To guide the development of this methodology, a preliminary application looked at the risk of general anesthesia for surgery patients. The analysis included the actions of the anesthesiologist and the effects of system management, and evaluated the risk reduction benefits of several proposed management changes. Lessons learned from that project were incorporated in a general risk analysis methodology that is applicable in any domain, and the resulting framework is demonstrated with an illustrative example that deals with the risk of hazardous materials transportation.



## ACKNOWLEDGEMENTS

I would first like to thank my family - my parents, Mike and B.J. Murphy, and my sister Kendra - for the lifetime of support and encouragement that they have offered me. I dedicate this dissertation to them. I also owe a great debt of gratitude to my wife Laura. Her comments and suggestions on the research itself have been valuable, but much more important have been the endless love and support she gives me, and her ability to make life beyond research simply wonderful.

I am grateful to my advisor, Elisabeth Paté-Cornell, for the enduring intellectual challenge and support she offered as this research progressed from fuzzy idea to final dissertation. Ray Levitt and Jim Jucker offered valuable comments and suggestions as readers. Thanks also to Sandy and Lori in the IE office, who have managed to put a very human face on an otherwise featureless bureaucracy; to CC, for all her help and support; and to the Industrial Engineering and Engineering Management Department, School of Engineering, and the Anesthesia Patient Safety Foundation for funding that supported this research. Pete Morris deserves credit for encouraging me to leave a lucrative career for the impoverishment of doctoral studies.

A special thanks to Erik Toomre, the officemate, housemate and loyal friend who has enriched the time in and away from the office with political and personal conversations, foolish mountain climbing adventures, and a dead-heat race to the finish. Also deserving mention are the fellow graduate students, especially Tom Grossman, Michelle Baron, and Linda Lakats, who have offered support and suggestions along the way.

Finally, I would like to thank Wile E. Coyote (without whose Saturday morning inspiration I might never have become an engineer), for instructive demonstrations of the hazards inherent in engineered systems.

## TABLE OF CONTENTS

List of Tables .....	viii
List of Figures .....	ix
<b>Chapter 1: Introduction and Problem Statement .....</b>	<b>1</b>
1.1 Overview of Research .....	1
1.2 The Problem .....	2
1.3 Research Goals .....	4
1.4 Organization of the Dissertation .....	6
<b>Chapter 2: Background and Related Research .....</b>	<b>8</b>
2.1 Probabilistic Risk Analysis: Background .....	8
2.2 Organizational Effects on Risk: Qualitative Research .....	9
2.3 Organizational Effects on Risk: Quantitative Research .....	11
2.4 Other Related Research .....	14
<b>Chapter 3: Preliminary Application - Anesthesia Patient Risk .....</b>	<b>15</b>
3.1 Introduction to the Anesthesia Project .....	15
3.2 The Anesthesia Risk Analysis Project .....	16
3.3 Implications for a General Methodology .....	22
<b>Chapter 4: Structure of the Methodology .....</b>	<b>25</b>
4.1 Structuring Human and Management Effects on Risk .....	25
4.2 Modeling Approach of This Framework .....	29
4.3 Human Error and Taxonomy of Its Causes .....	32
4.4 Modeling Human Action and Error .....	36
4.5 Summary of Chapter 4 .....	38
<b>Chapter 5: Modeling Action .....</b>	<b>40</b>
5.1 Expected Utility Theory and the Rational Model .....	40
5.2 The Bounded Rationality Model .....	50
5.3 The Rule-Based Model .....	60
5.4 The Execution Model .....	67
5.5 Management and Organizational Control Mechanisms .....	80
5.6 Choosing between Models of Action .....	89
5.7 Summary of Chapter 5 .....	93

<b>Chapter 6: Synthesis - Using the Framework to Manage Risk</b> .....	<b>94</b>
6.1 Linking Action to the Physical System .....	94
6.2 Implementing the Framework .....	99
6.3 Hazardous Materials Transport Example - Synthesis .....	103
6.4 Summary of Chapter 6 .....	114
<b>Chapter 7: Conclusion and Future Research</b> .....	<b>116</b>
7.1 Capabilities of the Methodology .....	116
7.2 Limitations and Future Research Directions .....	117
7.3 Conclusion .....	120
<b>References</b> .....	<b>122</b>

## LIST OF TABLES

Table 5.1: Probabilities of driver's choice in the bounded rationality model for two management changes .....	59
Table 5.2: Base case brake repair matrix, R, showing effect of rule-based action on brake condition .....	66
Table 5.3: Brake repair matrix, R, showing effect of inspection training on brake condition .....	67
Table 5.4: Brake repair matrix, R, showing effect of brake replacement policy on brake condition .....	67
Table 5.5: Accident probabilities for four driver types .....	78
Table 6.1: Brake repair matrix for transport example: Base Case .....	105
Table 6.2: Brake wear matrix for transport example: Base Case .....	106
Table 6.3: Base case model inputs for transport example .....	109
Table 6.4: Effects of risk management strategies - intermediate results and overall risk .....	111

## LIST OF FIGURES

Figure 3.1: Dynamics of an anesthesia accident sequence .....	17
Figure 3.2: Generic structure of the Markov Accident Sequence Models .....	20
Figure 3.3: Influence of anesthesiologist state on Accident Sequence Model parameters .....	21
Figure 4.1: Structure of human and management effects on risk .....	26
Figure 4.2: Generalized influence diagram of human and organizational effects on system risk .....	28
Figure 4.3: Taxonomy of error causes .....	33
Figure 5.1: A simple generic decision .....	43
Figure 5.2: Driver's decision tree for decision about driving speed .....	47
Figure 5.3: Probability tree illustrating evaluation sequence for three alternatives in a bounded rationality model .....	54
Figure 5.4: Probability tree of evaluation sequence for driver's decision .....	58
Figure 5.5: Probability tree illustrating uncertainty in a rule-based model .....	63
Figure 5.6: Rule-based model of technician's brake maintenance decision .....	66
Figure 5.7: Illustrative outcome functions for two actor types .....	71
Figure 5.8: Outcome functions for three possible outcomes (one actor type) ...	72
Figure 5.9: Outcome function displaying non-monotonicity .....	74
Figure 5.10: Probability distribution on task demand for the hazardous materials transport example .....	77
Figure 5.11: Outcome functions for four driver types for the hazardous materials transport example .....	78
Figure 5.12: Factors affecting choice of model of action .....	90
Figure 6.1: Subsystem schematic in which action (operator performance) is a failure mode event .....	96
Figure 6.2: Subsystem schematic in which action (maintenance) is an actor influence which affects failure mode event probabilities .....	97
Figure 6.3: Factors affecting accident probability in transport example .....	108

# Chapter 1

## Introduction and Problem Statement

### 1.1 Overview of Research

Complex, engineered systems such as nuclear power plants, chemical plants, aerospace and marine transport, etc., have the potential for catastrophic failures with disastrous consequences. In recent years, it has become increasingly recognized that human and management factors are often at the root of major failures in such systems. However, current probabilistic risk analysis (PRA) techniques are unable to handle these effects adequately.

The purpose of this dissertation is to address this problem: to develop a framework that extends the PRA methodology to incorporate human and management effects in a quantitative risk model, one that can evaluate risk management strategies that address organizational problems. Like current PRA techniques, the new methodology developed here will begin at the level of the physical system. Unlike current techniques, it will then provide a structure for incorporating first the actions of individuals that affect the physical system, and then the system's organization and management, which can affect human action. The final result is a methodology that can characterize the ways in which management and organizational factors affect the risk of system failure. This framework will be useful as a tool to *engineer the organization* for safety and reliability (and can also be used to identify how system design can counter organizational weakness).

An understanding of the ways in which management and organizational factors affect action is a crucial part of an effort such as this, and our knowledge of these processes is certainly far from complete. However, the central purpose of this research is not to develop or test new theories about how psychological and organizational forces affect the behavior of individuals, but rather to develop a tool that can make use of the limited understanding we do have of these effects, in order to improve the management of potentially risky technological systems. In this sense, this work is more "engineering" than it is "science."

The research presented in this dissertation is a part of an ongoing research effort in the Engineering Risk Analysis program at Stanford University, led by Professor M.E. Paté-Cornell. The methodology developed here is an extension of her work to incorporate management factors in quantitative risk analysis.

## **1.2 The Problem**

Technological disasters are nothing new, but as engineered systems grow larger and more complex, their destructive potential increases. Accidents can now affect populations across entire continents and for generations, as the Chernobyl disaster illustrated all too well when it showered nuclear fallout over much of Europe. There are, unfortunately, numerous examples of catastrophic accidents in complex technological systems; some of the more well-known are:

- |                           |   |
|---------------------------|---|
| • the Exxon Valdez: 1989  | Tanker grounding and massive oil spill              |
| • Chernobyl: 1986         | Meltdown of nuclear reactor                         |
| • Challenger: 1986        | Explosion and loss of space shuttle and crew        |
| • Bhopal: 1984            | Chemical leak causes 2,500 deaths; 200,000 injuries |
| • Air Florida: 1982       | Ice on wings causes crash into bridge in D.C.       |
| • Three Mile Island: 1979 | Severe core damage of nuclear reactor               |
| • Tenerife: 1977          | Runway collision of two 747 jets on takeoff         |

Researchers have begun to recognize that technological disasters are frequently the result of human actions that are influenced by management and organizational factors, rather than pure technical problems or isolated instances of human error. Accident reports for all of the system failures listed above, as well as many others, have acknowledged the importance of human and organizational factors. In the nuclear power industry in particular, the effect of management on system risk has been appreciated (e.g., Jacobs and Haber, in press; Wu, et al., 1991). Similar observations have been made in other fields, including marine and aviation accidents (Bea and Moore, 1991; Nagel, 1988), accidents in the oil and chemical industries (Wright, 1986; Bannister, 1988), and accidents in anesthesia delivery (Williamson, et al., 1993; Cooper, et al., 1978). Estimates of the fraction of accidents that are caused by human and/or organizational weaknesses range from 50 to over 90 percent (Senders and Moray, 1991; Williamson and Feyer, 1990; Kletz, 1985; Perrow, 1984). Embrey (1992) observes that

Studies of major accidents from a variety of industries ... indicate that they rarely arise from random failures of hardware as modeled by classical reliability theory. Usually the disaster arises from a combination of active and latent human errors in areas such as design, operations and maintenance.

Reason (1990b) calls this "the age of the organizational accident," where "hazards are seen to arise primarily from the as yet little understood interactions between the social and the technical aspects of the system."

As one example of how human and organizational factors can affect a system's risk, Perrow (1984) cites evidence that nuclear power plant operators sometimes disable

automatic systems and safety devices in order to meet production goals that are set by the organization. In any system, management and human action affect risk in a multitude of ways. Individuals may directly cause or prevent an accident, or influence its likelihood by strengthening or weakening components, changing the load on the system, etc. The organization can control, to some extent, the performance of individuals, through their qualifications and training, the incentives they face, resources at their disposal, etc.

Upon reflection, these observations about the role of human and organizational factors in system failure should not seem particularly surprising. The functioning of any modern, complex system is highly dependent on the actions of individuals in the system; these actions are (or should be) guided by the organization. In even the most highly automated systems, humans are responsible for design, construction, maintenance, and high-level operational control. If the individual and organization are charged with the responsibility for preventing system failure, then almost by definition, when an accident occurs, the cause will be traceable to human action and ultimately organizational factors<sup>1</sup>.

Despite the fact that human action and organizational and management factors do have a significant effect on risk, a persistent and often valid criticism of current risk analysis techniques is that they do not adequately capture these effects (e.g., Jacobs and Haber, in press; Freudenburg, 1992; Reason, 1990b; Apostolakis, et al., 1989; Perrow, 1984). While some nuclear power plant risk analyses have included the possibility of human error (e.g., THERP, Swain and Guttmann, 1983), these studies have focused on ergonomics – how the nature of the physical environment affects human performance – and sidestepped the nebulous issue of management and organizational effects. Current risk analysis techniques treat the effects of management implicitly, if at all. For instance, the effects of system management may be implicitly included in historical data for component failure rates (see Bley, et al., 1992). However, organizational factors can act as a common cause of failure; even if individual component failure rates do reflect organizational influences, the organizational dependencies between them are not modeled explicitly (Davoudian, et al., in press, a)<sup>2</sup>. Furthermore, since these implicit effects

---

<sup>1</sup> A counterargument to this position is that an accident may be the result of a calculated risk that happen to end in an unfortunate outcome. But an examination of technological accidents makes it clear that in the majority of cases, as Wagenaar and Groeneweg (1988, p.42) put it, "Accidents do not occur because people gamble and lose, they occur because people do not believe that the accident that is about to occur is at all possible." Accidents often occur by unforeseen pathways, because the magnitude of risk was misunderstood or denied; that is, because of a failing on the part of the actor and/or the organization.

<sup>2</sup> A common observation about accidents in complex systems is that they seldom result from isolated problems, but are usually caused by the concurrence of multiple problems. Organizational dependencies can make such concurrence much more likely.



cannot be separated out, it is impossible to determine how changes in system management would affect the risk of failure. While current PRA techniques expend extensive resources on the relatively straightforward, well-defined task of characterizing physical component performance, they do not address difficult, confounding issues associated with human and management factors.

The danger of leaving out the "qualitative" effects while modeling the physical system components in great detail is that it can lead to a false sense of accuracy. The very fact that human and organizational actions are unpredictable and difficult to model implies that they may be significant contributors to uncertainty and risk, and are thus essential elements in a comprehensive risk model. Apostolakis, et al., (1989) stated the problem as follows, in the context of nuclear power plant risk:

Thus far, we have worked very hard to pin down the risks stemming from equipment failures, human errors, and natural events such as earthquakes. We have turned the microscope of analysis and quantification on these factors. This has been very important and very good work, but, in another sense, it is the easy part of the problem. Now that we have the tangible parts more or less under control, we are beginning to admit to ourselves something that we have always known, i.e., that the main variables, the major determinants of plant safety, are not valve failure rates and such, but, rather, more amorphous and intangible entities that go by such names as morale, esprit de corps, management attitude, and so on. These factors are not currently included in PSA [probabilistic safety assessment, another name for probabilistic risk analysis], at least not explicitly.

A risk analysis methodology that includes human action and organizational factors would be an invaluable aid in risk assessment and risk management; developing a framework to implement such a methodology is the goal of this dissertation research. Recently, there has been a significant amount of interest in this problem, resulting in both qualitative research and some quantitative approaches; this research is discussed in the following chapter. While this dissertation addresses the same problem, it approaches it from an entirely different perspective.

### **1.3 Research Goals**

Since, as much evidence shows, human and management factors are a fundamental cause of risk, they should be a primary focus of risk reduction measures, and there is a need for a risk analysis methodology that can evaluate such measures, to guide the allocation of risk management resources. In spite of the criticisms of current PRA techniques, the basic approach of modeling the functioning of the physical system is an indispensable way to analyze risk in complex technical systems. A quantitative analysis of the physical

system provides valuable insight into how it can fail; without such an analysis, it is impossible to quantify the risk implications of changes to the system. Because of this, rather than discard the PRA methodology, I use it as a starting point. This research expands PRA's scope to support the formulation of an extended risk analysis model that explicitly includes models of human action and organizational effects. Such a methodology will provide two capabilities that go beyond those of current PRA techniques:

- 1) more accurate risk assessment, because a risk model developed by this new methodology will include more of the fundamental sources of risk; and
- 2) a quantitative risk management tool that guides the use of organizational control mechanisms to reduce risk, by providing a tool that can evaluate proposed organizational risk management strategies.

With difficulty, current PRA techniques might offer the first of these capabilities, by implicitly including the effects of human and management factors in component failure rates, dependencies, etc. What current techniques cannot do is the second – to evaluate organizational risk management strategies. Without expanding the scope of the risk analysis methodology to include human and organizational factors explicitly, it will be impossible to answer questions of the form: "If this organizational risk management strategy is implemented, what will be the effect on system risk?"

The extended PRA methodology developed here is a quantitative aid to "engineering" the organization itself, as well as the physical system that it operates. It can quantify, compare and prioritize risk management strategies, both organizational strategies and the more conventional hardware-based strategies that reinforce the physical system. Strengthening the physical system can be expensive, and quickly leads to diminishing returns. Organizational risk management strategies, in comparison, seem quite promising, judging by the frequency with which human and organizational problems are identified as the root causes of accidents in complex systems. A methodology that can directly compare strategies of the two types will allow risk managers to optimize the use of limited risk management resources.

#### Description of the New Methodology

The purpose of this research is to improve probabilistic risk analysis (PRA) as a tool for managing and reducing risk in real systems. It is an answer to Reason's (1990b) call for a new methodology (he spoke of nuclear power plant risk, but the same applies for other hazardous technologies):

[W]e should develop urgently a better understanding of the origins of organizational accidents and devise a more effective calculus for assessing plant risks. Current probabilistic risk assessment methods are unable to accommodate the organizational component.

This research makes no attempt to reach generalized conclusions about the ways in which management factors affect risk across systems and organizations, just as the current PRA methodology does not make generalizations about the system reliability effects of particular types of components. Indeed, there is little reason to expect, a priori, that such effects would be the same across different types of systems, or even for different systems of the same type. For example, it may be that centralized, authoritarian organizational control reduces risk in some cases, while decentralized management is better in others.

What this research does is to develop a framework for applying an extended risk assessment methodology to a particular system. This framework begins with a functional system model, and expands the analysis to include the human actions and the organizational factors that affect the physical system's performance. This methodology quantifies the risk implications of specific management changes by modeling their effects on human actions, and the effect of actions on physical system performance and thus on risk. This will support an informed allocation of management resources, allowing a comparison of various risk management strategies, and the tradeoffs between risk reduction and other dimensions such as cost, productivity, profit, environmental effects, etc.

In a sense, this approach is a logical extension of the PRA methodology to include a class of "components" (human actors) and the factors that affect their performance (organizational factors) that are currently excluded from the analysis. These human "components" are different from the physical system components for the same reason that they have often been ignored – they are extremely difficult to quantify and to predict. Human beings are not inanimate, unconscious objects whose performance is easy to characterize; they are thinking beings that respond to a wide variety of factors, many beyond the system itself. But since system safety depends critically on their performance, in order to be credible and useful, any risk analysis must take this extra step and characterize the effects of action and the influences of organizational factors.

#### **1.4 Organization of the Dissertation**

The remainder of this dissertation is organized as follows. Chapter 2 is a literature review

that covers other approaches addressing this same problem, and some related work that is applicable in the research here. Chapter 3 presents a brief overview of a preliminary project in this area, an analysis of anesthesia risk, that helped to focus this research. Chapters 4 through 6 develop the framework that expands the scope of the PRA methodology to include human and organizational factors: Chapter 4 lays out the basic structure of the approach; Chapter 5 develops several alternative models of human action and discusses management and organizational control mechanisms that can affect it; and Chapter 6 ties these pieces together, linking the new models to the existing PRA framework, and demonstrating the use of the methodology with an illustrative example. Chapter 7 takes a step back to look at some of the strengths and weaknesses of this approach, and identifies some topics for future research.

## **Chapter 2**

### **Background and Related Research**

#### **2.1 Probabilistic Risk Analysis: Background**

In the several decades since probabilistic risk analysis<sup>3</sup> (PRA) got its start, the methodology has undergone several shifts of focus, with each subsequent stage extending, rather than replacing, the previous (Greenhalgh, 1990). In the wake of the WASH 740 study (Atomic Energy Commission, 1957), which was intended to quell public fear about nuclear reactor safety but had the opposite effect, formal PRA developed out of earlier systems reliability techniques, and was used to analyze the risk of accidents in the nuclear power industry. The PRA methodology combined the quantification of component failure rates with fault- and event-tree analysis to calculate the probability of a reactor accident. Initially, it was concerned primarily with technical failures such as loss of coolant accidents (LOCA) and steam tube failures, which were addressed by adding engineered safeguards – physical reinforcements and redundancies. The WASH 1400 Reactor Safety Study (Nuclear Regulatory Commission, 1975) was the first comprehensive risk assessment for a nuclear power plant, and led the way for a shift of attention to the issue of human reliability and the man-machine interface, which intensified after the Three Mile Island accident in 1979. This second phase focused on behavioral failures (slips and lapses), and cognitive functions such as diagnostic errors, leading to improvements in procedures and in control-room design and technology. Most recently, the Chernobyl accident extended the focus of risk analysis again to recognize the effects of management in what Reason calls the socio-technical era. This places nuclear plant risks in the wider context of risk in any complex, well-defended system. From this perspective, the fundamental effects of organizational and management factors on risk can be seen in a series of disasters – Bhopal, Challenger, Chernobyl – in which management deficiencies were at the root of the actions, errors and technical problems that were the immediate causes of the failures of these systems. Having recognized the problem of organizational effects, however, there is not yet a methodology that is capable of addressing them effectively.

There has been some work on this problem, of course. Several organizational behaviorists have used qualitative methods to look at the effects of organizational and

---

<sup>3</sup> This methodology also goes under the names of probabilistic risk assessment and probabilistic safety assessment (PSA).

management factors on risk. Some of this work focuses on "high-reliability" organizations and their characteristics. A group of researchers at UCLA has begun to address the problem quantitatively in the context of application to nuclear power plants, though their methodology is still in the developmental phase. Paté-Cornell has begun to develop a general methodology based on PRA that can include the effects of management on risk; the work presented in this dissertation is an extension of this research. There is also some other research addressing related questions, such as research in human error and human factors, that is a source of useful ideas for my research. The remainder of this chapter reviews these other research areas.

## **2.2 Organizational Effects on Risk: Qualitative Research**

Quite apart from the technical, system-based modeling approach of PRA, some organizational theorists have studied the ways in which organizational factors can influence risk. Morone and Woodhouse (1986) analyze the management and regulation of hazardous technologies in the U.S., and argue that the organizational strategies employed have averted major catastrophes. A group of researchers based at Berkeley has made several studies of high-reliability organizations. Roberts (1990) and Rochlin, La Porte, and Roberts (1987), look at how nuclear aircraft carriers overcome inherent risks through redundancy, adherence to procedure, and a "self-designing organization" that constantly tunes and corrects potential problems. La Porte and Consolini (1991) and Roberts (1989) discuss the organizational structures and mechanisms that characterize such high reliability organizations. Also a part of the Berkeley team, Weick (1987) identifies a "high reliability organizational culture" that he claims is responsible for the very low failure rates of systems such as nuclear power plants and air traffic control. Perhaps contradicting himself, he analyzed the Tenerife disaster (1990), in which two 747's collided on the runway at takeoff, and identified strong hierarchy and tightly coupled system as key contributors to the accident.

An alternative perspective, taken by Perrow, disagrees with the premise that the failure rates in such systems are low<sup>4</sup>. He argues (1983, 1984) that technological systems, such as nuclear reactors, chemical plants, marine and aeronautical systems, are inherently dangerous because of the unpredictability of human and organizational elements that

---

<sup>4</sup> These positions are not necessarily inconsistent. Weick and Roberts, et al. make only the point that failure rates are low compared to what would be expected in such complex systems, while Perrow bases much of his argument on consequences rather than probability, and argues that these systems pose risks that are high relative to a standard of social acceptability.

control them, and has coined the term "normal accidents" to illustrate his view that catastrophic failures are inevitable. It is the unforgiving nature of complex technological systems that turns routine organizational failures into extreme consequences. "Our ability to organize does not match the inherent hazards of some of our organized activities" (1984, p10). He goes on to argue that many technological disasters are caused by human errors that are "forced" by the design and management of the system. His analysis of the accident at Three Mile Island (1981) serves to illustrate the type of failure that he views as inevitable in such complex systems. Some of the organizational problems that he identifies as significant contributors to risk are insufficient training and supervision, breakdowns in communications, and an emphasis on production that can compromise safety. Sagan (1993) provides an excellent contrast of the "high reliability" and "normal accidents" schools of thought, and in applying them to the problem of nuclear weapons safety in the U.S., comes down firmly and pessimistically on the side of "normal accidents."

In a pair of critical surveys of the field of risk analysis, Freudenburg (1988; 1992) finds that attention to hardware may address only the minority of the causes of technological risks, and social science factors that are often ignored may be more important. He identifies (*ibid.*, 1992) four sets of social factors that contribute to risk, but have received insufficient attention: individual-level human factors, organizational factors (by which he means group effects, including communication problems, diffraction of responsibility, and social pressures), the "atrophy of vigilance" over time, and poor allocation of resources.

In other work, Starbuck and Milliken's (1988) review of the Challenger disaster recognized the role played by the differing responsibilities of engineers and managers, and showed how repeated success could result in reducing safety margins until a serious failure occurs. Wahlstrom and Swaton (1991) develop a qualitative approach for identifying organizational factors that affect risk in nuclear power plants. Heimer (1988) looks at psychological research in risk perception and behavior and proposes how this might affect individual and organizational decision-making involving risk. While most of this qualitative research is not directly focused on quantifying system risk, it provides some insight into how the actions of individuals and organizations affect it.

### **2.3 Organizational Effects on Risk: Quantitative Research**

Two ongoing research efforts are currently working on the problem of bringing management effects into probabilistic risk analysis. For lack of better nomenclature, these are referred to as the "UCLA" approach and the "Stanford" approach, in recognition of their origins. The UCLA approach is being developed by a group at UCLA and Brookhaven National Laboratory, which is working with the USNRC to develop an applied methodology for nuclear power plants. The Stanford approach comes out of the Engineering Risk Analysis group at Stanford University, of which I am a part; this group is developing a general methodology that will be applicable in any domain. These two approaches are discussed below.

#### **The UCLA Approach**

A quantitative approach to the question of organizational effects on risk, described in Wu, et al. (1991), is being developed by a group of researchers based at UCLA and Brookhaven National Laboratory to look at nuclear power plant (NPP) risk, incorporating plant-specific data about organizational and management factors into the plant's probabilistic risk analysis. There are several variations on this methodology, probably reflecting the fact that it is still under development. One of the common features of these approaches is the use of the Nuclear Organization and Management Analysis Concept (NOMAC, Haber, et al., 1988) to characterize the structure of the NPP organization, define fundamental organizational functions, and measure management performance on them. This result is combined with an analysis of the key types of functions within the physical plant (e.g., maintenance quality, calibration of equipment, etc.) These two sets of variables – management performance and plant safety performance – are then related to one another to update the probabilities and dependencies in the work processes and component failure rates of the risk analysis model. This updating is done in one of two ways:

- 1) by subjective assessment (adjustment) of probabilities and dependencies in component failure rates, as suggested in Davoudian, et al. (in press, a, b) using the Work Process Analysis Model (WPAM-I and WPAM-II), or
- 2) by statistically correlating management and plant performance variables, as was reported in Jacobs and Haber (in press) and Haber, et al. (1991). A similar approach is taken by Okrent and Arueti (1990), comparing the performance of "good" and "bad" plants (as rated by the USNRC).

Thus, by updating the inputs to an existing risk analysis model according to measures of management factors, this methodology proposes to predict plant safety performance and overall risk more accurately.



The UCLA approach is promising in some ways, but also has some potential weaknesses. Thus far, this methodology has been used only to look at effects on a few subsystems, and mostly effects during normal times (as opposed to effects during an accident sequence, which may be significant). This is not a criticism of the methodology, but a reflection of its immaturity. There is in principle no reason that it could not be extended, though there may be practical obstacles to overcome. Also, it has been developed for and applied to nuclear power plants only; again, it may be possible to extend the approach to other domains, but it is difficult to draw conclusions at this point.

The UCLA group's published work addresses only the question of risk assessment. In addition to assessing the magnitude of risks, it would be useful to have a tool that could be used for risk management – one that is able to evaluate the effects of a change in management strategy. Presumably, it would be possible to use the UCLA methodology to evaluate a management change, if the effects of the change can be estimated in terms of data for the model. It is not clear how difficult this would be; that would probably depend on the data source used to quantify the link between management performance and plant safety performance – whether it is judgment or statistical data.

One of the problems with relying on statistical data is that it requires a relevant history of the phenomena to be analyzed. That is, in order to estimate the effects of a given management factor, there must be a historical record of that factor's use on which to base a statistical analysis. Observed statistical correlations in a different system, under a different organizational structure, or from a different time period may or may not be applicable to the system under analysis, and it may be difficult to know the difference. Thus it would be difficult to confidently predict the effect of a given management change unless there were significant experience with that management strategy in a comparable system.

A final weakness of the UCLA approach is that it operates at a very high level; while it may incorporate a plant's "safety culture" in the analysis, for instance, it does not necessarily distinguish the elements that make up such a culture. Since it does not model explicitly the ways in which such management factors create their effects (e.g., through incentives, supervision, information, etc.), it would be difficult to use this methodology to choose between these mechanisms for reducing risk. The effects of such mechanisms are included implicitly in this methodology just as overall organizational effects may be implicit in current PRA practice.

### The Stanford Approach

The Stanford group, led by Professor M.E. Paté-Cornell, has looked at the effect of organizational factors on system risk in several domains. Paté-Cornell and Fischbeck (1990) and Paté-Cornell (1989) focus on how organizational factors affect the handling of potential problems with the thermal protection system of the space shuttle orbiter. In Paté-Cornell (1990) and Paté-Cornell and Bea (1992), a taxonomy of error types is used to identify organizational issues (information, incentives, time pressure) that contribute to risks introduced in the design, construction and operation of offshore oil drilling platforms. The most recent application of this line of research is a study of anesthesia risk that laid the groundwork for this dissertation research. The anesthesia risk analysis project is described briefly in the next chapter.

The fundamental approach used in these studies is to develop a functional model of the physical system and then determine how management affects the reliability of elements of that model; in these studies, the effects of management factors were assessed directly utilizing expert judgment. This modeling approach traces the causal mechanisms by which organizational factors affect system performance and risk, by conditioning the performance of the physical system on the decisions and actions of individuals in the system, and then modeling how these decisions and actions depend on management factors. The research in this dissertation extends this basic approach to model the factors that affect risk in greater detail.

Like the UCLA approach, the Stanford approach is new and is still evolving; the research presented in this dissertation is the latest stage in that development. While it has been applied to several systems, the approach cannot yet be called a mature methodology. One of the strengths of this line of research is that it goes beyond measuring and assessing risk, and focuses on risk management – it provides a tool for evaluating and comparing different organizational risk management strategies. This serves as a basis for choosing among them in allocating limited risk management resources.

As with the UCLA approach, data to support the Stanford approach may be difficult to obtain, because such detailed data is not typically collected for systems. Because it models the fundamental causal mechanisms more completely, this approach relies less on statistical data, and may often require expert judgment because of a lack of other data. However, to the extent that statistical data is available, even if it is not at the detailed level of the phenomena being modeled, it may be quite useful in verifying the results of the model.

## **2.4 Other Related Research**

The question of organizational effects on risk falls at the boundaries of a number of fields. Because of this, there is some related research that is not directly focused on the problem addressed here, but that is nonetheless relevant and will be useful in developing this methodology. The work on human error, especially that of Rasmussen and Reason, is of primary importance among these. Rasmussen (1982, 1983) developed the skill-rule-knowledge framework to characterize three modes of task performance and corresponding error types. Reason (1990a, 1987a), builds on Rasmussen's framework to develop the Generic Error Modeling System (GEMS), a model of how individuals use Rasmussen's three modes to solve problems. This research is discussed in greater detail in Chapter 4 where it is used. Some of the more applied work on human error overlaps with the discipline of human factors engineering (ergonomics), which was developed largely to include operator error in risk analyses for nuclear power plants. This work focuses on designing systems for usability by relating task performance to human processes such as physical limitations, perception, cognition, and communication. However, this area of research typically does not explicitly consider management effects.

In addition to the work on human error, there is social science research modeling human action in a number of other fields: economics, behavioral decision theory, psychology, sociology, organizational behavior, etc. Some of this work (particularly expected utility theory, Simon's bounded rationality theory, and Rasmussen's rule-based model) provides a useful starting point for the models that make up the methodology of this dissertation. This research will be discussed in the development of the framework in Chapters 4 and 5.

## **Chapter 3**

### **Preliminary Application – Anesthesia Patient Risk**

#### **3.1 Introduction to the Anesthesia Project**

As a part of the research to incorporate the effects of human and management factors in quantitative risk analysis, I and several colleagues, Professor M. Elisabeth Paté-Cornell and Linda Lakats<sup>5</sup>, and Drs. David Gaba and Steven Howard<sup>6</sup> performed a preliminary project to analyze the risk of general anesthesia to surgery patients<sup>7</sup>. Detailed discussions of the analysis and results of this project are reported in Paté-Cornell, et al. (1994a, b, c). The purpose of this chapter is to give a brief overview of the anesthesia risk analysis project, and to place it in the context of the larger line of research whose goal is to incorporate human and management factors in quantitative risk analysis.

The anesthesia risk analysis project has a dual purpose. The first is to perform a pilot risk analysis of the anesthesia environment, since this is an area in which quantitative risk analysis techniques offer significant promise and have not previously been applied. The second purpose is more general, and is the reason for its discussion here – to serve as a preliminary test of the concept of explicitly integrating human and management factors into a quantitative risk analysis methodology, and to guide the development of a more refined methodology. The methodology and framework that are the primary product of this dissertation research grew out of ideas from the anesthesia project.

Anesthesia is in many ways an ideal application domain for studying the effects of human and management factors on risk, since human action and error – primarily the actions of the anesthesiologist – play a central role in the performance of the system. Section 3.2 presents a brief overview of the anesthesia risk analysis study, and section 3.3 discusses some of the strengths and weaknesses of the approach used in this project and their implications for developing a general methodology that models human and management effects on risk.

It is important to make clear the contributions of each of the project team members in this research. While all members of the team were involved in all phases of the project, I was

---

<sup>5</sup> Industrial Engineering and Engineering Management, Stanford University.

<sup>6</sup> Department of Anesthesia, Palo Alto Veterans Affairs Medical Center, and Stanford University School of Medicine.

<sup>7</sup> This project was funded by the Anesthesia Patient Safety Foundation, whose support is gratefully acknowledged.

primarily responsible for developing the risk model of the anesthesia environment, and collecting the Base Case data. Linda Lakats took the lead in identifying and quantifying the relevant states of the anesthesiologist, and developing, quantifying and evaluating proposed management changes. The overall project was supervised by Professor Paté-Cornell, and Drs. Gaba and Howard were our experts in the field, helping us to understand the anesthesia environment and providing expert judgment as data.

### **3.2 The Anesthesia Risk Analysis Project**

#### **Anesthesia Risk – Background**

The project reviewed here is a preliminary effort to quantify the risk of general anesthesia to surgery patients. This study considered only severe anesthesia accidents (those resulting in irreversible brain damage or death to the patient), and the analysis was restricted to relatively healthy patients in modern, Western hospitals, with anesthesia delivery by a physician anesthesiologist. While the base rate of anesthesia accidents is low (approximately 1 in 10,000), anesthesia accidents are nonetheless disturbing, particularly when the victim is a healthy patient undergoing routine surgery, and anesthesiologists and patients alike want to reduce the risk further. One of the most consistent conclusions from many anesthesia accidents studies is that they are caused primarily by human error on the part of the anesthesiologist (see, for example, Cooper, et al., 1978, 1984; Gaba, 1991; Runciman, et al., 1993a, b; Williamson, et al. 1993). Although some accidents are caused by gross negligence or incompetence, others occur with competent, capable anesthesiologists. These errors (or perhaps more accurately, actions, because not all are unambiguous errors) can be acts such as administering an overdose of an anesthetic drug, failing to detect a signal of a problem, or taking too much time to diagnose and correct a problem. The way to address these human causes of accidents and reduce the risk to patients is through the management of the anesthesia system; management mechanisms such as training, supervision, workload restrictions, etc., can affect the likelihoods of the errors and actions that can lead to an anesthesia accident. To address the question of how management influences risk through the anesthesiologist's actions, this study consists of three parts: 1) the development of a system-level risk analysis model of the anesthesia environment that includes the anesthesiologist's actions; 2) an analysis of the effect of the state of the anesthesiologist on actions, and therefore on risk; and 3) an assessment of how management factors affect the state of the anesthesiologist, and thus ultimately affect risk. Expert judgment is used

to assess the effect of anesthesiologist state on inputs to the risk model (probability of initiating event, time required to detect, diagnose and correct problems) in part 2, and also to assess the effect of management factors on anesthesiologist state in part 3.

The purpose of general anesthesia is to induce anesthesia (numbness), amnesia, and paralysis in a patient to facilitate surgery. While the patient is incapacitated by the anesthetic, the anesthesiologist, using an array of sophisticated equipment, takes over vital bodily functions (most importantly, breathing via a mechanical ventilator) and monitors and controls the patient's vital signs (heart rate, blood pressure and volume, etc.). During this time, there are a number of things that can go wrong that have the potential to cause irreversible harm to the patient. Because of the human body's natural resilience, the occurrence of one of these problems does not immediately result in harm, but causes the patient's condition to begin to deteriorate. An *anesthesia accident* occurs if the problem remains uncorrected for too long, and patient condition deteriorates to the point of irreversible brain damage or death. A crucial element of this dynamic process, as illustrated in Figure 3.1, is the actions of the anesthesiologist, who may initiate a problem in the first place, and is responsible for detecting, diagnosing, and correcting it. The management of the anesthesia system can affect these actions, and in doing so, affect the risk to the patient.

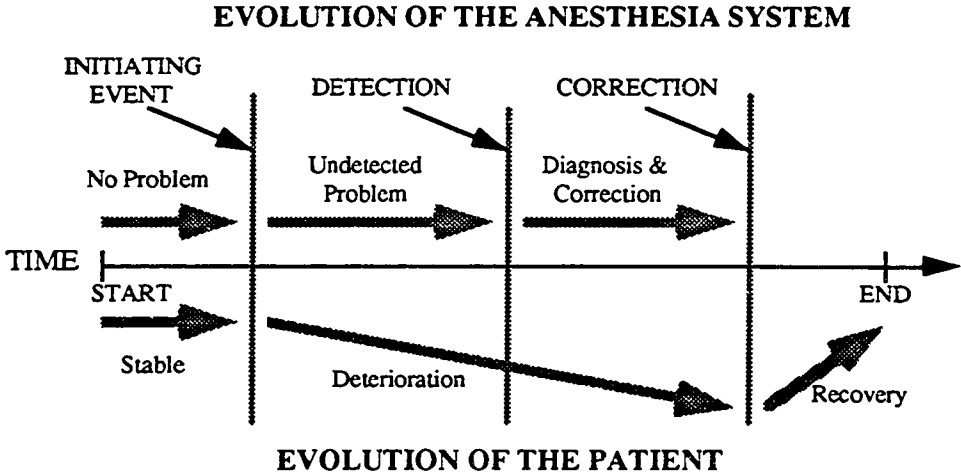


Figure 3.1: Dynamics of an anesthesia accident sequence.

Structure of the Anesthesia Risk Analysis Project

The structure of the three parts of this analysis can be illustrated with a set of equations that express the relationships between the performance of the physical system, the state of

the anesthesiologist, and the organization. The *state of the anesthesiologist* is an intermediate variable that is used to capture the effects of management on the anesthesiologist's actions. The following notation is used:

AA = anesthesia accident  
 IE<sub>i</sub> = the initiating event that begins an accident sequence  
 SA<sub>j</sub> = the state of the anesthesiologist  
 M<sub>k</sub> = the set of management factors

The first part of the study consisted of developing a risk model of the physical system. A number of different initiating events may start an accident sequence (e.g., a disconnect in the breathing circuit or an anesthetic drug overdose). The probabilistic risk analysis (PRA) framework (Henley and Kumamoto, 1981) is used to quantify overall risk as the sum of the products of initiating event and conditional anesthesia accident probabilities:

$$p(\text{AA}) = \sum_i p(\text{IE}_i) p(\text{AA} | \text{IE}_i)$$

The first term,  $p(\text{IE}_i)$ , is based on statistical data, adjusted by expert judgment where necessary; the second,  $p(\text{AA} | \text{IE}_i)$ , is calculated by the stochastic accident sequence model described below. The specific actions that the anesthesiologist takes during an accident sequence – detection of signals and combinations of them, possible diagnoses, and corrective actions – are embedded in this stochastic model, and affect the conditional accident probability.

The second part of the study analyzes how the state of the anesthesiologist affects these actions. Anesthesiologists are categorized into several possible states (e.g., fatigued, unsupervised resident, drug abuser, or none of the above). An anesthesiologist who is in one of the "problem states" may be more likely to cause an initiating event or to fail to detect and correct a problem quickly (e.g., a fatigued anesthesiologist may inadvertently administer a drug overdose, or take longer to detect a drop in blood pressure). Therefore, the state of the anesthesiologist, SA<sub>j</sub>, affects both the probabilities of initiating events and the conditional accident probabilities. Conditioning these inputs on the state of the anesthesiologist, the previous equation can be rewritten as:

$$p(\text{AA}) = \sum_i \sum_j p(\text{SA}_j) p(\text{IE}_i | \text{SA}_j) p(\text{AA} | \text{IE}_i, \text{SA}_j).$$

The third part of the study looks at how management procedures, policies, etc. affect the probabilities that the anesthesiologist is in each of the possible states (the overall

management strategy is designated  $M_k$ ). Management changes decrease patient risk to the extent that they reduce the likelihood that an anesthesiologist is in one of the problem states. For example, work schedule restrictions may reduce the probability that an anesthesiologist is fatigued, or strict supervision guidelines may decrease the chance that a resident anesthesiologist is unsupervised. The risk equation is rewritten once more, conditioning on the set of management factors:

$$p(AA | M_k) = \sum_i \sum_j p(SA_j | M_k) p(IE_i | SA_j) p(AA | IE_i, SA_j).$$

### The Markov Accident Sequence Models

The analysis considered several different initiating events, including breathing system failure (e.g., nonventilation or a disconnect in the breathing circuit), anesthetic drug overdose, and anaphylaxis (a severe allergic reaction). For each of the initiating events considered, a Markov Accident Sequence Model was constructed to characterize the accident sequence dynamics and determine  $p(AA | IE_j)$ , the probability of an anesthesia accident conditional on the initiating event<sup>8</sup>. While the details of each of the Markov models are specific to the particular accident sequence it represents (the states included in each model are those that are relevant to the corresponding accident sequence), they share a common underlying structure. Following the occurrence of an initiating event, several types of events occur in sequence:

- signals of the problem appear, either direct signals of equipment problems or from patient monitors
- the anesthesiologist may detect signals and combinations of signals
- the anesthesiologist diagnoses the problem based on signals detected
- the anesthesiologist takes corrective action(s).

In parallel with these events in the anesthesia system, the patient's condition begins to deteriorate; some of the signals depend on the patient's condition (e.g., a pulse oximeter indicates low blood oxygen). The appropriate corrective action will reverse this deterioration if it has not progressed too far<sup>9</sup>. This series of events is modeled as the interaction of two separate but highly interdependent subsystems: 1) the patient and 2) the anesthesia system. The parallel evolution of these two systems is represented by

<sup>8</sup> Unfortunately, the traditional risk analysis techniques of fault- and event-trees do not do a good job of modeling time-dependent system processes. In order to capture temporal effects, which are crucial in the anesthesia environment, it was necessary to go beyond the conventional risk analysis tools and develop a dynamic model.

<sup>9</sup> For some initiating events, misdiagnoses followed by ineffective solutions are possible. Also, in some cases there are interim actions that do not correct the underlying problem, but "buy time" by slowing the patient's deterioration.



embedded Markov models. For each initiating event, the relevant *states* of the patient (patient condition) and *phases* of the anesthesia system (signals detected, diagnosis, corrective action taken) are identified. At any point in time, the patient and the anesthesia system are each in exactly one state and phase, respectively. The *transition rates* between these states and phases determine the ultimate outcome. Over time, patient state changes at rates that depend on the phase of the anesthesia system, and transitions between system phases depend on patient state. Figure 3.2 illustrates the evolution of patient state and anesthesia system phase for a generic accident sequence. The number and identity of intermediate patient conditions and detections and corrective actions among anesthesia system phases depend on the initiating event being modeled. The shaded patient states (Recovered, Anesthesia Accident) are trapping states from which no further transitions occur.

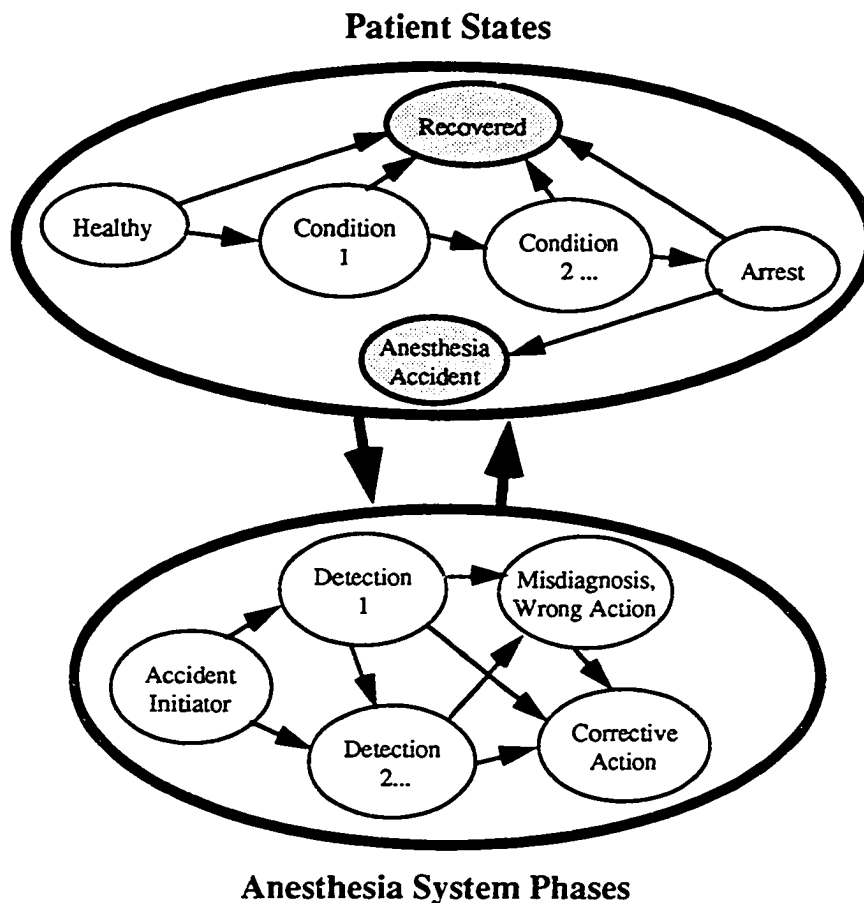


Figure 3.2: Generic structure of the Markov Accident Sequence Models.

### Data and Results

The data for the Markov Accident Sequence Models consists of the transition rates between patient states and anesthesia system phases. The model uses these transition rates to calculate the probability, conditional on the initiating event, of Anesthesia Accident. Also necessary for the overall risk model is data on the initiating event probabilities. All of this data was quantified conditional on each of the anesthesiologist states identified; the effects of anesthesiologist state on the risk model parameters are illustrated in Figure 3.3. By also quantifying the probability distribution on anesthesiologist states under status quo management factors, the overall anesthesia risk model calculates the Base Case anesthesia risk for the existing system. Although the risk differs for each anesthesiologist state, in general, breathing system failures account for about half of all accidents, anaphylaxis accounts for over a third, and drug overdose for about 10%.

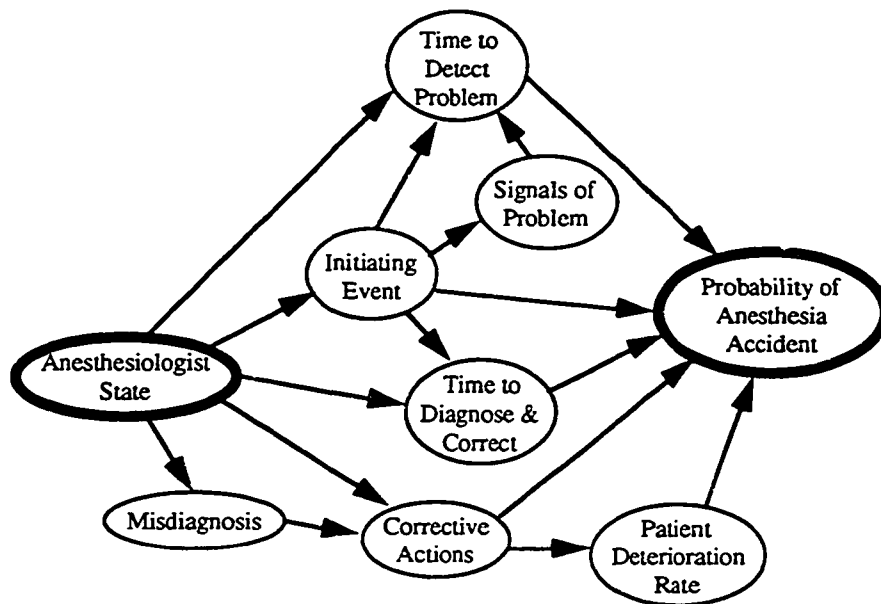


Figure 3.3: Influence of anesthesiologist state on Accident Sequence Model parameters.

A number of alternative management policies that address these issues were suggested by the experts we interviewed, and by analogy with other fields with similar problems, such as airlines<sup>10</sup>. These proposed management changes affect patient risk by decreasing the

<sup>10</sup> Anesthesiologists often view their job as similar to that of an airline pilot: a short, intense, and risky "take-off" (induction of anesthesia) is followed by a long, usually uneventful, and at times boring period of "level flight" (anesthesia maintenance), followed in turn by another short period of intense activity,

incidence of (relatively) high-risk anesthesiologist states. By assessing the effects of these proposed management changes on the distribution of anesthesiologist states, we were able to evaluate their effectiveness in terms of risk reduction.

### Conclusions of the Anesthesia Project

The results of this project suggested that training and recertification to ensure that anesthesiologists are knowledgeable and capable are promising strategies for reducing anesthesia risk. Regular recertification may reduce risk by 30%, and continuing regular training for practicing anesthesiologists reduces it by 15%. Close monitoring and supervision of resident anesthesiologists is also important, and may reduce risk by almost 15%. In spite of the concern they cause in the anesthesia community, drug and alcohol abuse do not seem to be large contributors to patient risk (on the other hand, they do pose serious risks to the abuser, and may be well worth addressing on those grounds). A complete description of the policies evaluated and their risk implications is reported in Paté-Cornell, et al. (1994b).

The results of this project should be considered preliminary. Because statistical data about anesthesia risk is so limited, it was necessary to rely heavily on the judgment of just a few experts. Further, this analysis was limited to healthy patients in large, modern hospitals. Nonetheless, this analysis has offered valuable insights into the problem of anesthesia risk, and demonstrated the value of a quantitative approach for developing effective strategies for improving anesthesia patient safety.

### 3.3 Implications for a General Methodology

The anesthesia project reviewed here demonstrates the essential feasibility of explicitly including human action and management factors in quantitative risk analysis for real systems. In this, it has served its purpose as a preliminary project in this area, and much of the same approach will be useful for other systems. However, the methodology of this project is not entirely generalizable, and it may not be applicable to all other systems and situations.

One reason for this is that the effects analyzed in the anesthesia project are more direct and obvious than they may be in other systems, in part because this project focuses on

---

increased risk, and heightened attention at "landing" (waking the patient). Not only is this analogy useful in understanding how anesthesiologists conceive of their tasks, but we found that some of the organizational risk management policies used by the airlines may also be applicable to anesthesia.

system operation, and not other phases such as design and maintenance. For example, the effect of a failure of the anesthesiologist to diagnose a problem has a clear and direct relationship to an accident. However, in a different system, relevant actions may occur long before a potential accident, and remain as "resident pathogens" that weaken the system and become apparent only when an extreme load causes failure<sup>11</sup>. Poor maintenance may increase component failure probability and the overall risk of system failure, but if the maintenance action is distant in time and space, its relationship to system failure is less obvious. Making this connection explicit will be a necessary part of a general model of human and management effects on risk.

Another reason that the anesthesia approach may not be generalizable is that it focuses on the actions of one individual (the anesthesiologist) in one type of situation (anesthesia delivery in the operating room). This is appropriate for anesthesia, because these actions are by far the largest source of risk and the immediate mechanism that influences it. However, in other systems, there may be multiple individuals whose actions affect risk, and relevant actions may occur in a wide variety of different circumstances. An important limitation of the approach used in the anesthesia project is that it depends on actor type to capture all effects. This approach would not be effective, for instance, to analyze a problem in which actions are influenced by the situation rather than by actor type. The anesthesia study focused on potential problems that could affect the anesthesiologist, but did not address the problems of the "problem-free" anesthesiologist, such as information, production pressure, schedule, incentives, conflict between the surgeon and anesthesiologist, etc. These problems may be a more significant factor in other systems than they are in anesthesia.

One important feature of the anesthesia project that will be retained in the general methodology is that the analysis begins with the physical system and the ways it can fail, and then proceeds to the actions that affect the system, and the management factors that influence those actions. Analyzing the problem in the other direction – looking first at management and action, then the system – would quickly become intractable, and make it impossible to identify and focus on the issues that most affect risk.

The remainder of this dissertation focuses on developing a general methodology that can be used to include the effects of human and management factors in quantitative risk

---

<sup>11</sup> Reason (1990a) uses resident pathogens associated with multiple-cause illnesses in the human body (heart disease, cancer) as a metaphor for latent failures in technological systems. He "emphasizes the significance of causal factors present in the system before the accident sequence actually begins" (p. 197, *ibid.*).

analysis for any system, addressing the issues raised here and others that will come up in the process. The anesthesia project treated human action and management effects as a "black box" – it did not address the question of why actors of different types perform differently, nor of the ways in which management mechanisms change the distribution of actor types. It also did not look at factors beyond actor type that affect action. Instead, it used expert judgment to assess the answers to these questions (the output of the black box) directly. While such an approach can be valuable, particularly in the absence of other data, a general methodology should not be limited in this way. One of the primary goals of the remainder of this dissertation is to open the black box of management and action, and to characterize what is inside.

## **Chapter 4**

### **Structure of the Methodology**

This chapter develops the basic approach and structure of a methodology that will allow human and management effects to be incorporated in a quantitative risk analysis model. Sections 4.1 and 4.2 lay out the basic structure of the relationship between human and management factors and system performance. Section 4.3 develops a taxonomy of the causes of human error which will aid in determining what types of models are needed to capture the relevant effects. From this, section 4.4 develops the structure of a framework to implement this methodology and identifies a set of models that will be included in the framework. These models will be developed in detail in Chapter 5, and Chapter 6 will integrate the pieces of the framework.

#### **4.1 Structuring Human and Management Effects on Risk**

The first step in modeling human and management effects on risk is to develop the structure of these effects in complex systems. I distinguish three levels of structure: the Organization, the Actor, and the Physical System, as illustrated in Figure 4.1. Management factors at the Organization level affect human performance and error at the Actor level, which in turn affects failure mode events at the Physical System level. Failure mode events are the basic events within a physical system that, in certain combinations, can cause system failure. These are the events that show up in current risk analysis models, such as the failure of an emergency cooling water pump in a nuclear power plant. In some cases, human action that is a direct part of system failure can constitute a failure mode event<sup>12</sup>. Once they are identified, current risk analysis techniques can successfully include these types of errors, because they are treated just like any other component failure. However, simply recognizing errors at this level may not help to prevent them, because as Perrow (1984) argues, human errors may be "forced" by organizational and management factors that are the actual root cause of the problem. And many of the human actions that contribute to system failure are not failure mode

---

<sup>12</sup> Such actions are often identified as "operator error". Well-known examples of system failure caused by such actions are the Exxon Valdez, where an inexperienced pilot navigated the tanker onto a reef, and the Union Carbide disaster in Bhopal, India, in which safety systems that might have prevented the release of methyl isocyanate gas had been shut down.

events at all, which makes it necessary to distinguish the second level of this structure, human performance at the level of actors (individuals) in the system.

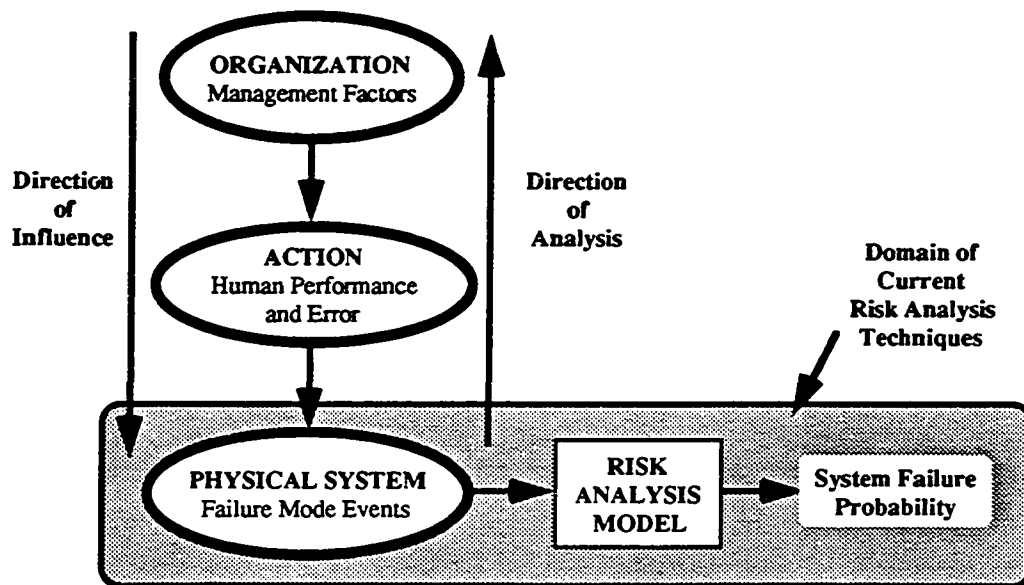


Figure 4.1: Structure of human and management effects on risk.

Human performance and error at the Actor level is not a direct part of system failure, and does not show up directly in current risk analysis models. Rather, actions at this level affect the likelihoods of and dependencies between failure mode events, through mechanisms such as system design, construction, operation and maintenance, or detection, diagnosis, and correction of problems. A failure of human performance at this level is neither necessary nor sufficient for the occurrence of a failure mode event or system failure, but may have a significant effect on their likelihood. Examples of major system failures in which human performance played a key role are the Chernobyl nuclear reactor, which was being operated in an unstable region when it failed, and the Space Shuttle Challenger, where the booster design increased the likelihood of critical component failure. This framework will create tools that can extend a risk analysis model to condition failure mode events on human performance at the Actor level. Of course, as with any variable that affects the probabilities of failure mode events in a risk analysis model, it is necessary to organize these actions into classes that are mutually exclusive and exhaustive (exactly one will occur).

At the top of Figure 4.1 are management factors at the Organization level that can affect the actions and performance of individuals in the system. These factors are the "control knobs" that management can use to affect human performance, including such mechanisms as incentives, training, selection and screening, policy and procedure, organizational structure and culture, etc. The effect of the organization may be to influence the state of the actor, as in the anesthesia project (e.g., fatigued, inexperienced, poorly trained, etc.), or it may affect the actor's situation (e.g., incentives, information, procedure, etc.). Of course, even with these control mechanisms at its disposal, management is not able to exert complete control over the actions of individuals, but this framework will help to capture its influence. In many major technological disasters, such as the Challenger, Bhopal, Piper Alpha, and the Exxon Valdez, it is possible to trace the root causes to management factors at the organization level that affect the performance of individuals in the system, which in turn cause or contribute to the failure of the physical system itself. In some cases, these management control mechanisms may inadvertently induce risky behavior, simply bringing these effects to light and eliminating them may cause significant improvements. More importantly, management may be able to reduce risk by employing these control knobs proactively to induce desired behavior and reduce the likelihood of detrimental actions. The types of organizational and management factors that are often involved, and the types of actions that can influence system risk, are illustrated in a generalized influence diagram of Figure 4.2. Of course, this diagram does not include all possible organizational and management factors, nor all possible types of actions, and the particular relationships between organizational, human and system factors depend on the system, but this figure gives a broad overview of the types of factors that are considered in this framework.

The structure of human and management effects on risk can be described by mathematically relating these events to the failure probability. Considering first just the physical system itself, the probability of system failure is the sum over the initiating events that could lead to failure,  $IE_i$ , of the failure probability conditional on the initiating event times the initiating event probability:

$$p(F) = \sum_i p(F | IE_i) p(IE_i)$$

As was argued above, events at the level of the physical system are influenced by the decisions and actions of individuals within the system. Conditioning the previous equation on decisions and actions,  $DA_j$ :

$$p(F) = \sum_i \sum_j p(F | IE_i, DA_j) p(IE_i | DA_j) p(DA_j)$$



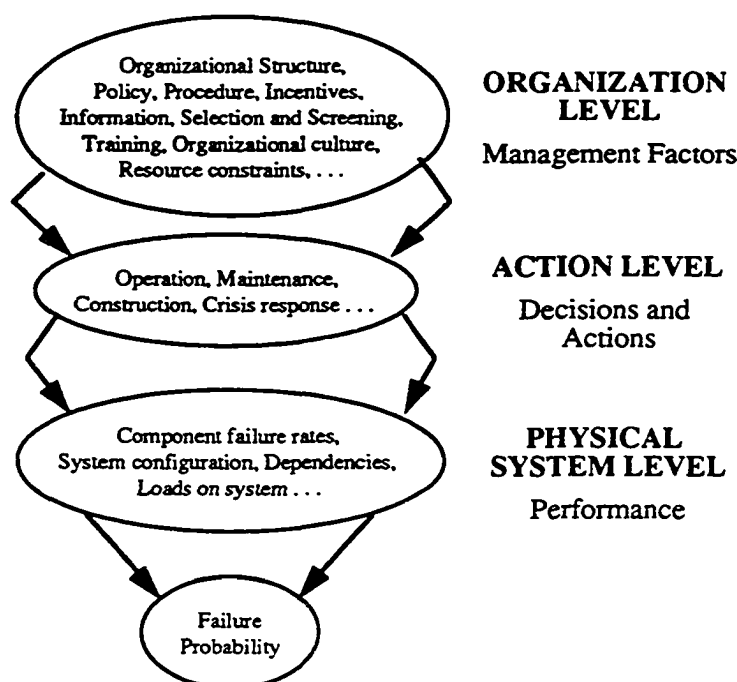


Figure 4.2: Generalized influence diagram of human and organizational effects on system risk.

Note that decisions and actions can affect both the probabilities of initiating events and the conditional failure probabilities, because individuals may cause or contribute to the occurrence of initiating events, and their actions in an accident sequence may also have a major effect on the outcome of the accident sequence. Calculating the conditional failure probability will often require a detailed model of the accident sequence. This second equation does capture the influence of the behavior of individuals on the physical system, but in order to determine the probabilities of the relevant decisions and actions, it is necessary to consider the third level, the effect of organizational and management factors on behavior. Conditioning on management factors,  $M_k$ :

$$p(F | M_k) = \sum_i \sum_j p(F | IE_i, DA_j) p(IE_i | DA_j) p(DA_j | M_k)$$

Note that the failure probability is conditioned on organizational factors, rather than averaging over possible factors, because this is a decision variable for the organization, not an uncertainty. The organization decides what incentives to offer, what procedures to establish, what selection criteria to use; these are not uncertainties beyond its control. In those cases where organizational strategy affects action entirely through its influence on the state of the actor, as was modeled in the anesthesia project, the previous equation

becomes:

$$p(F | M_k) = \sum_i \sum_j \sum_l p(F | IE_i, DA_j) p(IE_i | DA_j) p(DA_j | AS_l) p(AS_l | M_k)$$

If more than one different type or degree of system failure is possible, a set of equations of this form can be written for each.

In this framework, I will develop descriptive behavior models that can be used to capture the ways in which human performance depends on management factors, and methods to characterize the effects of human performance on the physical system. Once these links between management control mechanisms and physical system performance are developed, they can be used in conjunction with current risk analysis techniques to determine management effects on the overall risk of system failure. (For comparison, current risk analysis techniques, whose domain is encompassed by the shaded region at the bottom of Figure 4.1, include the effects of human performance and management factors only implicitly, if at all.) While the direction of influence in this structure is from the top down, from Organization to Actor to Physical System, the analysis will proceed in the opposite direction, from Physical System to Actor to Organization. The framework begins with a functional analysis of the physical system, identifying the ways in which it can fail, proceeds to the actions of individuals that can affect physical system performance, and finally identifies the management control mechanisms that can affect the actions of individuals. This bottom-up analysis will be used to develop top-down recommendations for management strategies that reduce system risk.

#### **4.2 Modeling Approach of This Framework**

The heart of this framework consists of models that describe the actions of individuals and the ways in which management can affect these actions. These models are used to predict the likelihood that an individual will take an action that has the potential to cause or contribute to failure of the system – inappropriate action, action taken at the wrong time, necessary action omitted, etc. It is natural to refer to such actions as error, or human error, but this may be a misnomer. Not all actions that contribute to system failure are necessarily errors, and even the designation of an action as "error" may be subjective. Particularly in a complex system, there are grey areas in which the distinction between "error" and "appropriate action" is unclear, as when goals conflict or the actor lacks information. Errors are a subset of human behavior, and it is often more fruitful to

model behavior in general than to try to draw what are sometimes arbitrary distinctions between errors and other behavior. As Rasmussen (1990, p.1198) makes this point,

Work in modern high-tech societies calls for a reconsideration of human error: research should be focused on a general understanding of human behavior and social interaction in cognitive terms in complex, dynamic environments, not on fragments of behavior called 'error'.

To avoid these problems, and to allow this framework to handle human and management effects that may not be unambiguous errors, this research concentrates on the factors that affect behavior and action more generally, and then looks at the ways in which these factors can lead to actions that may cause system failure. When the term "error" is used, it is not necessarily to place blame on the actor. In fact, the primary theme of this research is largely the opposite; that management problems are often the root causes of the errors of individuals.

In any complex system, there are many different situations in which individuals must take action. Different situations require qualitatively different types of actions, such as decision making, diagnosis of the situation, executing planned action, etc. Individuals act in different "modes" in these different types of situations, which is to say that the processes that determine action are different in these different types of situations. I will identify the modes in which individuals act, and develop models describing the processes that drive action in each of them. This framework will provide a set of models of action that serve as a "tool kit," from which the appropriate model can be selected to describe an individual's action (and potential error) in a given situation. There are typically many points in any complex system, or even within a single accident sequence, at which human action plays an important role and where this framework can be applied. At the end of the next chapter, I will develop some guidelines for which of these models may be most appropriate under what circumstances.

Unfortunately for risk managers trying to understand and control the behavior of individuals in a complex system, it is extremely difficult to characterize and predict the actions of human beings. This is evidenced by the fact that entire fields of social science research have been dedicated to modeling and predicting human behavior, with very limited success, at best. Human behavior is an extremely complex phenomenon that does not lend itself to mechanistic prediction. Recognizing this, I do not try to construct a grand model that can accurately predict human behavior – my attempt would certainly fare no better than previous efforts. Instead, I will recognize the limitations of our knowledge and the uncertainty this implies for predicting action, and treat the problem as

one of probabilistic inference from the perspective of system management. It is impossible with our current understanding of human action to develop a precise characterization of the modes in which individuals act, or of behavior within a given mode. However, without claiming to ascertain the ultimate scientific truth about how human beings behave (which may never be possible, and is certainly a long way off), we can go a long way toward characterizing our understanding of and uncertainty about it. The goal is to use the information that is available to make the best probabilistic prediction of behavior, not to predict a particular individual's behavior in a specific situation. This much can be very useful information in managing risk in a technological system, and will allow us to do a more complete and accurate job than is generally accomplished using current risk analysis techniques. In this sense, this framework is an *engineering rather than a scientific solution to the problem*. It addresses the question of how to manage the risk of real systems when faced with an incomplete understanding of the human behavior that is a key element in the functioning of these systems.

While it is often impossible to say with confidence how an individual will behave in a particular situation, there is certainly a wide body of knowledge, both specific to particular situations and more general, that may be relevant to the question. This information can be characterized probabilistically to reflect the extent and the limits of our understanding. It is the very incompleteness of our knowledge that makes probabilistic techniques so valuable in modeling action – they allow us to make good use of the knowledge that is available and yet still recognize and characterize the significant uncertainty that remains. I will model the different modes of action by drawing in part from existing behavior models developed in fields such as economics, psychology, behavioral decision theory, ergonomics, etc. These models will be adapted for use within this framework, and I will develop new methods where existing models are inadequate or unavailable. The parameters of these models will be factors that are under management control, such as incentives, information, selection and screening, etc. These models will allow a risk manager to make probabilistic predictions of action as a function of management controls, and ultimately, to evaluate the effects of these management controls on system risk.

Once the appropriate model of action has been selected for a given situation, the risk implications of a proposed change in management strategy can be estimated as follows. A change in management strategy modifies the factors that affect action – information, incentives, the actor's abilities, etc. The effects on action are captured by applying the appropriate model of action in a before-and-after analysis of the proposed management

change. By characterizing the effects of each possible action on system component performance, and using these results with a current risk analysis techniques for the system with and without the proposed change, the risk implications of the management strategy change can be estimated.

### **4.3 Human Error and Taxonomy of Its Causes**

Research on human error spans several fields, including psychology, ergonomics and human factors, and engineering, and a number of definitions and taxonomies of error have been proposed. Errors have been classified in many ways – as individual, management, or organizational errors, active or latent errors, errors of commission or omission, slips or mistakes, etc. A perspective similar to mine is that of Embrey (1992), who builds on Reason's distinction between active and latent errors to define management or policy errors as one category of organizational errors, which are themselves a type of latent errors. From this perspective, a management error is a policy, incentive, or other management action that "creates conditions which induce active human errors," (ibid., p. 199). I go a step further – rather than seeing management as just one of several types or causes of error, I see it as a fundamental influence that can potentially affect all human errors in complex systems. The management of a system does not affect the physical system directly, but through the individuals who design, construct, operate and maintain the system. Its effect on the actions of individuals within a system is pervasive – management (perhaps inadvertently) causes some human errors, and can prevent others. While it is impossible to eliminate human error entirely, management does have a fundamental effect on the actions and errors of individuals.

A classification scheme for human error will be helpful in developing this framework. While several other researchers have developed taxonomies of human error, (e.g., Norman, 1981; Paté-Cornell, 1990; Rasmussen, 1982, 1983; Reason, 1987b, 1987c; see a discussion of taxonomies in Senders and Moray, 1991), I take a somewhat different approach. Rather than categorizing error per se, I classify the factors that cause error. These are the same factors that affect action in general; it is problems with these factors that can lead to error. I take this approach because attention to the causes of error gives insight into the processes and parameters that drive action, and by identifying and logically separating these out, the sources of the problem can be addressed systematically. An understanding of the causes of error helps to determine what to do to prevent errors.

Figure 4.3 illustrates a taxonomy of the causes of errors. It is described below with a brief explanation of each of the elements and illustrative examples.

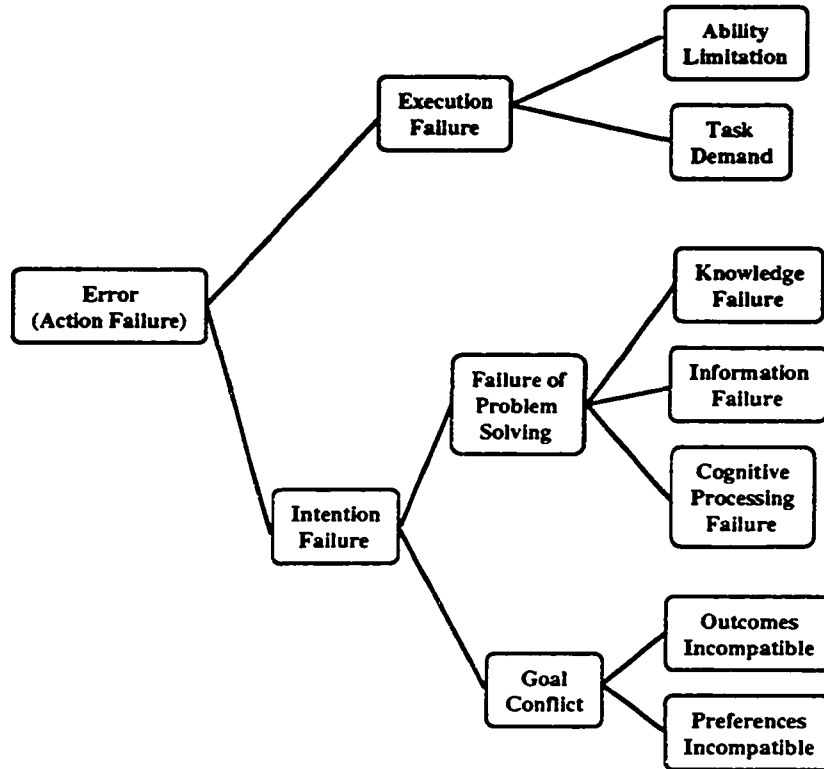


Figure 4.3: Taxonomy of causes of human error.

Action can be thought of as a two-step process consisting of first forming an intention that defines the desired action, and then executing that intention. This process can fail and lead to error at either step. This two-step process corresponds to the distinction between *slips* and *mistakes* that is fundamental in the human error literature (Norman, 1981). Execution Failures cause slips (and related lapses), which are actions that do not go according to plan. Intention Failures, problems in the process of intention formation, cause mistakes, where the plan of action is inadequate to achieve the desired outcome. As used here, intention is defined as a specific plan to take a well-defined action. In particular, values, preferences and goals are not intentions, though they are often involved in intention formation. Execution refers only to the capacity to carry out an intention once it is developed.

An Execution Failure can be caused by the limitations of the actor (Ability Limitations), as in the case of an operator who invokes steps in the wrong order because of a memory

lapse, or an inexperienced surgeon with an unsteady hand. Errors caused by Ability Limitations may be a result of the actor's fundamental inability to perform the necessary action, or may reflect a situation in which the actor has the basic capability to perform the intended action, but not the ability to do so reliably, and on a particular occasion makes an error. The latter is an example of the classic "slip" – a slip of the hand or mental lapse.

On the other hand, the configuration of the physical system may make it difficult or impossible to execute the appropriate action (Task Demand)<sup>13</sup>, as in a system whose controls are difficult to manipulate or distinguish. Of course, the distinction between Ability Limitations and Task Demand may be unclear – it is ultimately a mismatch between the requirements of the system and the actor's abilities that creates a problem, so it is the relative Ability Limitations and Task Demands that are relevant.

The other primary class of error causes is Intention Failure. As Reason (1990a, p9) points out in discussing mistakes, failures of intention are more subtle, more complex, and more difficult to detect and understand than failures of execution, and as a result may pose a greater hazard. One reason that an individual may develop an inappropriate plan of action is because of a Failure of Problem Solving, the classic "mistake," where for one of several possible reasons a deficient plan of action is developed. This can take the form of an Information Failure (lack of or incorrect information), or a Knowledge Failure (lack of or incorrect knowledge). I distinguish Information from Knowledge as follows: Information refers to the actors understanding of the current status of the system, environment, etc., and changes over time as the state of the system changes. Knowledge is more stable than information, and can include a number of different things, such as an understanding of how the system functions, knowledge of the alternatives that are available, and understanding of the outcomes that will result from a given action. An example of information is knowing that the pressure inside a steam line is currently 1,200 psi and increasing; knowledge is understanding that the line is likely to fail at pressures beyond 1,000 psi, and knowing what steps to take to reduce the pressure. Of course, both information and knowledge may be and often are associated with uncertainty, and most situations require that the actor have both accurate information and knowledge in order to take appropriate action to maintain system safety. The distinction between the two is important because the mechanisms for addressing them may be quite different.

---

<sup>13</sup> There may be legitimate dispute about applying the term "error" to such a situation. The point is that an excessively demanding task can cause an individual to fail in executing the intended action. In this sense, the action is an "error", though the true error may have occurred in the design or construction of the system.

The final way in which problem solving can fail is by a failure of cognitive processing of the available knowledge, information and alternatives (Cognitive Processing Failure). This could result from the application of flawed problem-solving methods, a cognitive processing error, or a problem whose size and complexity exceed the actor's cognitive capacity<sup>14</sup>. These are all cases where the actor has the sufficient information, knowledge and understanding to come up with the appropriate action, but fails to process the elements of the problem correctly.

The second type of causes of Intention Failure is Goal Conflict, which can occur when the individual goals that an actor pursues are in conflict with the organization's goals. This distinction between the goals of the actor and those of the organization, often overlooked in the risk analysis and human error literatures, is crucial to recognizing goal conflict. Several organizational behaviorists (e.g., Argyris, 1964; Thompson, 1967; Ouchi, 1979) identify goal congruence, the internalization of an organization's goals by the individuals within it, as an effective mechanism for controlling individuals' behavior. This literature usually implies that goal congruence occurs when actors adopt the organization's goals as their own, in spite of the fact that in doing so, they sacrifice their own self-interest for organizational ends. In contrast, I look at goal congruence and conflict from a somewhat more rationalistic perspective as the compatibility between individual and organizational outcomes and preferences<sup>15</sup>. If actions that lead to good organizational outcomes also lead to good outcomes for the individual, then goals do not conflict, but if actions that are good for the individual lead to risky or poor organizational outcomes, then there is goal conflict.

It is important to recognize that an "error" resulting from goal conflict is not an error at all from the perspective of the actor<sup>16</sup>. Goal Conflict may be caused by either the fact

---

<sup>14</sup> Note that the cognitive capabilities involved in forming proper intentions for action are considered here as part of intention formation, and not as part of the Ability Limitations discussed above that apply only to the execution of intentions once they are formed.

<sup>15</sup> There are certainly instances in which individuals pursue organizational goals that are contrary to their own self interest; this is common in strong religious and social organizations. However, for complex engineered systems that typically cannot demand the same sort of loyalty, it is far more prudent to make the individual's self interest compatible with organizational goals than to depend on actors' self-sacrifice.

<sup>16</sup> Here, I use the term error somewhat more broadly than others. Goal conflict is often overlooked, probably because it does not fit the classic description of an error. For example, Reason (1987c) characterizes mistakes as stemming from limited information processing and reasoning ability, but does not consider the possibility of rational, self-interested action that is inconsistent with organizational goals. While many researchers look only within the actor for the causes of error, this broader perspective enables me to include forces outside the actor, such as flawed incentives, that can encourage actions that may cause or contribute to system failure. (The same is true for the information failures discussed above, where the actor is not necessarily responsible for the information failures that may cause him to make an "error").



that the individual and the organization have Incompatible Outcomes (e.g., the individual is rewarded for behavior that creates risk), or because they have Incompatible Preferences (an individual's risk preference conflicts with the organization's, or the individual cares about different goals than does the organization). The problem of managing when there is a goal conflict is similar in many ways to the principal-agent problem from the economic literature, and the solution will be qualitatively similar – to adjust outcomes (and possibly preferences, though this is not addressed in the principal-agent literature) until the actor's self-interested behavior is consistent with the organization's goals. (This is discussed at greater length in the next chapter.)

This taxonomy logically separates the factors that are the immediate causes of errors. Any actual error may be the result of a combination of these causes. The root causes of error go much deeper than these, and are generally factors over which management may have some control. The taxonomy gives insight into the types of models of action that are appropriate; these will link action to the management factors that can reduce the likelihood of errors.

#### **4.4 Modeling Human Action and Error**

The taxonomy developed above makes clear the fundamental and distinct roles played by intention and ability in determining action and as potential sources of error. In order to develop the appropriate intention (appropriate from the organization's perspective), the individual must pursue goals that are compatible with the organization's, and must not make a mistake in doing so<sup>17</sup>. If an actor develops the intention to perform the correct action, then preventing an error requires only its proper execution, which depends on the match between the requirements of the system and the actor's ability.

While the taxonomy is useful in describing the causes of error and pointing out ways it may be reduced, by itself, it does not identify what types of models of action are appropriate. The work of Rasmussen and Reason is useful here. Building on Norman's slip-mistake dichotomy, Rasmussen (1983) has developed the Skill-Rule-Knowledge framework to describe human performance with three modes of action. Action in the Skill-based mode follows stored, preprogrammed instructions that make up a "script"

---

<sup>17</sup> I am assuming here that the organization's goals are clear. While that may not always be the case, this research does not deal with vague or conflicting goals at the organizational level; it assumes that the organization knows what it wants, at least in terms of system reliability and safety, and offers a way to help the organization achieve its goals.

(Schank and Abelson, 1977), a predetermined sequence of actions that proceeds without cognitive intervention. In the Rule-based mode, action is governed by stored rules that specify the appropriate action for given situations. The Knowledge-based mode describes an actual decision process in which the actor creates or selects an plan of action based on conscious analytical processes, using knowledge of the system to explicitly consider the effects of action and their desirability<sup>18</sup>. Reason (1990a) draws on this to create the Generic Error Modeling System (GEMS), which says that individuals operate at the lowest level possible, and move up to higher levels if the current level does not offer a solution (first Skill, then Rule, then Knowledge).

Skill-based errors, or slips, correspond loosely to execution failures in the taxonomy (slips are actually a special case of execution failures); an execution model is developed to characterize them. A rule-based model is developed to describe rule-based action and mistakes. To model action in the Knowledge-based mode, the rational actor model (expected utility maximization), is the obvious candidate. It is a useful model that applies to many situations, and a rational actor model will be developed for use with this framework. However, it would be a mistake (no pun intended) to use expected utility maximization as the only model of Knowledge-based behavior – this would imply that actors always make decisions rationally. Simon's bounded rationality theory recognizes that truly rational behavior is beyond the capacity of human beings in many real decision situations, and that actors often employ heuristics to simplify problems and approximate rationality. In addition to the expected utility model, then, a model based on bounded rationality is used to characterize knowledge-based decision making and mistakes. Thus, four models of action are developed for use in this framework:

- 1) an Expected Utility Model
- 2) a Bounded Rationality Model
- 3) a Rule-based Model
- 4) an Execution Model.

Together, these serve as a "tool kit" from which the appropriate model can be selected to describe action in a particular situation (the choice of which model to use in a given situation is discussed in section 5.6). The first three describe different ways in which an individual can develop an intention to take a particular action; the fourth characterizes the

---

<sup>18</sup> The same type of knowledge that is used in the knowledge-based mode is presumed to be encoded in the rules of rule-based behavior, and in the programmed instructions of skill-based behavior, so that these other modes just provide shortcuts that specify the same action as would a complete knowledge-based analysis of the situation. In reality, this may not be the case, and this is one of the ways in which rule-based and skill-based action can break down and lead to errors. This will be discussed in detail in the respective modeling sections in the next chapter.

execution of that intention. In principle, any action in any situation can be thought of as a process of in which the actor first forms an intention for which action to take, and then executes that intention. In fact, for some actions it may be appropriate to construct one model to describe intention formation (using the rational, bounded rationality, or rule-based model), and also to develop an execution model to depict the execution of the intention(s) thus formed. However, for most situations it will probably not be necessary to model both parts of this process explicitly. It is likely that only one part of the process will be susceptible to problems that can lead to an error that could contribute to system failure, and the other part can be assumed to proceed without difficulty.

In this framework, the purpose of these four models is to represent, from the perspective of management, the best available knowledge about the actions of individuals and the forces that affect them. It is important to distinguish knowledge about a process from the process itself. A model of coin tosses that characterizes outcomes as independent Boolean random variables with equal probabilities of heads or tails has little to do with the geometry of the coin and the physics of gravity, mass and inertia that actually determine the outcome, but it serves well to characterize our knowledge of that outcome. Similarly, the purpose of these models of action is not necessarily to represent the actual processes that drive action (though to the extent they do that, they may give more accurate results). These models characterize management's information and understanding of action, so that risk management strategies can be based on the best available information and beliefs. Further, while these four models were chosen because they seem to be the most general and most widely useful models, these are not the only models of action available. Some alternative models that might also be used to characterize action are discussed briefly in the following chapter. If, in a particular situation, a different model of action would provide better results, then it can be used in place of one of these models, but it would be implemented in framework in the same way as these models (expressing the probabilities of various possible actions in terms of parameters that characterize the situation, some of which management may influence), and the rest of the framework would be unaffected.

#### **4.5 Summary of Chapter 4**

In this chapter, I have laid out the basic modeling approach and structure of the framework using a taxonomy of the causes of error as an aid. However, while the study of error can offer useful insight, the framework is not structured rigidly around the

concept of error, and it will be able to treat actions that are not normally considered error. The framework will develop explicit, quantitative models linking management and organizational factors to actors' behavior, and then model the relationship between this behavior and possible system failure. Models of action and how it is influenced by management factors make up the heart of the framework. Four different models of action that apply in different situations will be developed: an expected utility model, a bounded rationality model, a rule-based model, and an execution model. Chapter 5 will quantify these four models of action so that they can be used in the risk analysis framework, and will also discuss some of the mechanisms by which management can influence action. Chapter 6 will develop and tie together the remaining elements of the framework, and discuss issues that arise in implementing it.

## **Chapter 5**

### **Modeling Action**

In sections 5.1 to 5.4 of this chapter, I develop and quantify the four models of action selected in the last chapter – the expected utility model, the bounded rationality model, the rule-based model, and the execution model. Section 5.5 is a discussion of the mechanisms by which management can influence the behavior of individuals, and section 5.6 gives guidance for choosing between the four models of action in a particular situation.

To illustrate concepts and models as they are developed, I will use examples related to a simple, illustrative system – a trucking firm that transports hazardous materials. The primary risk concerning the firm is the possibility that an accident in transit will release the hazardous material, risking exposure and health hazard to nearby residents, as well as environmental damage. Since human actions and errors are primary contributors to the likelihood of an accident in a system such as this, the firm would like to develop a way to characterize this risk and develop and evaluate risk management strategies that address it. Much attention will focus on the drivers of the trucks, because they are the individuals most immediately in control of the system and their actions often play a role in system failure, but actions in other parts of the system are also important. The examples that are used will illustrate many of the ways in which errors can occur, and some of the mechanisms that management can use to help prevent error. A model of the overall system that integrates these models of action is developed in Chapter 6. While I have tried to make the situations examined with this example as realistic as possible in the context of a simple illustrative example, the accident probabilities calculated here are not meant to represent the actual risks associated with transporting hazardous materials.

#### **5.1 Expected Utility Theory and the Rational Model**

Expected utility theory, also called subjective expected utility (SEU) is the classic rational model of human behavior, assuming that individuals choose actions that are in their own rational best interest. It extends the economic "rational man" concept to situations involving choice under uncertainty. Expected utility theory, which dates from Bernoulli's 1738 analysis of the St. Petersburg paradox (Sommer, 1954), has played a dominant role as a model of human action and intention formation since von Neumann

and Morgenstern (1947) axiomatized utility theory and Savage (1954) extended the model to include the decision maker's subjective probability distributions. The expected utility model is useful in describing decision-making behavior, because it is widely applicable, relatively simple, and individuals generally act more or less rationally. Particularly important for risk analysis applications is its ability to explicitly address uncertainty.

In spite of (or perhaps because of) its central place in the literature, the descriptive use of the expected utility model has been the target of extensive criticism<sup>19</sup>. Some critics claim that by focusing on explicit choice under uncertainty, it fails to represent many common situations, and others claim that even in the situations where it does apply, individuals systematically violate its predictions. However, to claim that individuals do not generally act (at least approximately) in their own rational self interest is to contradict intuition, experience, and experimental evidence. Bueno de Mesquita (1981; 1980) has had success in using an expected utility model to analyze the actions of national governments and leaders in international conflict; Zakay (1986) has had similar success using a multi-attribute utility model to predict individual behavior. While the notion of omniscient humans who use unlimited powers of calculation to maximize expected utility is clearly inappropriate, SEU is consistent with a more realistic picture of individuals who use incomplete information and finite cognitive powers to behave rationally within these limitations<sup>20</sup>. Though SEU is not likely to be a perfect predictor of action, when it is applied judiciously, it is a valuable model that can make sense of much human behavior: individuals are generally sensitive to incentives and probabilistic information, and usually react rationally to such factors. To characterize behavior with a different model that contradicts expected utility theory is to expect that individuals will consistently violate their own rational self-interest – a chancy proposition, at best. And to use such a model as the basis for managing individuals in a complex system that has the potential to fail catastrophically is to invite disaster. At the very least, the expected utility model can be used to ensure that individuals do not have rational incentives to take actions that increase

---

<sup>19</sup> The SEU model has had a dual identity, being applied in some cases as a descriptive model of actual behavior, and in others as a normative model of how individuals should choose between alternatives. I use it here in its descriptive capacity, for which it has received much of its criticism.

<sup>20</sup> Simon (1972; 1986) is one of the strongest critics of the rational model for its presumption of omniscience and global rationality, though even he has stated elsewhere (1957) that it may be possible for an individual to behave rationally with respect to his limited, possibly incorrect knowledge and beliefs about the problem. This "subjective rationality", in which the decision maker maximizes expected utility in the context of the problem as he understands it, is the sense in which the expected utility model is used here, and is consistent with the personalistic view of rational decision making originally advocated by Savage and Edwards.

risk. Even so, this model is not appropriate in all circumstances. Many situations simply do not fit the expected utility paradigm of explicit choice among well-defined alternatives, which is why it is necessary to develop two other models of intention formation<sup>21</sup>.

#### Quantifying the Expected Utility Model

Expected utility theory structures alternatives, information (including uncertain information), consequences and preferences in a rational model of choice behavior. The process an individual actually uses in making a decision can be thought of as an explicit process in which the decision maker actually maximizes utility, or as an implicit process in which the actor uses heuristics and intuition to approximate utility maximization. In either case, the expected utility model allows the individual's actual choice behavior to be predicted.

When using the expected utility model from the perspective of management to predict an actor's behavior, there are two levels of uncertainty involved. First is the actor's subjective uncertainty about outcomes, which is captured in the expected utility model. Since the alternative that maximizes expected utility is generally unique, the expected utility model's prediction of action is deterministic. However, as was mentioned above, behavior models are generally used in this framework to make probabilistic predictions of action. This second level of uncertainty appears because in using any behavior model to predict action, management may be uncertain about the parameters that affect the actor's behavior – the actor's preferences, state of information, probabilities, beliefs about consequences, etc. The uncertainty in these factors may stem from uncertainty about a given individual, (e.g., a given individual's level of risk aversion is not known) or from a distribution in a population of individuals, (e.g., a number of individuals who have different levels of risk aversion). It is important to recognize that objective measures of these parameters (even when that is possible) are not relevant; the actor's decision will be driven by her beliefs about them at the time the decision is made. So the uncertainties that must be quantified reflect management's beliefs about the actor's beliefs. These

---

<sup>21</sup> There are a number of psychological theories that address emotional and other psychological forces that drive human behavior (e.g., Asch, 1952; Deci, 1975; Festinger, 1957; Freud, 1966). These theories and the forces they model are not considered here, because they are typically not applicable to the types of actions and the forces influencing them that are important in the applications addressed by this framework. Actions such as that of a disgruntled postal worker turning a gun on co-workers, while potentially a very real and important risk, are best addressed by other methods (clinical psychiatry, in this case). On the other hand, risk analysis may be an appropriate tool for studying a security system designed to protect against such an assailant, and the behavior models developed in this framework may be reasonable descriptors of the actions of other individuals within the security system.

uncertainties translate into management's uncertainty over which alternative has greatest expected utility for the actor, and thus which action will be chosen. By characterizing management's uncertainty about the parameters of the expected utility model, a probabilistic prediction of action is generated.

Figure 5.1 uses a decision tree to illustrate a simple decision in which the actor chooses between two alternatives and faces only one uncertainty. In this situation, management may in fact be uncertain about several aspects of the actor's decision, such as the actor's subjective probabilities, or the form or parameters of the utility function. To understand the implications of this uncertainty for the actor's choice, management must specify its (management's) joint distribution on these uncertain parameters that characterize the actor's decision problem, and calculate the resulting probabilities that the actor will choose each of the alternatives. It does this by integrating the joint distribution on the parameters over the area corresponding to the choice of each alternative – the region in which the given alternative has expected utility greater than all other alternatives. In this example, the probability that the actor will choose alternative A is the integral of the joint density function over the region in which alternative A has higher utility than alternative B:

$$p(\text{Actor chooses Alt A}) = \int_{\substack{\text{region in which} \\ \text{EU(A)} > \text{EU(B)}}} f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n$$

where the  $x_i$  are the parameters of the actor's decision problem about which management is uncertain, and  $f$  is their joint distribution. The limits of integration are set so as to define the region in which the alternative A has the greatest expected utility. A similar equation, integrating over the complementary area, gives the probability that the actor will choose alternative B, and in the case of more than two alternatives, an equation like this can be written for each.

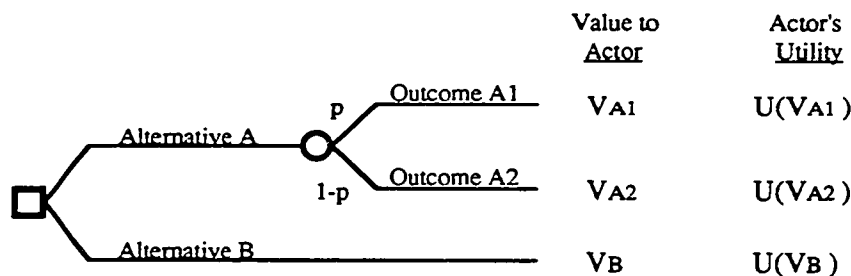


Figure 5.1: A simple generic decision.



### Management Controls in the Expected Utility Model

According to the expected utility model, choice, and thus subsequent action, is determined by four factors:

- 1) the set of *alternatives* considered,
- 2) the decision maker's *information* (subjective probability distributions) about the possible events or outcomes associated with these alternatives,
- 3) the *consequences* to the decision maker resulting from each combination of alternative and outcome, and
- 4) the decision maker's *preferences* for consequences.

By characterizing these factors, management can make reasonable predictions of an actor's behavior. Further, if management can influence these factors, they can be used as "control knobs" with which to influence the rational actor's choice and behavior. By changing the problem that the actor solves, management exerts some control over the solution that is reached.

In fact, each of these factors is at least partly under management control. Management may be able to add or eliminate some alternatives from the actor's decision problem, making them available for choice or eliminating them from consideration. Changing the resources available is one way it can do this – by increasing the resources within the organization and those at the individual's disposal, management may increase the set of feasible alternatives.

The expected utility model itself does not make the distinction between information and knowledge made above, but that distinction may be important for strategies designed to influence a rational actor's behavior. If a lack of information (in the sense of the current status of the system, etc.) is the problem, then better formal and informal information systems that aid in the discovery and communication of system state may be helpful. However, if a lack of knowledge (about how the system functions, etc.) is the problem, then different strategies, such as selection or training, will be more successful. Cognitive aids (management information systems, artificial intelligence systems, computers, etc.) may increase an actor's ability to deal with complex problems, effectively increasing her information and/or knowledge.

An obvious strategy for influencing the behavior of a rational actor is to change the consequences to the actor by associating incentives (rewards and/or punishments) with either actions themselves, if this is possible, or with system outcomes that are influenced by actions. Of course, this management technique is nothing new (no doubt it was well-established when it was employed in the construction of the Egyptian pyramids), but the

expected utility model demonstrates how and why it works, and can quantify its effects. Incentive strategies address the problem of goal conflict, which is essentially the same as the principal-agent problem in the economic literature (see, for example, Arrow, 1971). The agent (actor) acts to maximize her own utility, which does not necessarily lead to the action that would maximize the principal's (organization's) utility, since the agent has a different utility function and faces different outcomes. (For example, the individual and the organization face different outcomes when the individual is rewarded for productivity but faces no penalty for safety violations. It would not be surprising if an individual in such a situation were to sacrifice safety for productivity. This is not to say that individuals will intentionally cause failure in order to pursue self-interest, but that incentives may lead an actor to take actions that increases system risk, perhaps unwittingly.) The solution to the principal-agent problem is to align the consequences to the actor with organizational outcomes, setting incentives so that the agent, in maximizing her own expected utility under the incentive scheme, takes actions that also maximizes the principal's expected utility. In the one-dimensional principal-agent problem, this implies risk-sharing between the principle and agent – the agent bears some of the consequences of her actions, with the degree of risk-sharing depending on the relative risk attitudes of the principal and agent. In the more general situation, multiple outcome dimensions may be relevant (different forms of reward, for example). Also, the restriction that the principal cannot observe the agent's actions directly may not hold, and management may be able to observe and motivate behavior directly, which is more effective than rewarding only actual outcome. The general principle involved is to set the consequences to the individual so that actions that lead to good organizational outcomes also reward the actor, and actions that lead to bad outcomes or risk for the organization lead to undesirable outcomes for the actor as well.

A more bureaucratic management approach would be to simply prohibit the choice of undesirable alternatives. It may seem that such a prohibition would be difficult to model with an expected utility model, but in fact, to implement this strategy, the prohibition would need to be communicated to actors and enforced, with mechanisms for detecting and penalizing violators. This would change the structure of the actor's decision to include factors such as the chance of being caught and punished for violating the prohibition, but ultimately it consists of manipulation of the actor's outcomes and information, and can be evaluated with an expected utility model.

It may even be possible to change an actor's preferences, and thereby affect the solution to the decision problem. The socialization or indoctrination process, by which an

individual absorbs the principles, values, and beliefs of the organization, is common practice, and can cause the individual's preferences to shift so that they are more similar to those of the organization. This addresses the goal congruence issue in the sense in which it is discussed by organizational behaviorists, and is particularly apparent in military and religious organizations. Another method by which management can affect preferences is through selection, where instead of changing a given individual's preferences, the organization selects individuals whose preferences and values are already compatible with those of the organization.

**Example: Driver's Decision with the Expected Utility Model**

To illustrate the expected utility model, it can be used to predict the actions of the truck drivers in the hazardous material transport example. In this case, the expected utility model is used to predict the drivers choice of maximum driving speed. The result will be an expected frequency distribution for the number of drivers that will choose each possible maximum speed. Actual driving speed will vary, of course, depending on road and traffic conditions, and the maximum speed will actually be achieved only when road conditions allow. Drivers set their maximum speed in relation to the speed limit, and choose from among the following three alternatives:

- A: Speed Limit – maximum speed equal to the speed limit
- B: 10 mph Over – maximum speed 10 miles per hour over the limit
- C: 20 mph Over – maximum speed 20 miles per hour over the limit.

They base the decision on the advantages of finishing the route sooner and having a bit of extra time off, weighing this against the possibility of receiving a speeding ticket. Accident probability increases with speed, which is why this action is relevant to system risk, but drivers do not consider this in making their decisions. They value their time at \$10 per hour, and driving at 10 mph over the limit will save 3 hours, while going 20 mph over the limit will save 5 hours. A speeding ticket costs a driver \$500, because the company imposes an additional fine on the driver to help offset increased insurance rates. The driver's decision is illustrated by decision tree of Figure 5.2.

Management uncertain about the probability a driver will assign to receiving a ticket – the uncertainty is captured by a uniform distribution of 0.0 to 0.10 for the probability of receiving a ticket at 20 mph Over the limit, and half that for traveling at 10 mph Over the limit<sup>22</sup>. Management is also unsure about the driver's preference and risk attitude, but

---

<sup>22</sup> While it may seem unrealistically overconfident for the driver to assess a 0% chance of getting a ticket while traveling at 20 mph over the speed limit, if this could accurately reflect the beliefs on which the

feels that this uncertainty is adequately represented by an exponential utility function for the driver:

$$U(x) = 1 - e^{-\gamma x}$$

with the risk aversion coefficient gamma ( $\gamma$ ) ranging from 0.0002 to 0.002, uniformly distributed<sup>23</sup>.

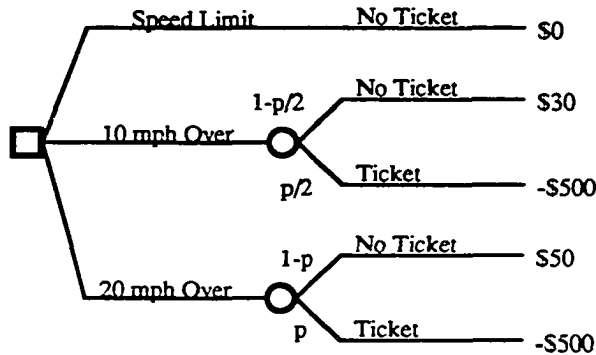


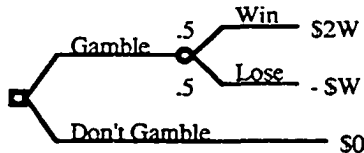
Figure 5.2: Driver's decision tree for decision about driving speed.

The two different levels of uncertainty associated with using the expected utility model to predict action are clearly illustrated here. The first is the driver's uncertainty about events: whether he will get a ticket. The second level of uncertainty is management's uncertainty about the driver's preferences and information: his subjective probability of getting a ticket ( $p$  at 20 mph over,  $p/2$  at 10 mph over, with  $p$  uniformly distributed from 0 to 0.1), and the risk aversion parameter in the driver's utility function, ( $\gamma$ , uniformly distributed from 0.0002 to 0.002). Management believes that these parameters are independent, so their joint density function is just

$$f(p, \gamma) = \frac{1}{(0.002 - 0.0002)(0.1 - 0.0)} ; 0.0 \leq p \leq 0.10; 0.0002 \leq \gamma \leq 0.002.$$

driver's decision will be based, it must be considered in predicting his behavior. Of course, if the goal is to change his behavior, rather than just predict it, then changing this belief may be an effective strategy.

<sup>23</sup> An intuitive way to depict the level of risk aversion implied by this is the following: Imagine a gamble with 50% chance to lose a given amount ( $W$ ), and 50% chance to win twice that amount ( $2W$ ) as shown.



An individual with the exponential utility function  $U(x) = 1 - e^{-\gamma x}$  would be willing to accept this gamble for any amount up to approximately  $1/\gamma$ . So a range of .0002 to .002 for  $\gamma$  implies that the driver would be

To calculate the probability that a given alternative will be chosen, this joint density function is integrated over the regions in which the combination of ticket probability and risk aversion coefficient correspond to the driver preferring that alternative. (In this case, the boundaries of these regions are found by setting the utilities of pairs of alternatives equal, and solving the resulting equation for  $p$  in terms of  $\gamma$ . These curves partition the parameter space into areas of known preference ordering of the alternatives. Integrating the joint density function over the appropriate areas yields the probabilities of each of the actions.) Doing this, the probabilities for each of the actions are:

<u>Action</u>	<u>Probability</u>
A: Speed Limit	0.15
B: 10 mph Over	0.32
C: 20 mph Over	0.53

This is an alarming result – that 85% of drivers make a rational decision to drive significantly over the speed limit – but on reflection, it is not terribly surprising. So this may be an effective point for management to intervene to reduce risk.

In general, management would like not only the ability to predict action, but also to change it (for example, to reduce the number of drivers who speed). There are a number of mechanisms at its disposal in a situation like this, and the expected utility model can be used to suggest them, demonstrate their effects, and estimate their consequences. One of the more obvious mechanisms is by using direct incentives. Unfortunately, because a driver's speed cannot be directly monitored, it is difficult to eliminate the positive incentive to the driver for speeding. But it is possible to increase disincentives for speeding by increasing the penalties to the driver for receiving a ticket. This will decrease the relative attractiveness of the more risky 20 mph Over alternative (similarly, but less so, for the 10 mph Over alternative), and the expected utility model can quantify the changes in the driver's choice probabilities. For instance, increasing the penalty associated with a ticket from \$500 to \$750 would significantly decrease the probability that the driver would choose to exceed the speed limit, reducing the probability of choosing 20 mph Over from 53% to 32%, and reducing the probability of choosing 10 mph Over from 32% to 19%. Increasing the penalty further, to \$2000, would nearly eliminate speeding, reduce these likelihoods to just 6% and 3%, respectively.

---

willing to accept this gamble for values of  $W$  up to about \$500 to \$5,000 - a plausible range of risk aversion.

The rational model also points out other mechanisms that can be used to influence the driver's behavior. Altering the driver's subjective probability of receiving a ticket would change the relative attractiveness of the alternatives and thus the likelihood that each would be chosen; an information and training program that provides information about the frequency of speeding tickets might change the range on the drivers' probabilities of being ticketed. If it can increase the lower bound of the range from 0.0 to 0.05 (leaving the upper bound and all other parameters unchanged) then the likelihoods of drivers choosing 20 mph Over and 10 mph Over change to 5% and 65%, respectively. This is actually a large increase in the number choosing 10 mph Over, but a sharp decrease in the number choosing 20 mph Over, and an overall decrease in the total number choosing to speed.

Another possible way for management to decrease the number of drivers who speed would be to selectively employ drivers who are more risk averse. Though this may take some time to implement because of the limited rate of driver turnover, it would eventually increase the risk aversion coefficient gamma ( $\gamma$ ) of the population of drivers. If, for example, by selecting for experienced drivers who have no speeding tickets, it were possible to get drivers who were twice as risk averse (gamma ranging from .0004 to .004, instead of .0002 to .002, as in the Base Case), then the fractions choosing 20 mph Over and 10 mph Over decrease to 39% and 25% (from 53% and 32%, respectively).

Although expected utility is widely accepted as a normative model of how individuals should choose to act, there has been considerable disagreement over its accuracy as a descriptive model of actual human behavior. A significant body of experimental research in behavioral decision theory has identified situations in which observed choice behavior appears to deviate systematically from the predictions of the SEU model, and has led to attempts to modify or replace the rational model in an effort to improve descriptive accuracy. Most of these alternative models are structurally similar to the SEU model; they differ by changing the functional form of the model and/or transforming the inputs (probability and value)<sup>24</sup>. I have chosen here to use the SEU model rather than any of these others, because it treats issues such as risk aversion and updating of information

---

<sup>24</sup> Kahneman and Tversky's Prospect Theory (1969) and Shanteau's (1975) Information-Integration Theory, as well as Karmarkar (1978) transform probabilities into "decision weights"; Prospect Theory also uses a value function defined on gains and losses in place of a proper utility function. Hogarth (1980), Luce and Raiffa (1956) and Bell (1982) use regret as an additional dimension of value, then either apply the SEU model or substitute a minimax criterion. Slovic and Lichtenstein (1968) and Payne (1973) find some support for an additive functional form (as opposed to SEU's multiplicative form) that directly sums weighted probabilities and outcomes. Coombs' (1975) portfolio theory treats risk itself as an attribute of value, trading it off against expected value.

more completely, and despite its critics, it is the most widely accepted model of choice behavior. Still, if there is compelling evidence that one of the alternative choice models offers greater predictive accuracy in a particular situation, it can be used in place of the SEU model, and it will be implemented within the framework in exactly the same way.

The next section describes a quite different approach to modeling decision behavior, using Simon's concept of bounded rationality. Rather than characterizing choice behavior as a rational optimization procedure, it describes decision making as a process driven by the actor's cognitive limitations. In a sense, bounded rationality is a substitute for the expected utility model, and a decision situation could be modeled by either. However, there are significant differences in the approaches of the two models that make them more complementary than competitive.

## **5.2 The Bounded Rationality Model**

Bounded rationality is a concept that evolved largely in reaction to the use of rational models to describe the behavior of individuals and organizations. The idea was first expressed by Simon, and developed by Simon, March, Cyert and others; most of the early foundations for this work were laid by Simon (1957, 1956, 1955), March and Simon (1958), and Cyert and March (1963). Simon (1957, p198) identifies

the principle of bounded rationality: The capacity of the human mind for formulating and solving complex problems is very small compared with the size of the problems whose solution is required for objectively rational behavior in the real world – or even for a reasonable approximation to such objective rationality.

Unable to act truly rationally, individuals instead employ simplified heuristics that allow them to make generally good, but not necessarily optimal, decisions, within the constraints of their abilities. Bounded rationality is not viewed as being at all irrational, and Simon was careful to distinguish it from social psychological theories of affective (emotional) forces that drive human behavior. "Human behavior in organizations is best described as 'intendedly rational'; and merits that description more than does any other sector of human behavior" (1957, p196). The use of heuristics is a "rational" response to the need to make decisions in an environment so complex that it is impossible to behave rationally in an omniscient, objective sense<sup>25</sup>.

---

<sup>25</sup> Hogarth (1980), and Janis and Mann (1977) have gone a step further in suggesting that there is a meta-rationality to the use of heuristic decision strategies: it may be objectively rational not to maximize

Despite several decades of development, bounded rationality is not a unified theory of decision. It "is organized in a set of conceptual vignettes rather than a single, coherent structure; and the connections among the vignettes are tenuous" (March, 1988, p. 271). Many of the fundamental features of the theory, however, are fairly well accepted by now. Briefly, bounded rationality makes several primary observations. Alternatives are not known in advance; they must be discovered by search, which requires significant effort. When a decision maker must search for new alternatives, they are sought in the "neighborhood" of old ones, so new alternatives are usually just variations on familiar ones. Decision making is not a process of optimization, but of "satisficing" (Simon, 1957) in which decision makers simply strive to satisfy a goal or "aspiration level." Alternatives are generated and evaluated in sequence; the search process continues only until a satisfactory (not necessarily optimal) solution is found. An alternative is judged satisfactory if it meets or exceeds the pre-defined aspiration level. While there are typically many dimensions of performance, rather than considering and combining an alternative's performance on all dimensions, as the rational model might suggest, behavior is often characterized by "sequential attention to goals" (Cyert and March, 1963). That is, only one goal (dimension) is considered at a time. Search is often prompted by failure to achieve a goal on a particular dimension, and the search focuses on that dimension, continuing until it discovers an alternative that satisfies the relevant goal. Over time, as performance improves on one dimension and drops off on others, the focus of attention shifts to follow the "squeaky wheel."

The bounded rationality theory does not predict which dimensions and goals will be considered nor which alternatives generated – to Simon and others, these are empirical questions about the structure and functioning of human cognition, to be answered by empirical research. On the other hand, intuition and some insights from the psychology of judgment and decision may give some guidance here. Alternatives that have been chosen in similar situations previously are likely to be considered first, particularly if they have led to good outcomes in the past; other familiar alternatives are also good candidates. Entirely new alternatives, to the extent that they are considered at all, will generally be relatively minor variations on familiar themes. The outcome dimension that will be considered in evaluating alternatives will tend to be a "squeaky wheel" dimension, where performance fails to achieve the goal. For example, recent problems with safety (an accident, or violations of safety regulations), are likely to focus attention on the safety

---

expected utility, because heuristic strategies usually perform quite well, and the improvement expected from optimizing does not justify the high cost of analysis.



of alternatives, ignoring other dimensions, at least until safety performance is adequate and some other dimension becomes the squeaky wheel. Failure tends to prompt search for new alternatives, whereas success leads to maintaining the status quo, repeating the choices that have led to success.

While this type of decision behavior has obvious advantages, it can be hazardous in the realm of low probability, high consequence events such as the catastrophic failure of a complex system, where failure may be so costly that it is unacceptable to allow this sort of "muddling through" behavior. The fact that the system has not yet failed does not necessarily mean that status quo behavior is acceptably safe. One way around this problem is to observe and learn from "near misses" – partial failures that are not catastrophic, but indicate problems in the system. Tamuz (1988) discusses some of the issues and difficulties that organizations have in learning from accidents and near misses in the airline industry, including difficulties in focusing attention and the value of retaining ambiguity in the accident reports.

In some instances, it is appropriate to use the principle of bounded rationality simply as a way to guide the application of the expected utility model – to ensure that the model constructed to represent the actor's decision process reflects simplifications the actor is likely to make. Because of cognitive limitations, the decision problem an actor solves will usually be a significant simplification of the actual problem. Outcomes that are readily available to the actor (in the sense in which Tversky and Kahneman, 1974, use availability, to refer to factors that are salient and easily called to mind) will be considered and will be judged as relatively likely, and others may be judged as unlikely or ignored altogether. Familiar information and alternatives are most likely to be included in the decision process.

But when a rational model is not appropriate for predicting action, the principles of bounded rationality can be operationalized to provide an alternative. Bounded rationality has many of the same implications as does the rational model, because individuals are at least intendedly rational. For example, increasing incentives or the probability of preferred outcomes will generally encourage and increase the likelihood of the associated behavior. However, bounded rationality is very different from optimizing behavior, and will often lead to different outcomes, for many of the reasons mentioned above.

#### Quantifying the Bounded Rationality Model

Unlike the expected utility maximization model, bounded rationality has not (as yet) been

developed into a single, precise theory that can be readily quantified to predict decision behavior. Therefore, I must create such a quantifiable model to capture the principles of bounded rationality as much as possible. I do not claim that this model is the only way to capture the features of bounded rationality, nor that it captures all these features (even if there were complete agreement as to what they are). Rather, it is one reasonable attempt to operationalize some of the major conceptual ideas of bounded rationality in a quantifiable decision model that can be used in a predictive fashion.

One of the distinctive features of bounded rationality is that rather than evaluating all alternatives and selecting the best, individuals *satisfice*; they evaluate alternatives sequentially and select the first that achieves a specified aspiration level. Another feature is *sequential attention to goals*; while there are typically multiple dimensions of value associated with a decision, actors focus their attention on one at a time. Quantifying this process to predict the decision maker's ultimate action leads to a model with two distinct parts: the first establishes the sequence in which alternatives will be evaluated; the second part evaluates the alternatives in the specified sequence, judging them according to their performance on one dimension, and continuing to evaluate alternatives in turn until an acceptable one is found. I use a probability tree to describe the sequence of evaluation, and develop a multidimensional evaluation model that characterizes uncertainty in attention allocation among goals.

#### - Evaluation Sequence

The sequence in which alternatives will be evaluated is typically uncertain, and can have a significant effect on the ultimate decision. A useful way to describe this uncertainty is with a probability tree – specifying the probability that each alternative is the first to be evaluated, and then at each subsequent branch in the tree, the probability that each of the remaining alternatives is evaluated next, as illustrated in Figure 5.3.

For a decision with  $k$  alternatives, there are  $k! = k (k-1) \dots (2) (1)$  possible evaluation sequences. The marginal probability that alternatives are evaluated in sequence  $S_i = \{w, x, \dots, z\}$  is the product of the probabilities along the corresponding branch:

$$p(S_i: \{w, x, \dots, z\}) = P_{w1} P_{x2 | w1} \dots P_{z(k) | y(k-1), \dots, x2, w1}$$

The probability tree structure allows the probabilities to be assessed separately at each branch, which gives maximum flexibility to tailor the model to the situation, if there is information to support such assessments. In some cases, however, it may be reasonable to define a few rules that describe the relative probabilities throughout the tree. For example, by specifying that

$$P_{Ai} = k_1 P_{Bi}; P_{Bi} = k_2 P_{Ci}; \text{ for all } i,$$

all the probabilities in the three-alternative tree above are determined by the two constants  $k_1$  and  $k_2$ . A scheme such as this may be used to characterize a situation in which some alternatives are judged more likely to be considered first because they are more apparent, or more "available."

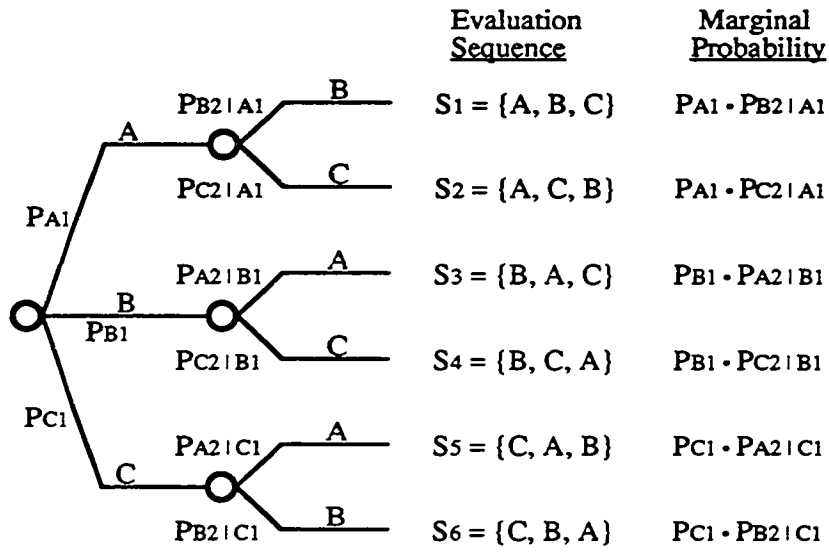


Figure 5.3: Probability tree illustrating evaluation sequence for three alternatives in a bounded rationality model.

- Evaluation Algorithm

For a particular evaluation sequence, alternatives are evaluated one at a time in the given sequence. An evaluation algorithm is used to determine whether a given alternative is acceptable, and when an acceptable alternative is discovered, the search stops and that alternative is selected. There are generally multiple dimensions of value; uncertainty over which of these the decision maker will use is characterized by probabilities  $p_j$ , the probability that dimension  $j$  will be used to judge alternatives. Alternative  $i$  will be accepted when the decision is based on dimension  $j$  if its value on that dimension,  $V_{ij}$ , is greater than or equal to the aspiration level for dimension  $j$ ,  $A_j$ ; if  $V_{ij}$  is less than  $A_j$ , then alternative  $i$  will be rejected, and search will continue with the next alternative in the evaluation sequence.

The way in which  $V_{ij}$  is modeled depends on how dimension  $j$  is valued. It may be possible to directly express an alternative's value on a natural or a constructed value scale, or if there is uncertainty in the outcome,  $V_{ij}$  can be defined as the expected value on

dimension  $j$  (in this case, it may be necessary to model preference by defining a nonlinear value function that is applied to the outcome, since strength of preference may not be linearly related to the outcome measure). Different dimensions need not be modeled in the same way.

The probability of accepting a given alternative, conditional on the evaluation sequence  $w, x, \dots, y, z$ , is equal to the sum over dimensions,  $j$ , of the probability that dimension  $j$  is used as the criterion for the decision, times a Boolean variable that indicates whether that alternative will be chosen in that situation. A given alternative will be chosen if it meets the aspiration level on the relevant dimension, and all alternatives before it in the evaluation sequence do not (it is also necessary to include an adjustment term,  $\epsilon_{ij}$ , explained below). This is given by:

$$p(\text{choose } w \mid \text{seq. } w, x, \dots, z) = \sum_j p_j \{ \delta_{wj} + \epsilon_{wj} \}$$

$$p(\text{choose } x \mid \text{seq. } w, x, \dots, z) = \sum_j p_j \{ \delta_{xj}(1 - \delta_{wj}) + \epsilon_{xj} \}$$

$$p(\text{choose } z \mid \text{seq. } w, x, \dots, z) = \sum_j p_j \{ \delta_{zj}(1 - \delta_{wj})(1 - \delta_{xj}) \dots (1 - \delta_{yj}) + \epsilon_{zj} \}$$

where  $p_j = p(\text{decision based on dimension } j)$

$$\delta_{ij} = \begin{cases} 1 & \text{if } V_{ij} \geq A_j \\ 0 & \text{if } V_{ij} < A_j \end{cases}$$

$$\text{and } \epsilon_{ij} = \begin{cases} 1 & \text{if } V_{ij} < A_j; \text{ and } V_{ij} > V_{kj} \text{ for all } k \neq i \\ 0 & \text{otherwise} \end{cases}$$

The term  $\delta_{ij}$  has value one if alternative  $i$  meets the aspiration level on dimension  $j$ , and zero otherwise. The product of  $\delta_{ij}$  and one minus  $\delta_{kj}$ , for  $k$  before  $i$  in the evaluation sequence, indicates whether alternative  $i$  is the first in this particular evaluation sequence to meet the aspiration level on dimension  $j$ . The adjustment term  $\epsilon_{ij}$  accounts for the possibility that none of the alternatives achieve the aspiration level on dimension  $j$ . If this is the case, the decision maker prefers the best among bad alternatives, choosing the one that comes closest to the aspiration value. To capture this, when no alternative achieves the aspiration level,  $\epsilon_{ij}$  has value one for the alternative which has the highest value on dimension  $j$ , and zero for all others. These calculations are performed for each possible evaluation sequence, and the results weighted by the probabilities of evaluation sequences

and summed, to find the overall probability of choosing each of the alternatives:

$$p(\text{choose } w) = \sum_{n=1}^{k!} p(\text{choose } w \mid \text{seq. } S_n) p(\text{seq. } S_n)$$

### Management Controls in the Bounded Rationality Model

Many of the management strategies that affect action in the expected utility model will have similar effects in the bounded rationality model, because the two models apply to similar situations, and because actors who follow a bounded rationality strategy are at least attempting to be rational. A consequence of the bounded rationality principle is that since individuals are intendedly but imperfectly rational, management may be able to improve their actions by providing assistance to overcome their cognitive limitations. For example, by providing information that the actor may lack, making it both available and salient, management can make it more likely that this information will be considered in the decision. Computational or cognitive aids that help the actor to process information can also be helpful, (e.g., simulating system functions to help the actor understand the consequences of actions), though the actor's abilities to assimilate and comprehend such information must be considered in judging its effect on the decision process.

There are several ways that management can affect the behavior of an actor who makes decisions according to the bounded rationality model described above. Management can influence the order in which alternatives are considered; it can change the likelihoods that each of the dimensions is used as the decision criterion; and it may be able to change the aspiration level on one or more dimensions. Some of the mechanisms at its disposal may affect more than one of these at a time. In broad terms, management "attention" to various goals and alternatives will make them more important in the individual's decision. Because the sequence in which alternatives are evaluated is so important in the bounded rationality model, simply making an alternative more familiar or salient can make it more likely to be considered first and thus chosen. Likewise, management attention to a particular dimension or goal makes it more likely that actors base decisions on that criterion. Management strategies that can influence the aspiration level (the standard of acceptability for alternatives) may also affect the actor's choice – for example, by stressing the importance of safety.

Even though actors do not treat them in the same way as in the expected utility model, changes to incentives and information may have similar effects under bounded rationality. For example, increasing safety incentives will increase the chance that safety

will be used as the decision criterion, and may make safer alternatives more likely to be considered by making them more salient. This will result in an increased likelihood of choosing a safer alternative, even though the incentives are not used in a rational calculation of outcome and expected utility. A difference between the rational and the bounded rationality models is that in rational decision making, since behavior is entirely self-interested, management must somehow affect the actor's self-interest in order to change behavior. In bounded rationality behavior, this is not necessarily the case; some mechanisms that do not affect the actor's self-interest may have a strong influence on the decision.

Example: Driver's Decision with the Bounded Rationality Model

The bounded rationality model can also be illustrated by applying it to the example of transporting hazardous materials. In this case, the bounded rationality model is used to model the decision of a driver who has a trip scheduled, but is faced with very bad driving conditions due to a major storm, which significantly increases the risk of an accident. The driver must choose one of the following three alternatives:

- A: *Go Now* and risk an accident
- B: *Wait* a day, until driving conditions improve
- C: Take an *Alternate Route* with slightly lower risk.

First, the sequence in which these three alternatives will be evaluated is quantified with a probability tree using simple equations that describe the relative probabilities throughout the tree. An alternative that is salient and familiar is more likely to be evaluated first in the bounded rationality decision heuristic. Since the Go Now alternative is the driver's "default" alternative, in that it has already been planned, it is most likely to be considered first. Next most likely is the Wait alternative, because it is the natural alternative to the Go strategy. The Alternate Route strategy is the least likely because it requires that the driver look beyond the more obvious alternatives to consider a new one. For this example, the probabilities for the evaluation sequence are expressed by:

$$p_{Ai} = 2 p_{Bi} ; p_{Bi} = 2 p_{Ci} ; \text{ for all } i,$$

whenever the corresponding alternatives have not yet been evaluated. This information implies that the probability tree for the sequence of evaluation is as shown in Figure 5.4, with the marginal probabilities of each sequence displayed at the ends of the branches.

The evaluation algorithm is used to determine the likelihood that the driver will choose each of the alternatives, conditional on each of the possible sequences in which they are evaluated. The driver may evaluate the alternatives on their performance on the

dimension of Safety, or on Schedule. The values of each of the alternatives on constructed scales for each of these dimensions are:

<u>Dimension</u>	<u>A: Go Now</u>	<u>B: Wait</u>	<u>C: Alt. Route</u>
Safety	-1	1	-.4
Schedule	1	-1	.8

These scales have been constructed so that the aspiration level for each dimension is zero. Scores on safety dimension are related implicitly to drivers' subjective understanding of the likelihood of an accident, and scores on the schedule dimension reflect the amount by which the alternatives delay the original schedule.

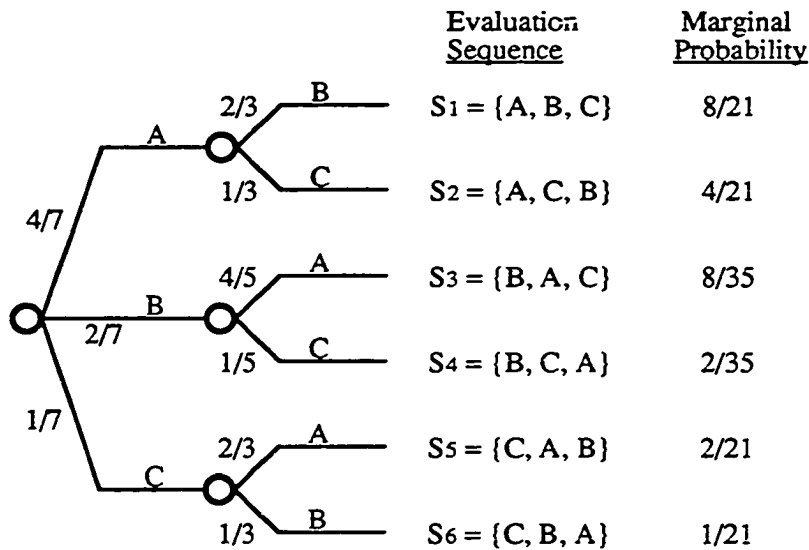


Figure 5.4: Probability tree of evaluation sequence for driver's decision.

Since drivers experience quite a bit of production pressure, there is a 0.65 probability that the driver will base his decision on Schedule, and a 0.35 chance that he will base it on Safety. Using the bounded rationality model developed above, the resulting overall probabilities of choosing each alternative are:

<u>Alternative</u>	<u>Probability</u>
A: Go Now	52%
B: Wait	35%
C: Alt. Route	13%

There are several ways that management can influence the driver's decision in the bounded rationality model. Management attention to safety will increase the chance that the driver's decision will be based on safety rather than schedule. This attention can

come in many forms: incentives may be effective, even though the driver's decision is not based explicitly on a rational consideration of outcomes. Simple management focus on safety (such as information or "propaganda" campaigns that stress the importance of safety in the organization) may also be effective. Emphasizing and communicating the acceptability of delaying a trip because of driving conditions will increase the likelihood that the Wait alternative will be considered.

While it may be difficult to be extremely precise in quantifying the effects of these management controls, it should be possible to approximate their effects on behavior. For this example, two management strategies are considered. The first is to eliminate much of the production pressure on drivers. This will have the effect of reducing the likelihood that the decision is based on schedule (because schedule is less of a "squeaky wheel"), reducing it from 0.65 to 0.55. It will also change the aspiration level on the schedule dimension so that the Wait alternative is acceptable under this decision criterion.

The second management strategy considered is to stress that safety is an important goal, and that it is legitimate to postpone a trip due to bad driving conditions. By explicitly recognizing the Wait strategy, management makes it more likely that the driver will consider it; the probabilities for the evaluation sequence change so that drivers are just as likely to consider Wait as Go, with Alternate Route unaffected:

$$p_{Ai} = p_{Bi}; p_{Bi} = 4 p_{Ci}; \text{ for all } i.$$

By making safety the squeaky wheel, this also increases the likelihood that the decision will be based on safety rather than schedule from 0.35 to 0.50. The effects of these management changes on the driver's decision in bad weather are shown in Table 5.1.

<u>Alternative</u>	<u>Base Case</u>	<u>Cut Production Pressure</u>	<u>Stress Safety</u>
A: Go Now	52%	31%	40%
B: Wait	35%	61%	50%
C: Alt. Route	13%	8%	10%

Table 5.1: Probabilities of driver's choice in the bounded rationality model for two management changes.

The results show that both these management strategies may be quite effective. Reducing production pressure may have a somewhat greater effect than stressing safety, even though stressing safety has a larger impact on whether the decision is based on safety. This is because simply increasing the importance of safety does not address the competing effect of production pressure. Relaxing production pressure makes it easier for actors to choose an alternative that performs well on the safety dimension.



### **5.3 The Rule-Based Model**

Another way to model the process of intention formation uses Rasmussen's concept of rule-based behavior (1983), in which the actor possesses a catalog of pre-established rules that prescribe appropriate responses to situations. The rule base can be thought of as a collection of "IF-THEN" directives that match action to the current situation, such as "IF situation is X, THEN do Y." An actor's rule-based decision process consists of identifying the situation, and selecting and applying the corresponding rule. An example of this type of behavior is a physician diagnosing and treating a patient's illness. The physician identifies symptoms, searches her rule base for a rule whose condition matches the observed symptoms, and the rule specifies the appropriate action. For example,

IF patient complains of a headache,  
THEN prescribe "two aspirin, call in the morning."

There may be chains of rules to diagnosis and solve a problem, such as:

IF patient complains of chest pain,  
THEN order a chest X-ray.  
IF X-ray indicates normal,  
THEN recommend rest, no treatment.  
IF X-ray shows signs of heart disease,  
THEN perform further tests to confirm diagnosis ...

...  
IF arterial blockage is confirmed,  
THEN recommend surgery ...

and so on, as complex as is necessary to adequately represent the system<sup>26</sup>. Clancey (1985) describes an artificial intelligence heuristic that applies rules at an abstract level that relates classes of problems to classes of solutions; the situation is abstracted to the problem class before applying the rule, and the solution refined to match the specific situation after. This may also serve as a reasonable model of how humans actually use rule-based reasoning.

The distinguishing characteristic of the rule-based decision mode is that the actor does not explicitly consider alternative actions and possible outcomes to reach a decision. Although an explicit decision process or trial-and-error learning might have been employed, by the actor herself or by others, to create the rule base in the first place, once it is in place, the actor simply applies coded rules to direct behavior. A rule-based decision process can be very efficient and effective in familiar situations; rather than engaging in the lengthy process of explicit decision making in every case, the knowledge

---

<sup>26</sup> At some point in such a set of rules, the actor may shift from rule-based decision making to another mode, such as bounded rationality. This is consistent with Reason's (1990a) GEMS model of problem solving.

behind such decisions can be encoded in a set of rules that are quickly and easily applied, reserving limited mental power for situations that demand it. It can result in errors, however, if a flawed rule specifies an inappropriate response, if an unforeseen situation occurs for which no rule exists, or if the situation is incorrectly identified and the wrong rule applied, leading to action that may be inappropriate for the actual situation.

While rule-based behavior is often efficient, there are some types of situations in which it may be particularly hazardous. There is some evidence to indicate that under threat, actors tend to revert to familiar, habitual responses, and avoid innovative solutions. This type of behavior, known as threat rigidity, may occur at both the individual and organizational levels, and is characterized by a restriction of information processing, a reliance on prior knowledge and the rigid use of existing procedures (Staw, et al., 1981; Zajonc, 1965). In his analysis of the collision of two 747's on takeoff at Tenerife, Weick (1990) identified "regression to more habituated ways of responding" as a factor in the accident. This implies that in crisis situations, individuals may tend to apply an inflexible rule-based decision process rather than explicitly evaluating alternatives. (This effect might also generalize to situations of high demand on actors, such as time pressure; this would have significant implications for the management of complex systems.) In such situations, a rule-based model may be a better descriptor of action. Even though rule-based behavior may not be as effective as a more reasoned approach in such situations, it is important to know that individuals tend to use it, and modeling such behavior may be able to help improve it. For example, ensuring that individuals understand that the rules of normal operation may not apply in exceptional situations, and providing alternative rules that are more appropriate for crisis situations, may improve outcomes.

The rule-based decision process described here is one in which the actor uses a pre-defined set of rules to direct decision and action – at the point of decision, the rules are not reconsidered. This is not necessarily the same as a situation in which the organization sets official policies and procedures (which can also be thought of as rules) that prescribe appropriate behavior for the actor. In such a situation, an actor may not actually make decisions according to a rule-based decision process. She may, for example, maximize expected utility, and consider violating the organization's rule, where the potential outcomes of this alternative include the possibility of being caught and punished. (Nonetheless, establishing an "official rule base" that defines appropriate action as a function of the situation may be one way for management to affect the rule base of an actor who does use a rule-based decision process, so it is not irrelevant here.) On the other hand, the actor may use a rule-based decision process, using rules developed

through experience or learned outside the organization, even if the organization does not establish official rules or policies.

### Quantifying the Rule-Based Model

A rule-based decision process is, in principle, straightforward to quantify. It is modeled as two separable parts: 1) the identification of the situation by the actor, and 2) the rules that specify appropriate actions for each particular situation. From the organization's perspective, it would be ideal if individuals in the system always correctly identified the actual situation, and had an encoded rule base that specified the organization's preferred action for each situation. Unfortunately, the reality is that not only do actors sometimes incorrectly identify situations and have rule bases that may specify responses that are inappropriate (for the organization, at least), but from the perspective of management, there may also be significant uncertainty at both levels of this process. It may unclear how an actor will identify a given situation, and it may be difficult to know what rules are encoded in the actor's rule base. These two levels of the model are described probabilistically to predict action.

We can characterize the actor's identification of the situation by defining the probability that the actor identifies the situation as  $S_j$ , given that the actual situation is  $S_i$ :

$$p_{ij} = p("S_j" | S_i)$$

The variables  $S_j$ ,  $j=1\dots m$ , denote possible situations, and the notation " $S_j$ " is used to indicate that the actor identifies the situation as  $S_j$ , whether or not that is the actual situation. Uncertainty about the actor's rule base – the rules used to determine action as a function of the situation (as it has been identified) – can be characterized using:

$$q_{jk} = p(A_k | "S_j")$$

where  $A_k$ ,  $k=1\dots n$ , denote possible actions. If the organization sets policies that define appropriate actions in various situations, then for each  $j$ , some  $A_k$  will match the organization's policy. However, this model allows the actor to choose an action other than that specified by the organization. These uncertainties can be displayed in the form of the two-stage probability tree of Figure 5.5.

The overall probability of a particular action,  $A_k$ , when the actual situation is  $S_i$ , is:

$$p(A_k | S_i) = \sum_j p_{ij} q_{jk}$$

In general, although the actor's identification of the situation is necessary for developing a probabilistic prediction of action, it will not affect the outcome except through the effect it has on the action chosen. That is, given that the actor takes action  $A_k$ , system outcome

does not depend on whether it was identifying the situation as  $S_i$ , or as  $S_j$ , which led to that choice of action.

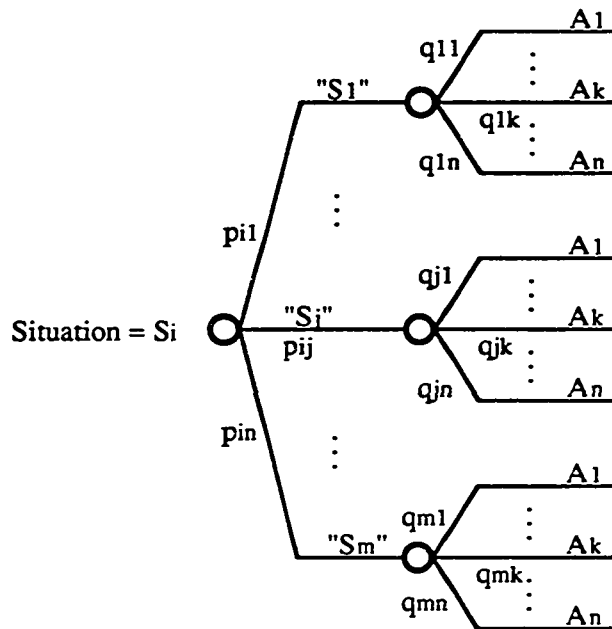


Figure 5.5: Probability tree illustrating uncertainty in a rule-based model.

#### Management Controls in the Rule-Based Model

There are a number of mechanisms that management can use to affect the behavior of an actor who uses a rule-based decision process. These can be divided into two basic type of strategies: 1) those that affect how situations are identified, and 2) those that affect the actor's rule base.

Strategies that are to affect the actor's identification of the situation must influence the actor's information and/or knowledge. The actor may identify the situation incorrectly because her information about the state of the system is incomplete or inaccurate, or because flaws in her knowledge lead to a misinterpretation of reliable information and an incorrect conclusion about the actual situation. Management can address information failures through information systems that provide the actor with more complete and accurate information about the state of the system, to the extent that such information is obtainable. It can change the actor's knowledge about how to interpret this information through mechanisms discussed in previous sections, such as selection and screening mechanisms, that ensure that the individuals in the organization have a good knowledge

base, and through training programs designed to improve the individual's knowledge base once they are a part of the organization.

To affect the actor's rule base, management must address the ways in which the rules are acquired and updated. Rules come from a variety of sources: they may be created by system management itself by setting policies, they may come from outside the organization, or they may encode ad hoc solutions developed by trial and error (Newell, et al., 1987). In the worst cases, rules may be based on nothing more than superstition (which may be particularly common in poorly understood systems) or they may be outdated habits that applied to an earlier system. One of the most obvious ways to affect an actor's rule base is for the organization to establish policies and procedures that create an "official rule base," and transfer this rule base to the actor. Especially in a situation where the consequences of alternatives do not affect the actor directly, establishing official policies may give guidance where there might otherwise be little to direct the actor's behavior. In such a case, part or all of an official rule base may be readily adopted, and may even encourage the actor to use a rule-based decision process instead of a different decision process. (Even if the actor does use some other decision process, such as expected utility maximization, organizational rules may be effective if the actor associates a disutility with violating them; however, the rule-based model would be an inappropriate way to characterize the actor's decision process.) Of course, an official rule base will be effective only to the extent that the actor is familiar with it, so the organization must have a mechanism, such as a training program, to communicate its policies to individuals in order to augment and improve their rule bases. In order to increase the likelihood that the organization's rules are actually adopted, it may help to justify them by communicating the consequences of actions, as in "When the reactor's primary cooling system fails, start the backup cooling system, *because otherwise there is a significant risk of a meltdown.*" This is especially important in systems that are so complex that actors cannot necessarily foresee all the consequences of their actions, and may help prevent the attitude that official rules are arbitrary.

Another way that management can affect the actor's rule base is by controlling which actors, and thus which rule bases, are in the system. Selection and screening mechanisms can help the organization acquire and retain individuals who have appropriate rule bases, as in the case of a hospital that selects interns on the basis of performance in medical school, on the presumption that this is an indicator of the quality of the rule-based knowledge that they will apply in diagnosing and treating patients.

An organizational strategy that is likely to be ineffective in influencing rule-based behavior is incentives. In the rule-based model, outcomes to the individual do not play a role in determining behavior. However, to take this too literally may overlook an important issue, and that is that while they do not actually enter the rule-based model, a significant change in incentives (or in some other aspect of the actor's decision context) may change behavior by causing the actor to reconsider the rules used, or by actually changing the mode in which decisions are made. A change in incentives may prompt the actor to abandon a rule-based decision process altogether, and begin explicit consideration of various possible alternatives and their consequences. New information, such as about the likelihoods of various consequences of an action, may have a similar effect, breaking the actor out of rule-based decision making into a more explicit consideration of alternatives by another decision making process.

**Example: Brake Maintenance with the Rule-Based Model**

The rule-based model can be used to investigate maintenance in the hazardous material transport example introduced above. Rather than the driver of the truck, this analysis focuses on the actions of technicians who are responsible for maintenance of the trucks, looking in this case at the effects of maintenance on the condition of the brake system, which may be important in preventing an accident. Trucks are serviced, and brakes inspected, every 10,000 miles. Brake condition at inspection is modeled as three discrete states: 1) Brakes Good; 2) Brakes Fair; 3) Brakes Worn. The technician uses rule-based reasoning to determine whether to replace the brakes. Based on the condition of the brakes as observed in the inspection, the technician decides whether to replace them using one of two possible rules: he will either replace the brakes if A) the inspection shows Brakes Worn; or B) the inspection shows Brakes Fair or Brakes Worn. Management believes that there is a 60% chance that the technician uses rule B) Replace if Fair or Worn, and a 40% chance that he uses rule A) Replace if Worn. Management also believes that if the actual condition of the brakes is Good, the technician's diagnosis will accurately reflect that; if actual condition is Fair, there is a 10% chance the technician will incorrectly identify it as "Good"; and if actual condition is Worn, there is a 5% chance the technician will diagnose it as "Good", and a 10% chance he will diagnose it as "Fair". Of course, whether the technician replaces the brakes depends on their condition as identified in the inspection, not on their actual condition.

A dynamic model that accounts for the deterioration rate of the brakes and frequency of service, as well as the actions of the technicians, is necessary to determine the effects on brake condition (this model is presented in section 6.3). Here, only the rule-based model

that describes the technicians' actions is presented. The probability tree of Figure 5.6 illustrates management's beliefs about the technician's rule-based decision process. The probabilities for the first level of uncertainty, how the technician will identify brake condition, are given for each of the three possible actual brake states: Good, Fair, and Worn.

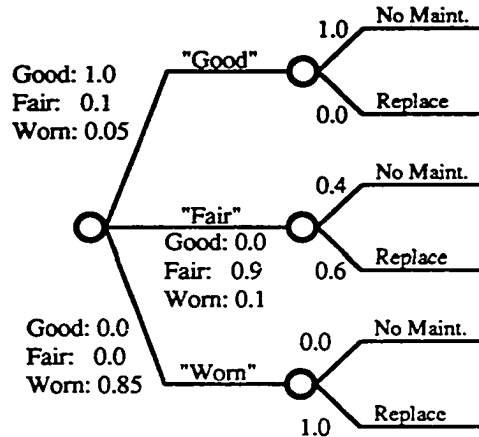


Figure 5.6: Rule-based model of technician's brake maintenance decision.

Table 5.2 shows the implications of this rule-based model for how the technician's behavior affects brake condition, illustrated as a Repair matrix, **R**. Element  $i,j$  of **R** gives the probability that the brake condition is  $j$  after service, given that it was  $i$  before service. This information will serve as an input to the stochastic model of brake condition that will be used to calculate the steady-state distribution of brake condition in section 6.3.

	Good	Fair	Worn
Good	1.00	0	0
Fair	0.54	0.46	0
Worn	0.91	0	0.09

Table 5.2: Base Case brake repair matrix, **R**, showing effect of rule-based action on brake condition.

Entry  $i, j = p(\text{state } j \text{ after service} \mid \text{state } i \text{ before})$

The two fundamental ways in which the organization can change the technician's action to improve brake condition and reduce risk are to improve the technician's ability to correctly identify brake condition in the inspection, and to increase the likelihood that the technician's rule base specifies that the brakes should be replaced when their condition is identified as Fair or Worn (as opposed to only when they are Worn).

One way that management can increase technicians' accuracy in brake inspection is through Inspection Training, a training program that teaches technicians to fully disassemble the brakes for inspection, rather than rely on a cursory check. Of course, this will not eliminate incorrect inspections, because not all technicians will follow the additional instruction, but management believes that it will cut the frequency of misdiagnosis by half. The effects of Inspection Training on brake condition after service are shown in Table 5.3.

	Good	Fair	Worn
Good	1.00	0	0
Fair	0.57	0.43	0
Worn	0.955	0	0.045

Table 5.3: Brake repair matrix, **R**, showing effect of inspection training on brake condition.

Another way that management could influence the technicians' action is by adopting an official Brake Replacement Policy specifying that brakes should be replaced more frequently, when they are in Fair or Worn condition, rather than waiting until they are Worn before replacing them. This policy is aimed at changing the technicians' rule bases, and would only affect action in those cases where brake condition is identified as Fair; management believes that such a policy would increase the probability that the technician replaces brakes in Fair condition from 60% to 90%. Table 5.4 shows the effect of Brake Replacement Policy on brake condition after service. To determine the effect of these two management changes on risk, the data in these tables must be used with the dynamic model of brake condition and a system risk model (see section 6.3).

	Good	Fair	Worn
Good	1.00	0	0
Fair	0.81	0.19	0
Worn	0.94	0	0.06

Table 5.4: Brake repair matrix, **R**, showing effect of brake replacement policy on brake condition.

#### **5.4 The Execution Model**

The second of the two primary types of error causes identified in the taxonomy of error causes is execution failure, the failure of the actor to properly carry out an intention. In an execution failure, the problem lies not with the formation of the intention or plan of action, but in the execution phase through which the plan is translated into action. The



idea of an execution failure is closely related to that of a slip or lapse from the human error literature, but it goes further to include actions and errors that would not normally be classified as slips or lapses.

Execution of an action, and possible execution failure, can be viewed in the context of the relationship between the abilities of the actor and the demands of the task to be performed, a measure of the difficulty of execution. A number of different dimensions of ability and demands may be relevant. While physical dimensions such as reaction time and strength can obviously play a role in system failure, demands on the cognitive abilities of actors – memory, vigilance, and capacity of mental processes – are often more important in complex systems. Complex technical environments that impose wide ranging demands can easily tax actors beyond their limits. While an operator may fully intend to carefully monitor and control a system, factors such as long shifts, heavy workloads, lack of training and experience and poor system design may make it difficult to do so reliably. Ironically, in a crisis, when it is most important that individuals be able to carry out intentions as planned, a sharp increase in demands on the actor may make it even more likely that the task demands will exceed the actor's capabilities.

The distinction between inadequate ability and excess task demand may not always be clear. Abilities can be inadequate only in relation to task demand, and task demand can be excessive only in relation to ability, so in a given instance, it may be impossible to say that one or the other is the cause of error. (Ideally, system design should balance task demands against the abilities of actors.) Despite this ambiguity, the distinction is important because it points to very different mechanisms that management can use to reduce the likelihood of errors – by making changes to the system to reduce task demand, or by changing the attributes of actors to improve their abilities.

The discussion of an actor's ability limitations is not necessarily restricted to those situations in which the actor is incapable of performing the desired action at all, though that is certainly one possible cause of an execution failure. The classic slip – a slip of the hand, mental lapse, etc. – is a common execution failure that is related to ability. In a slip, the actor possesses the basic ability to perform the intended action, but is not able to do so with perfect reliability, and on a particular occasion makes an error. While it may seem unusual to say that a lack of ability causes slips, since even the most capable actor is susceptible to them, the relationship between slips and ability will become clearer when the execution model is formalized.

There are several issues that the model of execution must be able to manage, some of which have already been alluded to. The model must be able to deal with errors caused by excess demands placed on the actor, as when design problems make it difficult or impossible for the actor to perform the required action. It must be able to address the question of the actor's ability, both the fundamental ability to execute a given action, and the possibility of slips. It must also handle the fact that different actors may face different responsibilities and different situations.

The model should also be able to consider several different outcomes of action in a given situation. In some situations, it will be sufficient to consider just two possible outcomes: successful execution and execution failure (error). In others, it may be necessary to describe more than two outcomes, as in different types or degrees of error, because they may have different implications for the possibility of system failure. An example of this latter case is the time required to execute an action, in a situation where system failure depends critically on time elapsed. In such a situation, the possible outcomes can be characterized with a distribution on the time required for completion. (The anesthesia application described in Chapter 3 looks at a system in which the times to complete actions are some of the crucial determinants of system failure, and uses continuous distributions to characterize these times.)

The final issue that the model should address is the possibility that the likelihood of an execution failure may not always increase monotonically with the demands placed on the actor. It is tempting to think of execution failures in terms of the classic engineering analysis of component failure, where failure depends on the relative values of component strength and the load that it must withstand, implying that the probability of error must increase with the demands on the actor. While this is certainly plausible, and has some experimental support (Berkun, 1964; Grinker and Spiegel, 1963), other psychological research suggests that the relationship may not always be so simple.

From their work on animals, Yerkes and Dodson (1908) formulated the inverted-U hypothesis, which asserts that task performance reaches a maximum at an intermediate level of arousal, falling off at both high and low levels. Duffy (1962) applied this hypothesis to human performance and found qualified experimental support. The effect may occur because at low levels of arousal, individuals lose interest or feel less motivated. If task demand were to be taken as a measure of actor arousal, this principle would imply that the probability of an execution failure may actually increase as task demand decreases below a certain point. While the inverted-U hypothesis has limited

empirical support, at best (Neiss, 1988, offers a good review and critique of this research area), it does offer a caution against automatically assuming a simple monotonic relationship between demands on the actor and error rate, and the model should retain the flexibility to capture such effects.

### Quantifying the Execution Model

The model of execution developed here will be used to predict the likelihoods of the various possible outcomes that may result from an actor's attempt to execute a given intention (perform a given task) in a particular situation. The modeling approach consists of defining *actor types* as necessary, distinguished by differing levels of ability, and describing their ability to execute the given intention (this same approach was used in the anesthesia project reviewed in Chapter 3). The different actor types are characterized by factors that affect the ability to perform the given task. While the factors that are relevant for distinguishing between actor types depend strongly on the characteristics of the system and the situation, there are some elements that are likely to be important in many domains, such as experience, training, and fatigue. If, in a particular situation, training and fatigue are the primary determinants of ability, then appropriate actor types might be Untrained, Fatigued, Untrained and Fatigued, and Normal (no particular ability limitations). The ability of an actor of a given type to execute the relevant intention is characterized by defining a mutually exclusive, collectively exhaustive set of possible outcomes – in the simplest case, successful or erroneous performance of the given task – and characterizing the likelihood of each of the possible outcomes using a probability distribution.

Ultimately, the execution model must calculate the probability distribution on outcome for each different actor type. This distribution is defined as the *outcome distribution*, denoted  $A_{i,j}$ , where  $i$  indexes actor types and  $j$  indexes possible outcomes;  $A_{i,j}$  is the probability that an actor of type  $i$  will perform action  $j$  when attempting to execute the given intention. In general, the actor's performance will be affected by the level of task demand, and the probability distribution describing the actions of each actor type is specified as a function of task demand. This is a measure of the demands that the system places on actors, such as the requirements imposed by the environment or by system design – a poorly designed system can make it more difficult to execute intentions, which may increase the likelihood of execution failures.

The outcome distribution,  $A_{i,j}$ , is found by first characterizing the outcome probability as a function of task demand, then specifying the probability distribution on task demand,

and integrating their product. The function that describes the abilities of each actor type is called an *outcome function*; it describes the outcome probability as a function of task demand and is denoted  $a_{i,j}(x)$  (note the distinction between the outcome functions,  $a_{i,j}(x)$ , and the scalar outcome distribution,  $A_{i,j}$ ). Thus  $a_{i,j}(x)$  is the probability that outcome  $j$  will result when an actor of type  $i$  is faced with the level of task demand  $x$ . Task demand may be a continuous or discrete variable here. The dimension of task demand is not specified here; in an actual application, the characteristics of the system and the task will determine which dimension is relevant. Typical dimensions will be measures such as task complexity and the availability of time or resources.

Figure 5.7 illustrates representative outcome functions for the simple case of successful or erroneous execution of an intention. The two curves represent two different types of actors, and give the probability of an error as a function of task demand. Only the error probability,  $a_{i,error}(x)$ , is illustrated here; for each actor type, the probability of successful execution,  $a_{i,success}(x)$ , is just one minus the error probability.

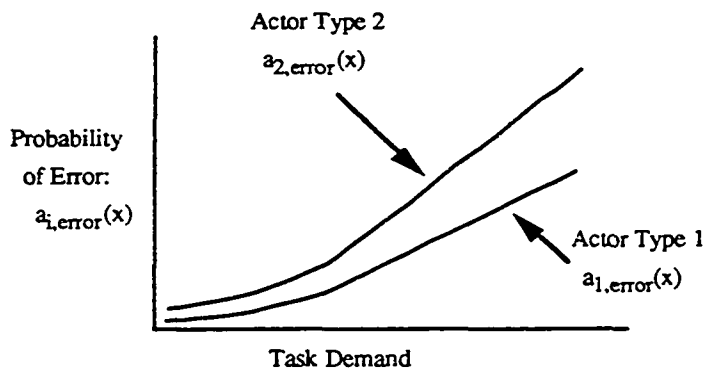


Figure 5.7: Illustrative outcome functions for two actor types.

If  $n > 2$  outcomes are possible, then instead of specifying only the probability of error, the probabilities of  $(n-1)$  outcomes must be specified (the probability of the  $n^{\text{th}}$  outcome is just one minus the sum of the probabilities of the others). Figure 5.8 shows a convenient way to display this information, with three possible outcomes: successful action, minor error, and major error. The probability of a given outcome at any level of task demand is just the vertical distance between the corresponding curves at that point. (While there is a natural ordering to the outcomes in this example, this is not necessarily always the case).

It is also necessary to define  $f_{TDi}(x)$ , the probability distribution of task demand that the system imposes on actors of this type in this situation. (This is a continuous or a discrete

distribution, to match the levels of task demand considered in the outcome functions.) The outcome distribution  $A_{i,j}$ , the probability that an actor of type  $i$  will perform the outcome  $j$  in the given situation, is calculated as the product of the distribution on task demand,  $f_{TDi}(x)$ , and the outcome function,  $a_{i,j}(x)$ , integrated over all possible values of task demand:

$$A_{i,j} = \int_{-\infty}^{\infty} f_{TDi}(x) a_{i,j}(x) dx$$

The probabilities of each possible outcome are calculated for each of the actor types<sup>27</sup>. This calculation must be performed for  $(n-1)$  outcomes, where  $n$  is the total number of outcomes possible. The probability of the  $n^{\text{th}}$  outcome is just one minus the sum of the probabilities of the others.

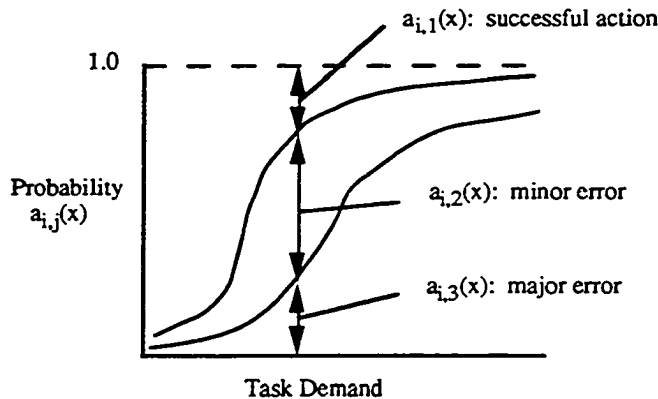


Figure 5.8: Outcome functions for three possible outcomes (one actor type).

The last piece of information required is the frequency of each of the actor types in the system, denoted  $q_i$ . The overall probability that an actor of any type will perform a particular action in the given situation is found by weighting the above results by  $q_i$ , the probabilities that an actor is of each different type, and summing over actor types<sup>28</sup>:

<sup>27</sup> If there are two or more distinct dimensions of task demand that affect the actor's performance, then it may be necessary to treat each of them explicitly. If so, the outcome functions can be specified as functions of all these dimensions, and the probability of outcome for each actor type would be found by integrating the joint probability distribution over all dimensions of task demand. For two dimensions, this is:

$$a_{i,j} = \iint_{y,x} f_{TDi}(x,y) a_{i,j}(x,y) dx dy$$

It is straightforward to generalize to more than two dimensions, though the computational burden may quickly become excessive.

<sup>28</sup> The probability distribution on actor type is characterized mathematically in the same way whether it characterizes the probability that a given actor is of each possible type, the fraction of time an actor spends

$$A_j = \sum_i q_i A_{i,j}$$

An actor's fundamental ability to successfully execute the intention affects the vertical position of the outcome function – as abilities increase, the probability of successful execution increases. In Figure 5.7, which graphs the probability of error, Type 1 actors clearly have greater ability. In Figure 5.8, an actor type with greater abilities than the one shown would have curves correspondingly lower, while one with poorer ability would have higher curves. (There is no reason why the ability functions for two different actor types could not cross, corresponding to one actor type performing better at some levels of task demand, and another performing better at other levels.)

Slips (and related lapses) are errors that are not caused by a particular lack of ability or exceptional task demand. However, it may be difficult to draw a clear distinction between errors that are slips and those that are caused by a lack of ability or an overly difficult task. Fortunately, such distinctions are unnecessary in this model – the outcome functions encode all the necessary information. Slips are represented by the fact that the probability of an error does not typically go to zero even for highly capable actor types at low levels of task demand, though they may be more likely for one actor type than another. The low, flat regions at the left of the outcome functions in Figure 5.7 characterize errors that would generally be considered slips.

The Yerkes-Dodson effect discussed above can also be captured by the outcome functions. In a situation where low task demand corresponds to a higher error rate, the outcome function for error will have greater values at lower levels of task demand. This yields a non-monotonic outcome function like that illustrated in Figure 5.9, which is treated the same as any other outcome function. The probability of successful execution, equal to one minus the probability of error that is graphed here, exhibits the characteristic inverted-U shape.

The distribution of task demand that is imposed on actors of a given type,  $f_{TD_i}(x)$ , captures the effects of system design and related issues – a poorly designed or constructed system that makes it difficult for an actor to carry out the appropriate actions will impose greater demands on the actor, shifting the mass of  $f_{TD_i}(x)$  toward higher values of task demand. For example, a poorly designed information system can overwhelm the operator with unimportant data, making it difficult to sort out crucial information. The

---

as each type, or the type distribution of a population of actors. Of course, the management implications of each of these situations may be very different.

distributions  $f_{TD_i}(x)$  can also differ by actor type,  $i$ , to reflect the fact that responsibilities may be allocated differentially according to actor type. For example, inexperienced actors may be assigned to situations that are less likely to impose high levels of task demand.

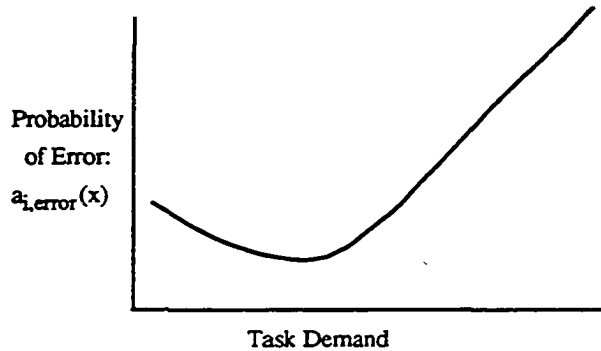


Figure 5.9: Outcome function displaying non-monotonicity.

#### Management Controls in the Execution Model

In one sense, execution failures are an inevitable consequence of human fallibility. But while it may not be possible to eliminate them entirely, execution failures certainly deserve the attention of risk managers, because it is often possible to reduce their frequency and severity. There are two basic ways in which management can influence task execution to reduce failures. The first is to reduce the difficulty of the tasks that actors face in order to reduce the probability of an execution failure<sup>29</sup>, altering tasks by changing the system itself or the ways in which it is operated. The second way is to increase the abilities of actors so that they are more capable of successfully executing the required tasks. This may mean changing the abilities of actors currently in the system, or eliminating low-ability actors and replacing them with others of greater ability.

One way for management to change task demand is to change the physical system itself to modify the tasks that must be performed (e.g., by automating a previously manual task). These physical changes can affect factors like task complexity and the available resources (such as the type, condition and availability of equipment). Improving the principles of good system design is beyond the scope of this research (Norman, 1988, discusses a variety of common design failures and offers some simple and useful

<sup>29</sup> With the caveat that in situations where the Yerkes-Dodson inverted-U hypothesis applies and outcome functions are non-monotonic, reducing task demand might increase error probability. In such circumstances, the goal would be to achieve whatever levels of task demand lead to the lowest error rate.

principles for good design). However, it is important to recognize here that system design, ergonomics, and the organization of the work environment can have major effects on system risk through the demands placed on actors (as well as, of course, the reliability and interrelationships of system components themselves, which is the domain of current risk analysis techniques). It is critically important that the design of the system takes into account the abilities and limitations of users and operators. This is not to say that every system can (or should) be designed to be foolproof, but that by accounting for the abilities and limitations of actors in system design, it may be possible to prevent later errors that have the potential to lead to system failure. The model described here provides a way to quantify the risk implications of alternative design strategies.

Another way that management can affect task demand is to change the ways in which responsibilities are allocated, thereby altering the tasks that individual actors must perform. Changes in the number and type of overall responsibilities for an individual may affect the difficulty of a given task, (e.g., increasing staffing levels may reduce the workload on each actor, allowing more time to complete a given task and increasing the chance of executing it properly). By changing the allocation of responsibilities among actors, management can affect the tasks and the levels of task demand that actors of each type face. This approach depends on the ability of management to distinguish between actors of different types. For example, it may be easy for management to identify inexperienced actors and organize tasks so that they are unlikely to face high levels of task demand, but may be more difficult to identify substance-abusers or actors who are stressed by events outside the work environment, and it will thus be difficult to control the levels of task demand that these actor types face. All of these management strategies that can affect task demand are reflected in the model by changes to the probability distributions on  $f_{TD_i}(x)$ .

Actors' abilities, on the other hand, can be affected by factors such as natural or innate ability, experience, training, fatigue, stress, distraction, and substance abuse. Management can influence actors' abilities in two ways. The first is to improve the abilities of the actors who are already in the system, by changing the factors that affect ability. The second is to change which actors are in the system, by identifying actors according to these factors, and selecting and screening them on this basis.

Mechanisms that improve the abilities of actors within the system, such as training and work-schedule changes (which affect fatigue and stress), move actors from one type to another, to decrease the number and frequency of actors of lower ability types and



correspondingly increase the types with greater abilities. In the model, these mechanisms may change whether or not an actor is of a particular ability type, but do not change the outcome functions that characterize the abilities of each type<sup>30</sup>. Selection and screening mechanisms change which actors are in the system, removing actors of types that have relatively poorer abilities and/or adding actors of types that have greater abilities. As before, the difficulty of distinguishing some actor types, (substance-abusers, etc.), may make it difficult to improve or eliminate them. The management changes that affect actors' abilities are captured in the model by changes in  $q_i$ , the distribution on actor types.

Example: Driver's Performance with the Execution Model

The execution model can be illustrated with the hazardous materials transport example used before, but it applies to different circumstances than the decision situations addressed by the previous models. The most obvious execution failure in the transport system is the driver's ability to avoid an accident<sup>31</sup>. The execution model developed here makes it possible to directly study the factors that affect accident probability. The possibility of an accident arises from the interaction between task demand (the demands the system imposes on the driver) and the driver's ability<sup>32</sup>. Four possible driver types are defined, based on whether the driver is experienced or inexperienced, fatigued or not fatigued. Shorthand notation is used to refer to these four types: I for Inexperienced, E for Experienced; upper-case F for Fatigued, lower-case f for Not Fatigued:

<u>Driver type</u>	<u>Description</u>	<u><math>q_i</math></u>
1. E-f	Experienced and Not Fatigued	.40
2. E-F	Experienced and Fatigued	.10
3. I-f	Inexperienced and Not Fatigued	.40
4. I-F	Inexperienced and Fatigued	.10

An Experienced driver is defined here to be one with more than five years of experience (not necessarily all with this company). A Fatigued driver is one who has driven more

<sup>30</sup> There is no reason in principle that the model could not account for management changes by adjusting the outcome functions for fixed frequencies of actor types, rather than changing the frequencies of actors of fixed types. However, it seems simpler and more natural to change the frequencies of actor types, and this causes no loss of generality, because it is possible to define as many types as necessary, some of which may have zero population before or after a management change.

<sup>31</sup> Here the model is used to look at an outcome that leads directly to (in fact, is) system failure. In many cases, it will be used to examine an outcome that could, but does not necessarily, lead to failure, such as whether a maintenance operation is carried out properly or is flawed. In that case, to find the contribution to risk, it is necessary to develop a system model that characterizes the ways in which the action (flawed maintenance) interacts with other events and components to lead to system failure.

<sup>32</sup> In fact, at this level, the driver does make decisions on a minute-by-minute basis, such as whether to change lanes, how hard to brake, etc., and then attempts to execute the intentions thus formed. But modeling decisions at this level is likely to require much effort for very little return, so all these "micro-decisions" are subsumed in this model of driver's ability.

than 16 of the previous 24 hours, or who has not had at least a 2 hour break in the last 8 hours. The company controls the overall work schedule, which can contribute to fatigue, but the hour-by-hour scheduling is left to the drivers, who often prefer to concentrate their driving as much as possible to get longer stretches of time off. Drivers are fatigued approximately 20% of the time, and about half of the drivers are inexperienced. These two factors are independent, leading to the given values of  $q_i$  that characterize the fraction of each of these types<sup>33</sup>.

Task demand is measured by a constructed scale of driving difficulty that includes visibility, road and weather conditions, etc. The distribution of task demand imposed on the actor by the system is characterized by  $f_i(x)$ , represented here as a triangular distribution to reflect the fact that more difficult driving conditions are less common. The distribution is the same for each of the four driver types, because all drive in the same conditions, so  $f_i(x) = f(x)$  for each driver type  $i$ , as illustrated in Figure 5.10.

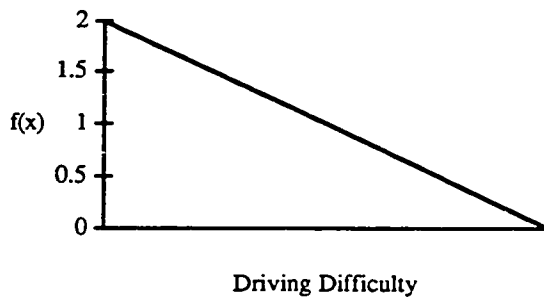


Figure 5.10: Probability distribution on task demand for the hazardous materials transport example.

The outcome functions describing the abilities of the four driver types are illustrated in Figure 5.11, which shows accident probabilities as a function of driving difficulty. At low driving difficulty, experience makes little difference in performance, but fatigue is important. As driving difficulty increases, the performance of inexperienced drivers quickly deteriorates, so that in very difficult driving, inexperienced drivers perform much worse, particularly if they are also fatigued. The performance decrement for experienced drivers is less extreme, even when fatigued. The functional form of these curves is

<sup>33</sup> In this example, experience and fatigue are assumed to be independent. In other situations, the relevant dimensions may be independent, or either positively or negatively correlated.

$$a_{i,a}(x) = k_1 e^{k_2 x}$$

with parameter values corresponding to level of experience and fatigue as follows:

	Fatigued	Not Fatigued
Experienced	$k_1 = .0002; k_2 = 1$	$k_1 = .0001; k_2 = 1$
Inexperienced	$k_1 = .0002; k_2 = 2$	$k_1 = .0001; k_2 = 2$

Accident probabilities for the four driver types are calculated by integrating the products of outcome functions and the driving difficulty distribution, and are given in Table 5.5<sup>34</sup>.

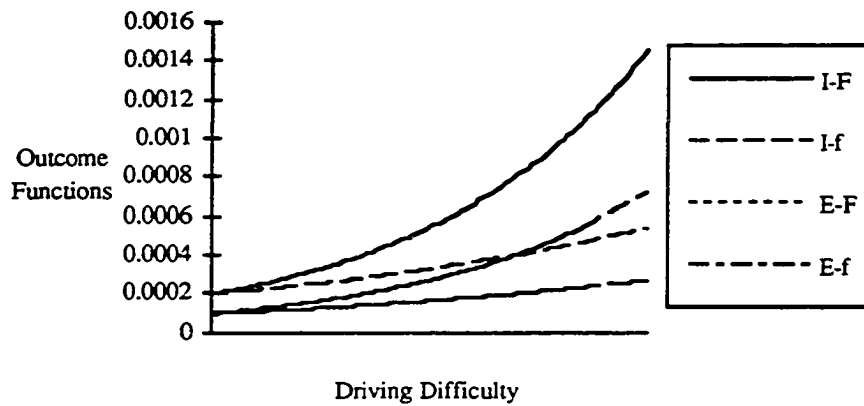


Figure 5.11: Outcome functions for four driver types for the hazardous materials transport example.

Driver type	$q_i$	$p(\text{accident})$
1. Experienced, Not Fatigued	.40	.000144
2. Experienced, Fatigued	.10	.000287
3. Inexperienced, Not Fatigued	.40	.000219
4. Inexperienced, Fatigued	.10	.000439

Table 5.5: Accident probabilities for four driver types.

Inexperienced, Fatigued drivers pose the largest risk – three times that of Experienced, Not Fatigued drivers. However, the risk caused by Inexperienced drivers is less extreme than might be expected, because very difficult driving conditions, where a lack of experience is the most problematic, are relatively uncommon. A driver who is Inexperienced and Not Fatigued poses less risk than an Experienced, Fatigued driver.

<sup>34</sup> In fact, accident probability also depends on the other factors modeled in previous examples - driving speed, weather conditions, and brake condition. The accident probabilities calculated here correspond to the best settings of these other variables, to illustrate the effect of driver type. The overall risk model that captures all these effects and their interactions is developed in section 6.3.

Fatigued drivers pose twice the risk for an accident as drivers who are Not Fatigued, regardless of experience.

The effect of risk management measures in this model will be to alter the fraction of drivers of each type; the accident probability associated with each type will not change. For example, efforts to reduce fatigue will not affect the abilities of a driver who is fatigued, but will reduce the fraction of drivers who are fatigued. This model can also be used to evaluate prospective risk management measures. Two proposed management strategies are described below, and each of these risk management measures will change the vector of driver types as noted.

Work schedule changes can be designed to reduce the probability that a driver is fatigued. This includes reducing the total number of trips that drivers must make, as well as requiring breaks and time off to restrict the number of consecutive hours that drivers may be on the road. While this strategy should significantly reduce the number of fatigued drivers, it will not eliminate them entirely, because drivers may violate these policies occasionally to increase their consecutive hours off. This change in work schedules will decrease the frequency of fatigued drivers from 20% to 10%, changing the vector of driver types from [.4, .1, .4, .1] to [.45, .05, .45, .05].

By improving compensation and promotion policies to attract experienced drivers and reduce turnover, the fraction of drivers who have significant experience can be increased. Because it takes some time to acquire new drivers, the effects of this strategy will not be felt immediately. This policy change will ultimately increase the fraction of drivers who are experienced from 50% to 60%, changing the vector of driver types from [.4, .1, .4, .1] to [.48, .12, .32, .08].

This simple example serves to illustrate the execution model and the kinds of issues that it can address, such as the importance of the interaction between the abilities of different actor types and the demands placed on them by the system. Some applications may require more detail to accurately represent the situation - for example, it may be appropriate to look at a greater number of actor types, such as those whose ability is impaired due to the use of alcohol or drugs, or those who simply have poor natural ability (slow reflexes or easily distracted). It may be useful to look at more than two possible outcomes, perhaps distinguishing between minor and major accidents, and the distributions on task demand may depend on actor type. Any of these analyses would proceed in the same way as the example has, but tracking the larger number of variables

required to reflect these finer distinctions. Of course, such a model can also be used to evaluate a wide variety of risk management strategies (and variations on those strategies) in addition to those suggested above.

### **5.5 Management and Organizational Control Mechanisms**

This section looks at the relationship between management and action, as did the previous modeling sections, but is organized according to the perspective of management control mechanisms, in part because many of these are familiar to managers of complex systems. There are a number of mechanisms that management can use to influence the actions and errors of individuals, some of which affect action in more than one of the modes modeled in the previous sections. Clearly, which management strategy is best for preventing error depends on the intention and ability processes that are at the root of it. If ability limitations cause error, then strategies that address intention formation, (e.g., incentives) will be ineffective; if the problem is caused by goal conflict, then training to improve the actor's abilities will not help, though indoctrination that changes preferences might. The following mechanisms have been identified in the previous sections and are discussed more fully here:

- Selection and screening
- Training
- Policy and procedure
- Work demands
- Information systems and cognitive aids
- Incentives
- Organizational culture
- Design of system and equipment
- Resource constraints

The number of mechanism that management can use to influence the actions of individuals in a complex system may be large, and this is certainly not an exhaustive list, but it does give a useful overview of many of the more important ones. These management mechanisms may in some cases be blunt tools, because they can affect action at different points in the system simultaneously, even action in different modes that are governed by different processes. While it may be possible to target some mechanisms, such as incentives, at particular actions that management would like to influence, other mechanisms, such as selection and screening, which can change the individuals who are in the system, may affect a wide range of behavior in many different situations. Management mechanisms can affect actors' abilities to execute intentions,

which can affect action in many situations; they may also affect preferences and risk attitudes, influencing expected utility and bounded rationality decisions in other situations; and they can affect the rule bases that govern behavior in still other situations. It is possible that these effects will complement one another by reducing risk at each point, but this will not necessarily be true. For example, while selecting for actors of greater experience and skill may be expected to improve performance on some dimensions (through greater abilities, knowledge, etc.), such actors may also be more likely to ignore organizational policies that conflict with their own judgment, which might increase risk. In any case, it is important to look at all the effects of a proposed management policy when evaluating it, to prevent under- or over-estimating its effects.

Of course, even with these control mechanisms at its disposal, management is not able to exert total control over the individuals' actions. At best, these mechanisms are imprecise and incomplete, which is in part what makes the probabilistic techniques of this framework so useful. For example, although selection and screening mechanisms allow management to influence the abilities of individuals by selecting for those with high ability, this process is inexact (it is difficult to measure an individual's capabilities), and does not give complete control (it is impossible to find individuals whose abilities are unlimited). In addition, these mechanisms can be costly; not only is the selection process itself expensive, but more highly qualified individuals invariably demand higher compensation. However, these mechanisms can nonetheless have an influence, and this framework can help to capture it.

### Selection and Screening

The selection of individuals for participation in a system and screening to eliminate those who are unsuited can be important tools for risk management, because the attributes of individuals can often affect the actions they will choose and their ability to execute those actions. Selection and screening criteria may be based on education, experience, interview, job performance, simulation tests, etc. By maintaining good employee relations and compensation, the organization may also be able to limit turnover, retain valuable employees, and increase the average level of experience and ability of its employees.

The most obvious effects of these approaches may be on individuals' abilities to execute actions. Some individuals are simply better suited for a given task than others, as a result of greater natural ability for the task, or of greater experience with similar tasks. Selecting individuals for their ability is nothing new, but explicitly accounting for its

effect on risk may lead to criteria that focus on factors that are important contributors to risk.

To affect actions that are rule-directed, criteria can also be established to select individuals who have a well-suited rule-base. The medical profession is a good example of such an area. Many actions are determined by medical protocols, which specify appropriate tests and treatments according to patient characteristics and symptoms. The decision maker is not necessarily explicitly aware of the alternatives that may be available or their possible outcomes, but simply follows the protocol. Medical protocols are often quite elaborate and typically exist in written form, but physicians often base decisions and actions on their mental rule-base. Physicians and other medical personnel with up-to-date, accurate rule bases are crucial to safe and effective health care delivery. While it is difficult to observe an individual's rule base directly, information about it can often be inferred from education and experience or selective testing.

While it may be difficult to attempt to change a given individual's preferences, it should be possible, at least in principle, to select individuals whose preferences are best suited to the requirements of the system. In practice, it may be difficult to do this with much precision, but it is certainly something that can be and is done qualitatively. A conservative investment fund may rightly have little interest in hiring a junk bond speculator, even a very good one, because her taste for risk would likely be incompatible with the firm's.

There is a tradeoff between selecting relatively unskilled vs. highly qualified individuals. The appropriate solution depends on the system and the requirements of the position to be filled, and each strategy has advantages and disadvantages. Unskilled, inexperienced individuals come at lower direct cost, and with significant monitoring and direction, their actions may be more predictable. This may be desirable, particularly in a rote position where it is important that actions be standardized. On the other hand, highly qualified individuals who are allowed the freedom to decide how best to pursue goals may perform much better in situations where it is necessary to develop good solutions to novel situations. Disaster can result from selecting unqualified individuals and allowing them too much freedom with insufficient direction, as in the case of the grounding of the Exxon Valdez, where an inexperienced, unqualified pilot navigated the supertanker onto a reef as the captain slept below (Moore, 1994).

### Training

Similar to the effects of selection and screening, training can increase actors' knowledge, information, and abilities, though in this case, by changing the attributes of existing actors, not replacing them with different actors. Training is often the most effective way to tailor the attributes of the individual to system requirements, though it cannot substitute entirely for selection and screening mechanisms. It can be valuable for both new and long-time employees. Where behavior is rule-directed, training may augment or amend the actor's rule base, allowing the individual to perform quite well in a system that is too complex to understand fully. For expected utility or bounded rationality decision making, they can improve the actor's understanding of the system so that decisions are based on more accurate knowledge. Training can help compensate for a lack of experience, and can offset a lack of natural ability, though it may not be able to entirely counteract weaknesses in these areas.

Of course, the type of training that is most effective depends on the system, the responsibilities of the actor, and the types of situations they might encounter, but some general principles apply. On-the-job experience can effectively teach some skills, but explicit training programs, such as simulation and training drills, are particularly effective for teaching individuals about rare but important situations and for those in which trial and error learning are unacceptable, as in technological systems where errors can be disastrous. Simulation exercises allow individuals to gain experience and competence in situations where errors do not have dire consequences. While simulation may be more difficult to implement, it can also be more effective than other training procedures, because it offers hands-on experience.

### Policy and Procedure

The setting of policies and procedures consists essentially of management instructing individuals in how they should act. This can be thought of in the context of a rule-based model: management develops a set of rules which it transfers to the actor, specifying which action to take in what circumstances. This raises the basic problems associated with rule-directed action discussed above: errors can occur because of flaws in the rules, or because a situation is identified incorrectly, leading to the application of the wrong rule. The latter can be addressed in part by building into the rule base the characteristics that differentiate situations and providing the actor with the information necessary to make this distinction.



However, a problem with setting policies and procedures to direct action is that an individual may not comply with the organization's rules at all, choosing instead to determine action in a different way, such as maximizing their own utility, or by following their own set of rules. So along with the rules that specify appropriate actions, management may have to monitor action and create an explicit or implicit incentive structure (rewards and punishments) to induce individuals to follow its rules. When policies and procedures specify actions that are consistent with or do not affect the actor's self interest, monitoring behavior and incentives may be relatively unimportant.

Providing a rule base to guide behavior can be very helpful, as in situations where a system is so complex that individuals would be unable to determine the appropriate action (leading to cognitive processing failure). Rule books that specify how nuclear plant operators should deal with unusual events are a good example of this.

### Work Demands

Excessive demands placed on the actor by the work environment can interfere with the proper execution of intentions. This may be a matter of excessive task demands: a task that is too difficult, or simply that the quantity of work is too great for the time available, causing stress and fatigue. Also, factors not directly related to the task to be performed, such as outside stress, distraction, fear, etc., may nonetheless interfere with it. This interference may take the form of inducing slips and lapses that might not otherwise occur, or it may simply making the required task too difficult to perform at all.

In addition to affecting the execution of intentions, the demands of the work environment may also influence the formation of intentions for action. As noted in the bounded rationality model, individuals often do not have the resources necessary for objective rationality in real-world situations. When the demands of the work environment increase, time pressure and performance pressure may spread the actor's cognitive resources even thinner, compromising the quality of decisions. In the extreme, the actor may retreat to overly simple or flawed strategies, ignoring feedback from the system, until system failure is imminent. The other side of this is that a work environment which does not place undue demands or stresses on individuals will allow them to perform their functions more successfully. Note, though, that the Yerkes-Dodson hypothesis implies that reducing the demands of the work environment improves performance only to a point, below which a lack of motivation, arousal, and challenge can decrease effectiveness.

### Information Systems and Cognitive Aids

An actor's information and cognitive abilities can have considerable effects on decisions and action in the three modes of intention formation considered in this framework – rule-based, bounded rationality, and expected utility maximization<sup>35</sup>. Information that is unavailable or incorrect can lead the actor to misinterpret the state of the system, causing action that is inappropriate for the circumstances. Limitations on actors' cognitive abilities can have similar consequences. The explosion of the space shuttle Challenger is a highly visible instance in which a failure to communicate crucial information allowed a launch decision that caused the death of all 7 crew members and a crisis of faith in the organization (Report of the Presidential Commission on the Space Shuttle Challenger Accident, 1986). Management strategies that improve the actor's information and cognitive abilities will lead to better decisions. To achieve this, management can use information systems (systems or procedures for discovering, storing, and communicating information) and cognitive aids (mechanisms for improving and extending actors' cognitive processing abilities). While information systems and cognitive aids are conceptually distinct, they are discussed together here because they can often be addressed with the same or similar mechanisms.

It may be possible to improve the actor's information about the current state of the system at the time of a decision with physical monitoring and information transmission systems, and by better organizational communication channels. These include formal and informal information organizational structures for storing and transmitting information, as well as modern technological solutions, such as a management information system. A caution raised by the bounded rationality theory is that more information is not always better. Humans have a limited ability to absorb and process information, and will sometimes ignore it altogether if there is too much information or if it is not organized in a useful way, so the effective organization and presentation of information may be as important as its availability.

Cognitive aids encompass a wide range of mechanisms, such as management information systems, artificial intelligence systems, normative decision systems (e.g., Regan, 1993), computers, calculators, etc.; as well as systems for the organization and processing of information, such as an accounting or record-keeping system. They can significantly

---

<sup>35</sup> Knowledge, as distinct from information, can have very similar effects, but is primarily affected through selection and screening mechanisms and training, which are discussed elsewhere in this section.

increase the reliability of cognitive processing performance, and can greatly extend human abilities to handle complex problems and information.

### Incentives

Incentives may be one of the easiest factors to manipulate and are one of the traditional tools that management uses to affect action; rewards and punishments encourage desired actions and inhibit others. The problem of determining the incentives that will induce an individual to act in the best interest of the organization is analogous to the principal-agent problem discussed in section 5.1.

Rewards and punishments may be associated with good and bad outcomes in the overall system, with partial failures and near misses, or directly with the desired and discouraged actions. The more closely incentives can be associated with actual behavior, (as opposed to system outcomes that may be only probabilistically associated with behavior) the more effective they will be. This is because system outcomes can be affected by a wide variety of factors, so incentives associated with them are not as closely related to behavior, and have less effect on it. Unfortunately, associating incentives more closely with behavior usually comes at a cost. This question comes down to trading off the greater effectiveness of rewarding outcomes more closely related to behavior against the cost and accuracy of monitoring (see Ouchi, 1979; Eisenhardt, 1985). Associating incentives directly with system failure may be ineffective, because system failure is (hopefully) a rare event, and even behavior that significantly increases failure probability will usually not lead to failure (Levitt, 1975).

The expected utility model illustrates another important consideration in developing an incentive system, which is that the incentive be large enough to overcome whatever intrinsic incentives may induce undesirable behavior. Often, the reason incentive programs are necessary is to counteract the effects of natural incentives, such as an all-too-human inclination to do things in the easiest way rather than the safest.

In addition to formal incentives that actually change the consequences to the actor (bonuses, punishments, performance evaluations), informal incentives, such as peer pressure, social approval and disapproval for specific behaviors, and informal rewards may also be effective (these are often thought of as components of organizational culture, discussed below). These act in much the same way as formal incentives, though they pose an additional difficulty in that they are typically more difficult to quantify and may be difficult to influence. If informal incentives are strong compared to formal ones, then

it is particularly important to include them in the analysis; if they are much weaker, it may be safe to ignore them.

### Organizational Culture

Organizational culture can have a significant effect on system risk, but almost all the research done in this area has been qualitative (see, for example, the work on high reliability organizations by Roberts, 1990; Weick, 1987). Organizational culture is difficult to measure, and its effect on risk even more difficult to quantify. Much more work in this area is necessary before the effects can be fully understood (if that ever happens), but since it may be significant, it is worth considering here, even if the treatment must be approximate. The effect of culture on the physical system, like the effect of management, must occur through the actions of the individuals in the system.

One of the plausible ways in which culture may affect action is through the risk attitudes of actors. A "safety culture," like that often discussed in the context of the nuclear industry (e.g., IAEA, 1991), may be characterized by individuals who act rationally and are risk averse; who prefer to act more conservatively and are unwilling to take risks. This can be quantified in the risk preferences and utility functions of an expected utility model. Indoctrination and socialization processes are one way to affect an actor's preferences, by transferring the organization's values and preferences to the individual (this is goal congruence in the organizational behavior literature). These processes can be particularly important in military organizations, for example, but they are present to some extent in almost all organizations. Alternatively, a strong safety culture may be better characterized as one in which action is rule-directed. In such a culture, individuals' behavior is governed by rules that are established by the organization, and actors resist violating those rules – making it an "everything by the book" organization. Of course, an organization such as this is only as good as its "book"; if rules are incorrect or if new situations arise, a by-the-book culture may be of little help.

Organizational culture is difficult to define and difficult to control, certainly much more so than implementing a procedure or an incentive program. Nonetheless, while it may be difficult to turn around an entrenched culture, it is not impossible to make some changes. Serious management attention to issues like safety and risk can bring about a change in the attitudes of the individuals in the system. This may involve many of the other mechanisms that have been discussed here, such as punishing risky behavior, setting and training for appropriate safety procedures and changing work demands to make it

possible to observe them, as well as selection and screening criteria designed to retain individuals who are committed to safety and eliminate others.

### Design of System and Equipment

The design of the physical system and equipment affects system risk, of course, through the reliability of components and their interactions, but these interactions are adequately captured by traditional risk analysis techniques. What is not captured as well is the effect of design on the abilities of actors to develop and implement appropriate plans of action. While design is not the focus of this research, and I must leave the development of the fundamentals of good design to others, it is important to recognize that system design can cause errors or help to avoid them. Norman (1988) attributes many system problems to design, and Perrow (1984) argues that many errors in complex systems are "forced" by the system – that the design of the system makes them all but inevitable.

While it is probably not possible, and perhaps not even desirable, to attempt to design systems to prevent all errors, it may be possible to prevent many potentially disastrous errors by accounting for the limitations of users and operators in system design. This can be as simple as attention to ergonomic and human factors issues in design, such as making monitors and equipment visually distinct and easy to understand. It can also mean using automated systems to assist in or take over some of the functions of individuals, as with the information systems and cognitive aids discussed elsewhere in this section. This may be particularly attractive for tasks such as monitoring, where automated systems can be very reliable and human beings are notoriously error-prone. System design can also affect how responsibilities are allocated, through the placement and accessibility of system controls. Unfortunately, the financial forces associated with constructing large, complex systems can encourage shortcuts and cost saving measures in design and construction. While this is not necessarily bad, the effects of this on the requirements for operation are too often ignored. In a well designed and constructed system, it may be possible to cut some corners in operation (like reducing staff) without severe consequences, but a physical system that is not particularly robust can be extremely vulnerable to weaknesses in operation.

### Resource Constraints

Resource constraints, especially budget constraints, can affect almost all of the other management controls that have been discussed, because almost all of them require resources to implement. A lack of resources can lead to selection and screening criteria that accept less-qualified individuals to save labor costs, and can affect the workload and

time constraints on actors, if staffing is cut to reduce costs. Budget constraints can limit the resources available for incentive programs, training, implementing policy and procedure, and information and communication systems, as well as forcing compromises in the design and inherent robustness and safety of the physical system and equipment. Limited resources may even affect organizational culture, leading to a culture of cutting corners, lack of commitment, etc. Constraints on the resources available to actors may limit the number and quality of alternatives from which they can choose when faced with a decision. Resource constraints may inadvertently make good alternatives infeasible; conversely, increasing the resources available to the actor may make attractive alternatives feasible. In general, tight resource constraints limit the organization's ability to carry out its functions, one of which is to reduce the risk of system failure. Research on organizational slack (resource surplus) and effectiveness, while it does not focus specifically on risk, generally finds that slack increases effectiveness, though in the long run it can encourage complacency and decrease performance.

### **5.6 Choosing between Models of Action**

As discussed at the start of Chapter 4, three levels of structure are important in this framework – the physical system, the actions (and errors) of individuals within the system, and system management. Management factors influence the actions of individuals, which in turn affect the performance of the physical system. The analysis, however, proceeds in the opposite direction, starting with the physical system, then identifying actions that affect it, and finally the management factors that can influence action. But since individuals act in several different modes, it was necessary to develop four different models – expected utility, bounded rationality, rule-based, and execution models – to characterize the link between management and action. So before modeling management effects on a given action, it is necessary to determine in which of the four modes the individual acts, in order to select the appropriate model. As discussed in Chapter 4, it may in some cases be appropriate to model both the intention formation and the intention execution processes associated with the same action; however, in most cases, that will probably not be necessary. While it is difficult to establish ironclad rules for which model is most appropriate in a given situation (for many of the same reasons that it is difficult to model human behavior to begin with), it is possible to develop some guidelines that will help. Not surprisingly, this is similar to the discussion of section 4.4, which led to the selection of these four models. Figure 5.12 illustrates the factors affecting the choice between models of action in a given situation.

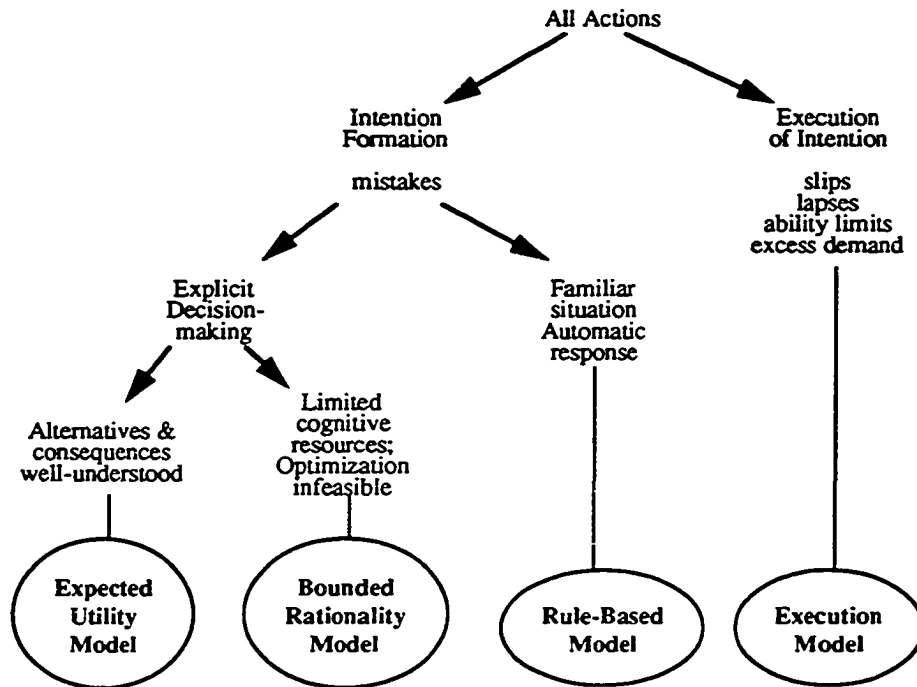


Figure 5.12: Factors affecting choice of model of action.

The first step is to determine whether the action to be modeled involves the formation of an intention or the execution of a given intention. The distinction is not always clear; the execution of high level intentions may require the formation and then the execution of lower level intentions, as when the decision to shut down a nuclear reactor necessitates subordinate decisions about the sequence of steps in shutting down the system, etc. This really comes down to the question of setting the level of detail in modeling; intention formation is modeled when it is warranted – when doing so captures important effects, such as incentives that might be changed to affect the behavior. Below that level, behavior is modeled simply as the execution of the higher level intention.

Action in any situation can be thought of as a two-step process of developing and then executing an intention, but in most circumstances it is not necessary to model both parts of this process. It is usually possible to identify a priori which step in this process is likely to break down in a way that can cause risk to the system, and thus which step must be modeled explicitly. Making the distinction between situations in which intention formation is key and those in which the execution step is more important is equivalent to determining whether the errors that might occur are mistakes or execution failures (again,

I use execution failures as a general category within which slips and lapses are special cases). This can also be viewed in light of Rasmussen's Skill-Rule-Knowledge framework. If the action occurs in Rasmussen's skill-based mode, in which the actor follows a pre-defined script that specifies actions for routine tasks, the possible errors can be classified as slips, lapses, ability limitations, or excess demand, and the execution model should be used. This model is appropriate for situations in which the actor's intention is not at issue, but only whether she will successfully execute it; where the actor's ability or the demands put on the actor by the system, the environment, and the situation may interfere with correctly executing an intended action.

On the other hand, if the error to be modeled is a mistake, in which the actor may develop and implement the wrong plan of action, then one of the other three models will be appropriate, and the actor's process of intention formation must be considered<sup>36</sup>. If the process involves explicit decision-making, corresponding to action in the knowledge-based mode, then either the expected utility or the bounded rationality model will be appropriate. The expected utility model will be most appropriate in situations where the full set of available alternatives is readily apparent to the actor, and when the choice of action has clear and significant effects on consequences to the actor. The rational model is particularly appropriate where the "mistake" may actually be a reasonable choice from the actor's own perspective; this is an instance of goal conflict.

The bounded rationality model is appropriate for situations in which cognitive demands are so high or cognitive resources such as time and attention are so limited that optimizing the decision is not feasible. It also addresses the need to define alternatives as well as select among them, and may be a more appropriate descriptor of action when alternatives are not well-defined, where decision-making requires that the actor first discover or develop alternatives.

The rule-based model is likely to be a good descriptor of action in many situations. Reason (1990a, p65) points out that "human beings are strongly biased to search for and find a prepackaged solution at the RB [rule-based] level before resorting to the far more effortful KB [knowledge-based] level." Much, if not most, behavior in familiar situations is rule-based; Newell (1987) and Cohen (1991) both describe individual and organizational learning as the acquisition and refinement of rules that direct action. For

---

<sup>36</sup> It is possible to have a situation in which both a mistake and an execution failure are possible. In this case, it may be necessary to use two models: one of the three models of intention formation, to determine the likelihood that the actor will develop the correct intention, and also the execution model to determine the likelihood that the intention will be successfully carried out.



example, physicians routinely employ rule-based strategies in diagnosing and treating patients: sets of rules – called protocols – specify how to deal with particular situations. They seldom explicitly consider alternatives, probabilities and consequences to reach a decision, because in most cases it is unnecessary, and in fact, rule-based decision making may be more effective because of the amount of experience encoded in the rules. The nuclear industry makes rule-based behavior even more explicit, with established, written catalogs of rules that specify the appropriate response to a wide variety of possible situations. Actors in many other situations employ similar rule-based techniques, even if they are not so explicit. Rule-based behavior may be particularly likely when action does not have a significant effect on consequences to the actor, or when the link between actions and outcomes are not clear, making explicit decision making processes difficult.

Reason's Generic Error Modeling System (GEMS, 1987a, 1990a) may also help to shed some light on the mode in which the actor will perform in a given situation, and thus which of the models is most appropriate. According to the GEMS model, humans try to work in the lowest possible of the three levels: skill, rule, and knowledge. So if there is a "script" to support it, an actor will work at skill level, errors will be execution errors, and the appropriate model for the situation will be the execution model. If no script is available, or if the actor must choose from several alternative scripts before implementing one, then the actor will try to use rule-based reasoning to determine action, and the rule-based model will be the most appropriate model. If all else fails, if there is no script and no set of rules that the actor can use to guide behavior, then as a last resort, the actor will turn to the difficult knowledge-based mode of reasoning, and either the bounded rationality or the expected utility model will be the most appropriate.

An interesting twist on the choice of models is the fact that in some cases, management may actually be able to influence the mode of behavior in which the individual acts, and thus which of the models is appropriate. This is particularly true among the three modes of intention formation: rule-based, bounded rationality, and expected utility maximization. For example, instituting new incentives may prompt a shift to utility maximization, or eliminate them may shift actors away from utility maximization and into a rule-based decision mode, with the rules provided by the organization. By selecting for experienced actors who have encoded a thorough understanding of the system as rules, it may be possible to keep actors out of the error-prone knowledge-based (expected utility or bounded rationality) mode. Simply providing a management-sanctioned rule base may have a similar effect. In order to evaluate a strategy in which

management influences the individual's mode of action, it will be necessary to construct a new model of action that is consistent with this new behavior mode.

### **5.7 Summary of Chapter 5**

In this chapter, I have developed four models of action that characterize the ways in which management factors affect the behavior of individuals. These four models, the expected utility model, the bounded rationality model, the rule-based model, and the execution model, can be used to model behavior in situations involving both intention formation and execution, ranging from fully rational decision making to skill-based performance. Each of the models was illustrated with an example of an action that affects risk in an illustrative hazardous material transport system.

This chapter also discussed the ways in which management control mechanisms can influence action, and some issues involved in implementing these control mechanisms. It concluded with a discussion of how to choose which of the four models of action is most appropriate in a given situation. Chapter 6 will develop the remaining elements of the framework and pull together all the pieces, illustrating the use of the framework by combining the examples of this chapter into a system model that characterizes the combined impact of multiple actions on system risk.

## **Chapter 6**

### **Synthesis – Using the Framework to Manage Risk**

This chapter completes the development of the framework. Section 6.1 develops the link between action and the physical system, the final quantitative piece of the framework. Section 6.2 is a brief, step-by-step set of instructions for applying the framework in an actual application. Section 6.3 uses the hazardous material transport example introduced in the previous chapter to illustrate the integration of the various pieces of the framework.

#### **6.1 Linking Action to the Physical System**

The final piece of this framework is the link between the action of an individual and the physical system. In making this link, it is not possible to construct general models like those that relate Action and Management, because every system is different, and the effects of actions on the physical system are unique to that system and situation. Fortunately, such new models are not necessary, because the effects of actions can be handled satisfactorily with the mathematical techniques of current risk analysis practice. The first part of this section describes the two methods used to model the link between action and the physical system, and the second part discusses how an understanding of the primary points at which individuals interact with the system can help in identifying actions that may affect the performance of the physical system.

The basic components of the probabilistic risk analysis methodology are:

- 1) a functional model of system configuration
- 2) failure mode events – individual component reliability
- 3) probabilistic dependency in component failures
- 4) external events – loads on the system
- 5) system dynamics – deterioration in crisis situation

Together, these components model the functioning of the physical system to calculate the probability of its failure. The decisions and actions of individuals in the system, however, can affect the system at each of these points. For example, an operator may shut down a redundant component, poor maintenance can affect component reliability, and dependencies can be introduced through similar operating and maintenance policies, an actor may make an error in operations, and actions taken in a crisis may affect the dynamics of the accident sequence.

The actions of individuals that affect physical system performance fall into two categories, depending on how they affect the system. An action may actually be a *failure mode event*, a direct element of system failure analogous to an individual component failure. Or it may act in the same way as an *external event*, affecting the likelihood of component failures without itself being a failure mode event. While they are treated mathematically in exactly the same way as conventional external events, actions that fall into this second category are typically not "external" to the system. To make the distinction clear, human actions of this type are referred to as *actor influences*. Action in any of the four modes of Chapters 4 and 5 (action described by any of the four models) may be either of these – a failure mode event or an actor influence that affects the likelihood of failure mode events.

Failure mode events are the basic events that, in particular combinations, cause system failure (like a pump failure that is a part of a system failure mode). They are the elements of minimum cut sets in a fault tree analysis. An example of an action that is a failure mode event would be an operator shutting off a safety system, making it unavailable in case of a crisis in the system – shutting off the safety system is an integral part of the accident sequence; without it, the system will not fail (at least, not by that failure mode). Actions that are failure mode events can be incorporated directly into a risk analysis model just like any other failure mode event. Management changes that affect the likelihood of such an action will change the corresponding failure mode event probability, which is straightforward to incorporate in the risk model; it is just like changing the failure probability of a physical system component.

The second category of actions, actor influences, are not failure mode events and are not a direct link in system failure, but can influence the likelihoods of failure mode events, like an external event does. A classic example of an external event is an earthquake, which can simultaneously put extra loads on many system components, increasing their failure probabilities, but does not necessarily cause any of them to fail, and thus is not an element of a failure mode. An external event does not show up directly in the risk model; rather, the failure probabilities in the model are conditioned on it. Actor influences will be treated in the same way, as conditioning variables that affect the probabilities of the system's failure mode events. The handling of these two types of actions may be best explained with the use of simplified examples.

#### Action that is a Failure Mode Event

The effects of an action that is a failure mode event are handled just like any physical

system failure mode event. Consider the simple schematic example of Figure 6.1, a subsystem in which there are two parallel paths that can prevent system failure – the first is primary automatic equipment, and the second is a system operator who can engage a manual backup to perform the same function if the primary system fails.

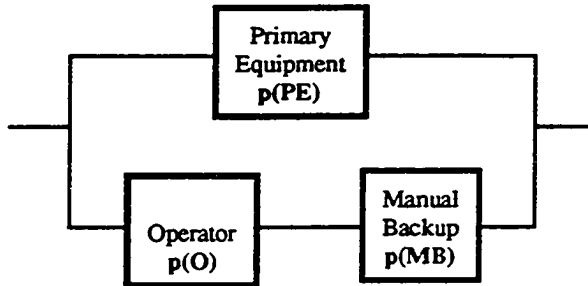


Figure 6.1: Subsystem schematic in which action (operator performance) is a failure mode event.

In this example, the action that affects system safety is whether the operator will successfully engage the backup system or make an error. If the probability of operator error,  $p(O)$ , is assumed to be independent of the other probabilities in the problem, then the subsystem failure probability is

$$p(F) = p(PE)[p(O) + p(MB) - p(O)p(MB)]$$

The action in this case, operator error, is treated exactly as a physical component failure would be. Management changes that affect the probability of operator error would simply change the value of  $p(O)$  in this equation, similar to the effect of a technical change that affects the failure probability of a piece of equipment. Actions that have multiple outcomes or those that affect the system at several points are all handled in the same way as a physical component failure with similar effects.

#### Action that is an Actor Influence

To illustrate how actor influences affect the likelihoods of a system's failure mode events, consider the alternate schematic of Figure 6.2, in which the primary equipment is in parallel with an automatic switch that will turn on the backup equipment in case of primary equipment failure. In this case, the actor influence is the quality of equipment maintenance; it does not show up directly, but affects the reliability both of the primary equipment and of the backup.

The equation for the failure probability of this subsystem has the same form as in the previous example, but here, both  $p(PE)$  and  $p(BE)$  depend on the quality of maintenance,  $M_i$ , which can be good ( $M_g$ ) or poor ( $M_p$ ). Thus, the probability of subsystem failure is

the weighted sum of the conditional probabilities of failure:

$$p(F) = \sum_{i=g,p} p(M_i) \{p(PE|M_i) [p(AS) + p(BE|M_i) - p(AS) p(BE|M_i)]\}$$

Note that the terms in brackets correspond exactly to the equation from the example above, except that they are conditioned on the quality of maintenance. In this case, management changes that affect the probability of good vs. poor maintenance change the values of  $p(GM)$  and  $p(PM)$  in this equation, which is analogous to having some influence over the external event probabilities. Of course, an actual system may well contain both the action of an operator, which is a failure mode event, and the action of maintenance, which is an actor influence, and there may be several actions of each type that are relevant to system risk.

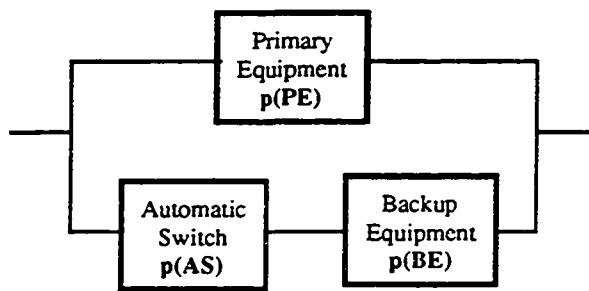


Figure 6.2: Subsystem schematic in which action (maintenance) is an actor influence which affects failure mode event probabilities.

#### Points at which Actions Affect the Physical System

The variety of different kinds of complex systems is immense, and the number of specific actions by individuals that may contribute to their failure is even larger. Action has the potential to influence the performance of the physical system, and thus the associated system risk, at any of the numerous points where an individual interacts with the system. However, there are some common ways in which humans interact with a system, and by which their actions can affect the physical system. Identifying and categorizing the ways in which actions affect systems will help in applying this framework to develop a risk analysis model that incorporates human and management effects. Individuals design, construct, operate, and maintain the physical components of a systems, and troubleshoot to detect, diagnose and correct problems when they occur. An error at any of these points may have the potential to cause or contribute to system failure.

#### - Operation

Normal system operation is usually the focus of attention in risk analysis, in part because

operation is the most prominent phase of system life, but also because problems in operation are often responsible for system failures. In normal operation, many complex systems deal with materials or processes that are inherently hazardous (e.g., nuclear power plants, chemical plants, transportation systems), and careful operation and control of these systems are necessary to prevent major disasters. An error in system operation can cause system configurations that are unstable and dangerous. Many airline disasters, including the Tenerife disaster that claimed 583 lives, are caused by errors in normal operations. The Chernobyl meltdown was also caused by errors in operation, though these occurred during testing rather than in normal operating mode. On the other hand, errors in normal system operation are not the only human actions that can affect a system's risk, and other actions should not be neglected.

#### - Maintenance

The purpose of inspection and maintenance is to service, repair, and replace system components to prevent or correct failures due to normal wear, so that the system continues to function safely. A distinction can be drawn between maintenance strategies: maintenance on schedule involves regular inspection and preventive maintenance; maintenance on demand corrects problems only after they appear. Obviously, the choice of strategy has implications for risk and the tradeoff between productivity and safety (Baron, 1994). A maintenance error may allow a crucial system component to fail, or can even cause a failure that would not have occurred otherwise. Maintenance can also interfere with normal system function, as when backup or safety systems are taken out of commission for maintenance with the system still in operation. A maintenance error that left a pump out of commission without warning was responsible for the explosion and fire that destroyed the Piper-Alpha offshore oil platform, and a similar error with valves in the emergency feedwater system may have contributed to the Three Mile Island meltdown.

#### - Design and Construction

Complex systems are created by humans who design and construct them. Errors in design and construction may make the system weaker than it was intended to be, or may cause system components to be less reliable or to fail under particular conditions. Although the examples used to illustrate the development of this framework generally deal with the operation of an existing system, the same techniques can be applied with equal effectiveness to the design and construction processes used to create a system. The O-ring problem that led to the explosion of the space shuttle Challenger is an example of a design error that was the direct cause of system failure.

### - Troubleshooting

Troubleshooting is the process by which problems (often during normal operation, but also in construction, maintenance, etc.) are detected, diagnosed, and corrected before they can cause significant damage. There are a number of ways in which the troubleshooting process can break down, where an error can prevent the selection and execution of the appropriate corrective action, and potentially lead to a system failure. If a problem is not detected at all, or if a detected problem cannot be diagnosed, then corrective action may not be taken. An inappropriate action may be taken if the problem is diagnosed incorrectly, or if the wrong response is selected following a correct diagnosis. Even if the appropriate action is selected, the actor may make an error in executing it. The Three Mile Island nuclear meltdown was caused in part by the fact that the operators did not properly diagnose and correct an initially minor problem.

## **6.2 Implementing the Framework**

This section pulls together the pieces of the framework developed in this dissertation into a brief checklist of the steps to be followed in applying the framework to an actual system. This checklist can be used to guide the development of a risk model that includes the effects of human and management factors for a particular system, and can support the development and evaluation of management strategies whose purpose is to reduce risk.

Briefly, to implement this framework for a particular system, the steps that must be followed are:

- Step 1: Construct a System-Level Risk Model
- Step 2: Link Effects of Actions to the Physical System
- Step 3: Select and Develop Models of Action;
- Step 4: Define and Quantify the Base Case
- Step 5: Define and Evaluate Management Changes

### **Step 1: Construct a System-Level Risk Model**

Begin by constructing a system-level risk model; like any risk analysis model, this should consist of the events at the level of the physical system (failure mode events) that in various combinations can cause failure of the system (these combinations of events are the system's failure modes). If the system fails in ways that are highly time-dependent, necessitating the use of a dynamic risk model to accurately portray the system, the classic definitions of failure modes and failure mode events may not strictly apply. But it still will be possible to construct a risk model containing the events that are directly involved



in the possible accident sequences. If a formal risk analysis model already exists for the system, it may provide a good starting point for the application of this framework<sup>37</sup>. Some of the failure mode events that appear in this model may be actions by individuals in the system, such as operator error. These action-events should be identified for later reference. Quantify the risk model to the extent possible, particularly the failure probabilities for physical system components that do not depend on actions. Probabilities of events that are actions of individuals need not be quantified here; they will be addressed in later steps.

### Step 2: Link Effects of Actions to the Physical System

Identify actions that are *actor influences* (discussed in Section 6.1) – actions that may influence the likelihoods of the events in the system-level risk model. Consider actions both before and during a possible accident sequence, actions associated with operating, maintaining, designing and constructing, and troubleshooting the system. The actions considered here should not be limited to the responsibilities that are assigned to actors by the organization. Often, actor influences are actions the organization discourages or does not address, such as cutting corners on critical tasks. Consider the actions that individuals must perform to ensure proper system functioning, and the possibility that they may fail to perform these actions. Also consider other actions that individuals might take that could cause system failure, ways that they might interfere with the system's proper functioning.

Organize actor influences into sets that are mutually exclusive and collectively exhaustive (that is, exactly one action from each set must occur). For example, if maintenance of a particular component affects its failure probability, this maintenance should be organized as a set of possible actions, such as 1) correct maintenance, 2) inadequate maintenance, 3) no maintenance. Because of how these actions are structured, one and only one of them must occur. Different sets of actions need not have any particular relationship to one another; e.g., there is no particular connection between the probabilities of the maintenance action and, say, an actor's choice of the production level at which to run the system, though their effects on risk may interact significantly. The interaction of their effects will be captured in the risk model, and need not be considered here.

---

<sup>37</sup> It may be easiest to approach this modeling task iteratively, first defining a greatly simplified model and progressing all the way through the framework with it, and then going back and adding detail and complexity to it in stages. For the first iteration of the framework, the analysis may be more tractable if relatively little detail is included in the risk model of the system. In any case, the risk model will probably change in the course of the analysis; later steps in this framework may bring to light additional failure modes or events that require revision of the initial risk model.

Condition the risk analysis model on the occurrence of these actor influences. This is identical to the process of conditioning a risk analysis model on the occurrence of an external event. The probabilities of events in the risk analysis model that depend on these actions must be conditioned on their occurrence. With the maintenance example above, where the component failure probability depends on maintenance, the probability of component failure is conditioned on each of the possible maintenance outcomes:

$p(\text{component failure} \mid \text{correct maintenance}),$   
 $p(\text{component failure} \mid \text{inadequate maintenance}),$   
 $p(\text{component failure} \mid \text{no maintenance}).$

### Step 3: Select and Develop Models of Action

For each actor influence identified in Step 2, and each failure mode event action identified in Step 1, select the appropriate model of action to characterize it. Refer to the discussion of section 5.6 to help select from the four models presented in Chapter 5.

Consider the intention and execution processes that lead to action: which of the models most accurately reflects the process that determines the actions of interest? In general, there may be several different relevant actions, and a separate model must be constructed for each. The models that are appropriate for these different actions may be of different types – expected utility may be the best model for some actions, while a rule-based or execution model is better for others in the same system.

Construct a model of the type selected for each relevant action, relating management and environmental factors to the actor's behavior. Refer to the section of Chapter 5 corresponding to the appropriate type of model for help in model construction.

### Step 4: Define and Quantify the Base Case

At this point, all the necessary pieces are in place to support a comprehensive risk analysis that includes the effects of human and management factors on system risk. The models of action capture the effects of management controls on the actions of individuals, through probability distributions on action. The effects of these actions on the physical system are included in the system-level risk model either directly (if the actions are failure mode events) or indirectly (if they are actor influences). The probability of system failure is calculated by this system-level risk model.

Define a Base Case to represent the system as it currently exists<sup>38</sup>. The Base Case should include the current state of the physical system (characterized by the system-level risk

---

<sup>38</sup> If the analysis is being conducted for a system that does not yet exist or is not yet in operation, then the choice of a "Base Case" is somewhat arbitrary. Since the point of the analysis is to compare the relative

model), as well as the settings of management variables, which serve as inputs to the models of action. Use the integrated model to determine the Base Case probability of system failure.

#### Step 5: Define and Evaluate Management Changes

The integrated model that has been developed with this framework can also be used to determine how changes in management factors affect individual actions, and how these changes in action will affect the overall failure probability.

First, the management change that is to be evaluated must be defined. Ideas for management changes may come from many sources, including similar systems elsewhere, management strategies used in different systems or industries, etc. Use insights from the model to suggest management changes – look for actions that are major contributors to risk, and management factors that have a significant influence on action. Then quantify the effects of the management change in terms of the inputs to the models of action<sup>39</sup>. The effects of the management change on actions are determined by running the models of action with these new inputs. The new probability distributions on actions are used with the system-level risk model to calculate the overall system failure probability under the proposed management change. The difference between this probability of system failure and that under the Base Case is the risk-reduction benefit of the proposed management change.

Repeat this step to evaluate other management changes. The joint effects of several management changes implemented at the same time can be evaluated similarly: define the management changes, quantify their joint effects on the models (which may differ from the combination of their individual effects), run the integrated model to determine overall system risk, and compare the result.

While the focus of this methodology is to assess the impact of management factors on risk, it is important to remember that risk is not the only dimension that is relevant in risk management decisions. Cost, productivity, etc., are important, and may also be affected

---

risks associated with different management strategies, it does not really matter which is identified as the Base Case; it only establishes a convenient reference point.

<sup>39</sup> In some cases, a management change may alter not only the inputs of the models, but also their structure or the choice of models of action. For example, if a proposed management change calls for the implementation of incentives in a situation where actors previously made decisions in rule-based mode, this might switch the actors into an expected utility maximization mode, necessitating a corresponding change in the model that is used to describe action. In such a case, the model itself must be altered in order to evaluate the management change.

by risk management strategies. If the risk management budget is fixed, then resources used by one risk management strategy are not available for others, and resource expenditures must be considered in order to maximize the total risk reduction. Beyond this, risk reduction must be traded off against other goals, such as cost reduction, productivity, environmental effects, etc., to make the best management decision.

Used as outlined here, the framework supports the development of an integrated risk model that captures the effects of human and management effects on risk. This integrated model can be used to evaluate the risk effects of a wide variety of proposed management strategies. This information about risk effects, together with information on the other costs and benefits of the various strategies considered, will allow management to prioritize its risk management measures, and optimize the allocation of resources.

Because of the complexity of many engineered systems, and the analytical resources that are required to characterize human and organizational factors in a model of this sort, it may be necessary to restrict the use of this approach to those situations in which human action has a particularly large potential effect on risk. However, this is not any different from the development of any risk analysis model, or any model, for that matter. Any modeling or analysis effort focuses on the aspects of the system that are the most important to the phenomena being studied, and the exercise of judgment is required to determine which aspects of the system will be included, and what is to be left out. In any particular application, the choice of the appropriate level of detail must be left to the analyst, who must trade off model complexity and accuracy on one hand and transparency, simplicity of use and availability of data on the other. The tradeoffs that are appropriate will depend on the system being modeled and the questions to be answered. A model that has needless complexity and detail will quickly become unwieldy, may require unavailable data, and may lack credibility because it is difficult to understand. On the other hand, a model that is too simple may fail to capture important effects and may give misleading results.

### **6.3 Hazardous Materials Transport Example – Synthesis**

It is difficult to give general instructions about how to integrate models of action into a system risk model, because the answer depends so much on the structure of the system and on the ways in which actions affect it. So to demonstrate the synthesis of action models into an overall system risk model that links human and management factors to the

risk of system failure, I will use an example, integrating the four models of action in the hazardous materials transportation example of the previous chapter into an overall risk model for the system. I will then demonstrate how this model can be used to evaluate risk management strategies.

This is a simple example developed for the purpose of illustration. The four actions that are modeled here, while they are important factors that affect accident probability, do not necessarily make up an exhaustive list of all the actions that might have an important effect on risk in a system such as this. But this will serve well to illustrate the synthesis of action models into a system risk model, because additional actions would be integrated in just the same way as these.

#### Developing the Global Risk Model

In developing a global model for the hazardous materials transport system, as for any system, the distinction between actions that are failure mode events and those that are actor influences is crucial. In this system, the (unintentional) action Accident, modeled by the execution model, is a failure mode event, an action which is a direct part of system failure. In fact, this particular action is system failure, which simplifies system modeling; the execution model itself will serve as the basis for the system risk model. The other three actions modeled are actor influences; none of them are necessary or sufficient for system failure, but they may have a significant effect on its likelihood. Their effect is captured through the execution model, as discussed below.

As shown in Section 5.4, the execution model determines overall accident probability by integrating the outcome function, which specifies the probability of an accident as a function of task demand, and the probability distribution on task demand. The example in that section showed how different driver types could have different outcome functions. Similarly, the effects of the other actions that influence risk will be to alter the outcome function, and in some cases, to alter the probability distribution on task demand as well. For example, driving speed affects the outcome function (the accident probability as a function of driving difficulty), and weather conditions alter the probability distribution on task demand (bad weather makes difficult driving more likely). Brake condition, speed, weather, and driver type all affect the parameters that define the outcome function and the probability distribution on task demand in the execution model.

The probability distribution on driver type is specified directly, and the distribution on speed is the output of the expected utility model. The results of the bounded rationality

and rule-based models, which describe the driver's decision about whether to delay in bad weather and the actions of the brake maintenance technician, respectively, are not used directly, but require further calculation to yield the probability distributions on weather and brake condition.

The first of these is quite simple – the model requires the fraction of trips made in each weather condition, Good Weather, Bad Weather on an Alternate Route, and Bad Weather on the Default Route. To get this from the probabilities of the driver's decision from the bounded rationality model, which assumed bad weather, the probability of Bad Weather is required – for this example, the probability of a storm is assumed to be 5%. The probability that the trip occurs in good weather, then, is just the probability of good weather when the trip is originally scheduled plus the probability that there is a storm and the driver chooses to delay:

$$p(\text{Good Weather}) = 1 - p(\text{storm}) + p(\text{Wait} \mid \text{storm}) p(\text{storm}).$$

The probability that the trip occurs on the alternate route or on the default route in a storm are thus given by:

$$\begin{aligned} p(\text{Alt. Route; Storm}) &= p(\text{Alt. Route} \mid \text{storm}) p(\text{storm}) \\ p(\text{Default Route; Storm}) &= p(\text{Go} \mid \text{storm}) p(\text{storm}). \end{aligned}$$

The final part of the model relates the actions of the brake maintenance technician, modeled by the rule-based model, to the probability distribution on brake condition (Good, Fair, or Worn) using a stochastic model of the deterioration and servicing of brakes<sup>40</sup>. The rule-based model generates a Repair matrix,  $R$ , which specifies the probability that a truck's brakes will be in each possible condition after service, given their condition before service. Table 6.1 repeats Table 5.2, giving the Base Case Repair matrix for the transport example; entry  $i,j$  in the matrix is the probability that brake condition is  $j$  after service, given that it was  $i$  before service.

	Good	Fair	Worn
Good	1.00	0	0
Fair	0.54	0.46	0
Worn	0.91	0	0.09

Table 6.1: Brake repair matrix for transport example: Base Case.

The evolution of brake condition is controlled by the rate of brake deterioration and the frequency and quality of service. To find the average time spent in each condition, first a

<sup>40</sup> As with the dynamic model used in the anesthesia project, in order to model the stochastic, time-dependent nature of brake state, it is necessary to go beyond conventional risk analysis tools.

stochastic model is developed to find the steady-state distribution of brake condition immediately after service. This is done by creating a Wear matrix,  $\mathbf{W}$ , shown in Table 6.2, that describes the degradation of brake condition in use, based on a constant wear rate.

	Good	Fair	Worn
Good	0.61	0.30	0.09
Fair	0	0.61	0.39
Worn	0	0	1.00

Table 6.2: Brake wear matrix for transport example: Base Case.

The matrix product of the Wear and the Repair matrices,  $\mathbf{WR}$ , characterizes transitions among brake conditions during a single cycle of wear and service. The equilibrium distribution of brake condition immediately after service,  $[P_G, P_F, P_W]$ , is given by the steady state distribution for this stochastic system:

$$[P_G, P_F, P_W] = [q_G, q_F, q_W] \lim_{k \rightarrow \infty} (\mathbf{WR})^k.$$

This has a unique solution that is independent of the arbitrarily chosen initial distribution,  $[q_G, q_F, q_W]$ .

This distribution on brake condition immediately after service is not the same as the average time spent in each state (in fact, this is the best brake condition distribution; brakes only get worse as they wear). Since the constant wear rate of the Wear matrix implies exponential transition times, a Poisson process is used to describe the deterioration process. This allows the expected time in each state to be calculated, using the steady state vector  $[P_G, P_F, P_W]$  as the starting point.

Any given combination of brake condition, speed, weather condition, and driver type is called a scenario, and all combinations are possible. An example of one possible scenario is an Experienced, Fatigued driver traveling at 10 mph Over the speed limit with Fair brakes, in Good weather. The overall system risk is the sum of the products of the scenario probabilities and conditional accident probabilities:

$$p(\text{Accident}) = \sum_{\substack{i = \\ \text{driver type}}} \sum_{\substack{j = \\ \text{brake}}} \sum_{\substack{k = \\ \text{speed}}} \sum_{\substack{l = \\ \text{weather}}} p(i, j, k, l) p(A | i, j, k, l).$$

The term  $p(i, j, k, l)$  is the probability of the scenario in which driver type is  $i$ , brake condition is  $j$ , driving speed is  $k$ , and weather condition is  $l$ ;  $p(A | i, j, k, l)$  is the probability of an accident conditional on the occurrence of scenario  $i, j, k, l$ .

In this example, probabilities of scenario elements (driver type, brake condition, etc.) are assumed to be conditionally independent, given a particular management strategy, so scenario probability,  $p(i, j, k, l)$ , is the product of the probabilities of the elements of that scenario:

$$p(i, j, k, l) = p(A_i) p(B_j) p(S_k) p(W_l)$$

where  $p(A_i)$  is the probability of driver type  $i$ ,  $p(B_j)$  is the probability of brake state  $j$ ,  $p(S_k)$  is the probability that the driver travels at speed  $k$ , and  $p(W_l)$  is the probability of driving in weather condition  $l$ . The probabilities of each of these scenario elements come from the action probabilities calculated by the expected utility, bounded rationality, and rule-based models of action, which all take into account management effects. The fact that these probabilities are all conditioned on management strategy means that this model does capture the common cause effects of management on risk. This answers a persistent criticism of probabilistic risk analysis – that it underestimates the likelihood that management problems will simultaneously increase the failure probabilities of multiple system components. Conditioning the entire risk model on management captures these dependencies. For example, if selection mechanisms choose drivers that are both more experienced and more cautious, the probability of inexperienced driver and the probability of exceeding the speed limit decrease together, as can be seen in the example below. If there were additional probabilistic dependence due to factors other than management, it would be taken into account as it would be in any probabilistic analysis, by conditioning the probabilities of the scenario elements on one another as necessary.

The failure probability conditional on a given scenario,  $p(A | i, j, k, l)$ , is calculated using the execution model with inputs appropriate for that scenario. These inputs define the outcome function and distribution on task demand for that scenario by adjusting the nominal parameter values for the outcome function and task demand distribution. However, because the execution model is nonlinear, multiplicative changes to the input parameters do not lead to proportional changes in the overall risk. The structure of the relationships among management factors, human decision and action, and the physical system, and the effects of all these on system risk in this example, are illustrated in the influence diagram of Figure 6.3.

The task demand distribution is a triangular distribution that can be characterized by its maximum value,  $m$ , and is given by

$$f(x) = \left(\frac{2}{m^2}\right)x + \frac{2}{m}; \quad x = 0 \text{ to } m.$$



The outcome function, which gives the probability of an accident as a function of task demand, is exponential, and is characterized by two parameters,  $k_1$  and  $k_2$ :

$$a_a(x) = k_1 e^{k_2 x}$$

The values of the parameters  $m$ ,  $k_1$ , and  $k_2$  for any given scenario are equal to the nominal parameter values times the multiplicative factors that correspond to the elements of that scenario, as given in Table 6.3. For a given scenario, each parameter is scaled by the product of the factors corresponding to that scenario. The probabilities of these individual scenario elements are also given in Table 6.3. These Base Case probabilities of action match those from the illustrations of action models in Chapter 5.

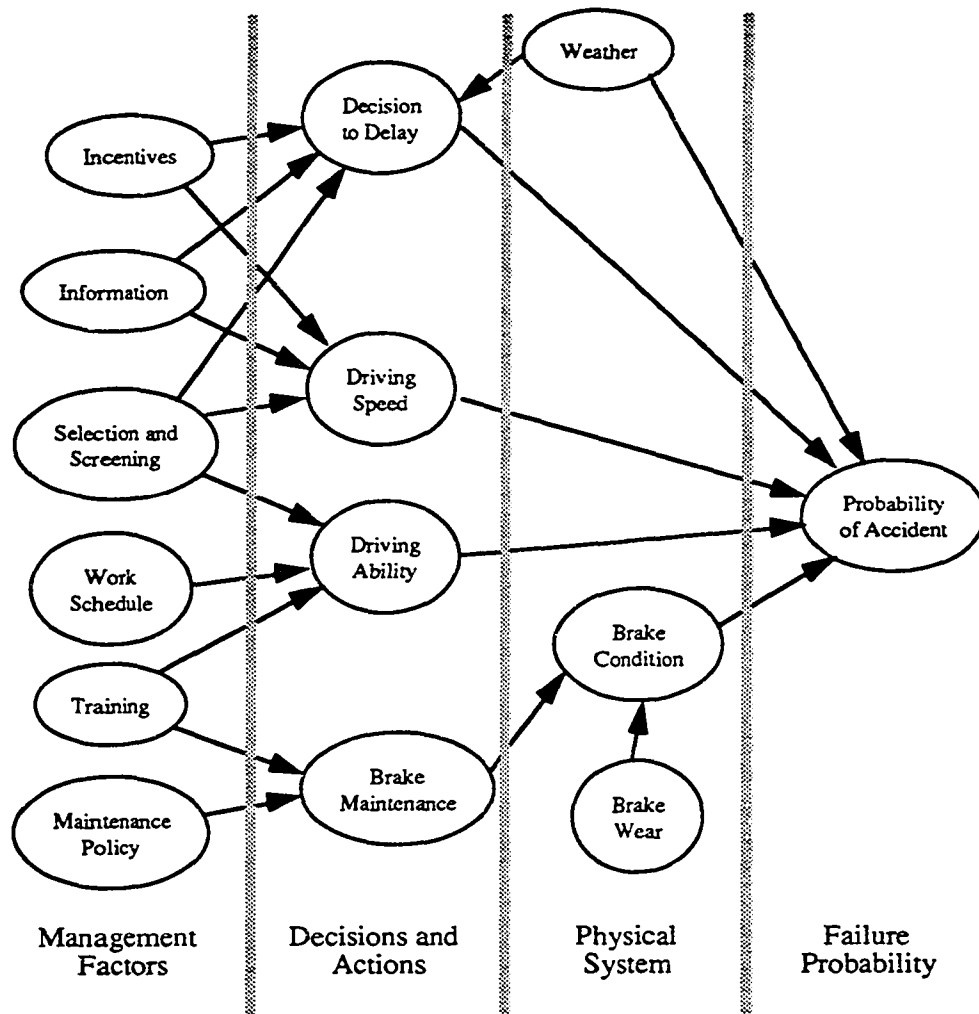


Figure 6.3: Factors affecting accident probability in transport example.

<u>Nominal Value</u>		<u>m</u>	<u>k<sub>1</sub></u>	<u>k<sub>2</sub></u>	<u>Probability</u>
		1	.0001	1	
<u>Driver Type</u>	Exp., No Fat.	-*	-	-	.4
	Exp., Fatigue	-	2x	-	.1
	Inexp., No Fat.	-	-	2x	.4
	Inexp., Fatig.	-	2x	2x	.1
<u>Brake j</u>	Good	-	-	-	.650
	Fair	-	1.1x	-	.275
	Worn	-	1.5x	-	.075
<u>Speed k</u>	Limit	-	-	-	.152
	10 Over	-	1.5x	-	.323
	20 Over	-	2.5x	-	.525
<u>Weather l</u>	Good	-	-	-	.967
	Bad - Alt. Rte	3x	-	-	.007
	Bad - Default	4x	-	-	.026

\* parameter unchanged; multiplicative factor = 1

Table 6.3: Base Case model inputs for transport example.

As described in Section 5.4, the conditional failure probability for a given scenario is the integral of the task demand distribution and the outcome function for that scenario:

$$p(A | i, j, k, l) = \int_{-\infty}^{\infty} f_{ijkl}(x) a_{ijkl}(x) dx$$

The overall failure probability for the system is the probability-weighted average of the conditional probabilities, as given above.

Calculating the overall risk using this model yields an accident probability of  $7.80 \times 10^{-4}$ . Performing a simple sensitivity analysis on this result (by decreasing the probabilities of higher risk cases by half, one at a time) reveals that driver type, driving speed, and weather conditions are all major contributors to system risk. These sensitivity results indicate that risk would decrease by 29%, 24%, and 21% respectively, if the probabilities of higher risk cases of driver type, driving speed, and weather conditions could be cut by half, so management strategies that are able to have a significant effect on these probabilities might be effective risk reduction measures. Brake condition turns out to be a much smaller contributor; the sensitivity analysis result for it showed only a 3% decrease in risk, so even a very effective strategy for improving brake maintenance would not cause a significant risk reduction.

A disproportionately large fraction of risk in this system is caused by scenarios in which there are multiple problems. Although accidents are possible in all of the scenarios, those scenarios with problems on all four dimensions account for fully 15% of the system's risk, even though together they account for just over 0.5% of the total scenario probability. This is consistent with the observations of a number of researchers (Reason, 1990a; Rochlin, et al., 1987; Perrow, 1984) who observe that in complex systems, failures seldom result from a single problem, but are usually caused by the combination of multiple problems that unfortunately coincide to allow catastrophic system failure.

To illustrate the reason for this, the worst-case scenario in this example (a Fatigued, Inexperienced driver, driving at 20 mph Over the limit with Worn brakes in Bad weather) is more than 480 times as risky as the best-case scenario (an Experienced, Not Fatigued driver, driving at the Speed Limit with Good brakes in Good weather) and almost 90 times as risky as the average over all scenarios. (Fortunately, there is only a 1 in 10,000 chance that this worst-case scenario will occur.) While this is clearly an unacceptable level of risk, it is usually impossible to single out and address such high risk scenarios directly. Since it is generally even more difficult to identify the joint occurrence of multiple problems than it is individual problems, the best way to address the risk caused by combinations of factors may be to simply reduce the likelihood of each of the individual problems, thereby also reducing the joint likelihood. Of course, many of these causes may have common roots in system management, so management solutions to these problems may not only decrease the likelihoods of individual problems, but also decrease the dependency between them to provide added risk reduction benefits.

#### Measuring the Effects of Risk Management Strategies

While developing and quantifying a Base Case model for a system such as this can be instructive, and can help to identify the primary contributors to risk, the real value of a model like this is that it can be used to evaluate the effectiveness of proposed management changes that are designed to reduce risk.

In this model, changes in management policy do not affect the accident probabilities associated with scenarios, but change the scenario probabilities. So once the Base Case has been developed, the conditional failure probability for each scenario,  $p(A | i, j, k, l)$ , is not affected by management changes, though the scenario probabilities,  $p(i, j, k, l)$ , will change. For example, a management policy that reduces drivers' workload will affect the probability that a driver is fatigued, and thus reduce the probabilities of scenarios that include a fatigued driver; however, this policy will not change the performance of a

fatigued driver, so the conditional accident probabilities associated with fatigue scenarios is unaffected. The effects of management changes are captured by quantifying management factors in terms of the inputs to the behavior models (e.g., changing the incentives in the expected utility model) to determine their effects on the likelihoods of the actions modeled, and then propagating these effects through the system risk model to determine the resulting effect on risk.

To illustrate this process with the transport example, I look at four proposed risk management strategies, each consisting of several parts. These strategies are designed to address a particular aspect or weakness in the system; they are:

- Strategy 1 – Improve Maintenance: Brake Condition
- Strategy 2 – Select for Better Drivers
- Strategy 3 – Improve Drivers' Decisions
- Strategy 4 – Improve Drivers' Abilities

and they are discussed below. The intermediate results (probability distributions on driver type, brake conditions, and decisions) and the risk reduction implications of these strategies are shown in Table 6.4 below.

		<u>Base</u>	<u>1 Maint</u>	<u>2 Select</u>	<u>3 Dec'n</u>	<u>4 Ability</u>
<u>Driver i</u>	1 E,f	.4	.4	.56	.4	.54
	2 E,F	.1	.1	.14	.1	.06
	3 I,f	.4	.4	.24	.4	.36
	4 I,F	.1	.1	.06	.1	.04
<u>Brake j</u>	1 Good	.650	.748	.650	.650	.650
	2 Fair	.275	.208	.275	.275	.275
	3 Worn	.075	.044	.075	.075	.075
<u>Speed k</u>	1 Limit	.152	.152	.262	.497	.152
	2 10 Over	.323	.323	.285	.187	.323
	3 20 Over	.525	.525	.453	.316	.525
<u>Weather l</u>	1 Wait/Good	.967	.967	.970	.980	.967
	2 Alt. Rte	.007	.007	.006	.004	.007
	3 Go/Bad	.026	.026	.024	.016	.026
<u>Risk</u> (x10 <sup>-4</sup> )		7.80	7.64 (-2%)	5.69 (-27%)	5.25 (-33%)	6.44 (-17%)

Table 6.4: Effects of risk management strategies – intermediate results and overall risk.

• Strategy 1: Improve Maintenance – Brake Condition

A new Brake Replacement strategy addresses the actions of the maintenance technicians who inspect and maintain the trucks' brakes. This proposed strategy includes a policy

that specifies that brakes should be replaced when in Fair condition, rather than waiting until they are completely Worn, and also includes inspection training for repair technicians, which improves their ability to correctly identify brake condition. The new brake replacement policy increases the probability that the technician's rule-base specifies that brakes should be replaced when their condition is identified as "Fair," from 0.6 in the Base Case to 0.9 under the Improved Maintenance policy (note that it is still possible that the technician will not follow the organization's policy). The inspection training for maintenance technicians improves their ability to diagnose brake condition, reducing the probability of misdiagnosing brake condition by half. These changes cause a significant improvement in the performance of the maintenance function; under the Improved Maintenance policy, just 15% of brakes that are Fair and 3% of those that are Worn are returned to service without replacement, compared with 46% and 9% in the Base Case. This causes a change in the probability distribution on brake condition (Good, Fair, and Worn, respectively), from [0.650, 0.275, 0.075] in the Base Case to [0.748, 0.208, 0.044] under Improved Maintenance.

Unfortunately, while this is a fairly significant improvement in brake condition, it does not have a very large effect on risk – under this proposed policy, overall risk of an accident is reduced by only 2%, from  $7.80 \times 10^{-4}$  to  $7.64 \times 10^{-4}$ . This is because Fair brake condition does not have a large effect on risk, and the probability of Worn brakes is already low in the Base Case, so cannot be reduced by much. On the other hand, even though the benefit is not large, this strategy may still be worthwhile if it is simple and inexpensive to implement.

- Strategy 2: Select for Better Drivers

By changing the selection processes by which drivers are acquired and retained by the firm, it is possible to improve drivers' performance, on average. This new strategy would consist of basing selection decisions for hiring drivers more heavily on the driver's level of experience and previous driving record. Its effect would be to increase the fraction of drivers who are Experienced (increasing drivers' ability to avoid accidents), to make drivers more risk averse on average (decreasing the number who choose to exceed the speed limit), and to increase the likelihood that the drivers decision to delay in bad weather is based on Safety rather than Schedule. As a result of these changes, the fraction of drivers who are Experienced will increase from 0.5 to 0.70, and the range of the risk aversion parameter, gamma, increases from .0002-.002 to .0003-.003 in the expected utility model. The probability that drivers will base decisions on Safety in the bounded rationality model increases slightly, from 0.35 to 0.40. Because the turnover

rate of drivers is limited, a strategy such as this may take some time to display its full effects. The effects analyzed here are the new equilibrium conditions after the change.

The effects on drivers decisions and on driver type are shown in Table 6.4, and the overall effect on the accident probability is to reduce it by 27%, from  $7.80 \times 10^{-4}$  to  $5.69 \times 10^{-4}$ . This is a significant risk reduction, about two-thirds of which is due to the effect on Driver type (increasing the probability that drivers are Experienced).

- Strategy 3: Improve Drivers' Decisions

Drivers' decisions about driving speed and whether to delay a trip because of bad weather play an important role in the probability of an accident, and there are several mechanisms that can be used to help ensure that these decisions are prudent. The strategy proposed here is to reduce speeding by increasing the disincentive associated with a speeding ticket, and to reduce schedule pressure on drivers, so that they are more likely to delay a trip when driving conditions are poor. An additional penalty of \$250 is imposed on drivers who receive a speeding ticket, as in the expected utility example in Section 5.1, and the reduction in production pressure will have the same effect as the bounded rationality example in Section 5.2 – to increase the probability that the driver will base his decision on Safety from 0.35 to 0.45, and to reduce the aspiration level on the Schedule dimension so that Waiting is an acceptable alternative.

The effect of these changes on drivers' decisions is significant (see Table 6.4); they reduce the risk of an accident by 33%, to  $5.25 \times 10^{-4}$ . The effect of the disincentive for speeding and that of the reduction in production pressure each account for about half of the total risk reduction.

- Strategy 4: Improve Drivers' Abilities

One of the major factors affecting accident probability is driver ability, as influenced by Experience and Fatigue. This final policy changes drivers' work schedules to reduce fatigue, and uses driver training to increase the number of drivers whose abilities are comparable to those of an Experienced driver (while training does not change drivers' actual experience, the effect is the same as replacing some inexperienced drivers with experienced ones). The changes in work schedule reduce the fraction of drivers who are fatigued from 20% to 10%, and the effect of driver training is to increase the number of Experienced drivers from 50% to 60%.

This strategy has a significant effect on the distribution of driver type (see Table 6.4), and reduces risk by 17%, to  $6.44 \times 10^{-4}$ . This is still quite significant, but a smaller effect

than the previous two management strategies examined. Just over half of this effect is due to the reduction in the number of fatigued drivers, and the remainder is due to the effects of training.

These examples serve to illustrate the ways in which management strategies can influence risk through their effects on human decisions and actions, and how the framework developed in this research can be used to evaluate the effects of such strategies.

Obviously, there is an almost unlimited number of different risk management strategies and combinations of them that could be proposed, but for any of them, evaluation proceeds by identifying the effects of the management strategies on the decision and behavior models that predict action, using those models of action to predict the effect on behavior, and then incorporating the behavior changes in the system risk model to determine the effects on system risk. The framework measures the effect on risk of proposed risk management strategies, information that is necessary for making the cost-benefit tradeoffs necessary for optimal allocation of risk management resources.

Of course, it must be made clear that risk is not the only dimension on which the quality of a management strategy should be judged. Other dimensions, such as productivity, cost, environmental impact, etc., are also important. They may also be affected by risk management activities, and any decision about management strategy must take them into account, balancing risk reduction with other relevant dimensions to choose the best overall management strategy. Similarly, management strategies that are designed to affect dimensions other than risk (such as cost-cutting programs, or efficiency enhancements) may have unintended effects on risk through their influence on the behavior of actors in the system. This methodology can be used to examine the risk implications of such strategies before making decisions about them.

#### **6.4 Summary of Chapter 6**

This chapter began by developing the link between the actions of individuals and the performance of the system, the final quantitative connection in the management-action-system chain modeled by this framework. It then provided a brief description that summarizes the steps one would go through in implementing the framework in a real application, beginning with a risk model of the physical system, identifying and modeling the effects of action on the system, and finally the effects of management on the relevant actions. The use of this framework is illustrated by tying together the illustrative

examples of the previous chapter into a system risk model for hazardous material transport. This risk model characterizes the risk implications of the actions modeled, and demonstrates the evaluation of several different risk management strategies that might affect these actions and thus the overall risk of the system.



## Chapter 7

### Conclusion and Future Research

#### **7.1 Capabilities of the Methodology**

This dissertation has developed a framework to implement a PRA-based quantitative risk analysis methodology that can explicitly include organizational and management effects. Since management does not affect the system directly, the framework models the actions of individuals in the system as an intermediate. The innovative feature of this approach is that it extends PRA techniques beyond the physical system to include explicit models of human actions and the organizational factors that affect those actions. It develops explicit models to predict human action in a given situation, and avoids the pitfalls of attempting to predict a phenomenon as inherently uncertain as human action by using probabilistic techniques. It can therefore utilize limited information about the factors that drive human behavior, without claiming false accuracy. In addition to offering a more accurate risk assessment tool, this framework's ability to evaluate the effects of organizational change makes it useful for risk management and risk reduction; it can identify how management and organizational factors can be used to make technological systems safer and more reliable. This approach can be used to address organizational and management effects at all stages of a system's life: in design, construction, maintenance, and decommissioning, as well as in system operation.

The four models of action that have been developed for this framework are not necessarily the only or the best models available. But one of the features of this methodology is that the behavior models are "modular" components of the framework, in that they can easily be substituted for one another as the situation demands. All of the models of action are used predict the probability of various possible actions as a function of management factors. This means that any other model of action that offers descriptive or predictive capabilities beyond the four developed in Chapter 5 can easily be used with this framework, as long as it can be quantified to calculate a probability distribution on possible actions. Examples of alternative models that might be adapted to this framework include Kahneman and Tversky's prospect theory (1969), Bell's theory of regret (1982), an alternative structure for the bounded rationality model, or an entirely new model that captures effects beyond those now considered (see section 7.2 below).

The final form of the methodology developed here has not yet been tested in an actual application. Earlier, less detailed versions of this basic approach have been applied to

risks associated with the space shuttle, offshore oil drilling platforms, and most recently, anesthesia. The final form of the framework has been applied to the illustrative example of hazardous material transport, but that is admittedly a simplified problem. It is likely that some practical problems will arise when the framework is applied to a real system, but none of these seem insurmountable. The next step is to apply the more detailed methodology developed here in other domains. It would probably be best to apply it first to a relatively simple system with a few, easily identifiable points at which action affects the physical system, before attempting to use it with a more complex system. The primary obstacle encountered in applying this methodology is likely to be identifying the appropriate level of modeling detail: in determining what is important, which effects can be left out, and how to capture the important effects while keeping the analysis tractable. In this respect, this methodology is no different from any other modeling effort. Experience with applications will be helpful, and perhaps further research could simplify the modeling of some effects.

## **7.2 Limitations and Future Research Directions**

There are a number of issues that this framework has not necessarily been designed to address. It might be possible to handle some of these with little or no change to the methodology; others cannot be addressed without significant extensions; still others probably cannot be managed with this approach at all. A collection of these issues is discussed briefly below; it serves as a list of potential future directions for extending this line of research.

### **External Events**

The effects of external events (e.g., an earthquake), while not discussed in the development of this framework, are handled adequately by the current PRA methodology. To include the effects of an external event along with organizational effects in this methodology, the analysis of action and management effects must be conditioned on the occurrence (and non-occurrence) of the external event, and it proceeds in the same way as an external event analysis in the current PRA methodology. An interesting note is that some external events, such as fires, may actually be caused by human and management effects, and the approach developed here can offer a powerful way to understand and influence such effects.

### Long-term Effects

The models in this framework do not explicitly address the question of long-term effects, as when an actor's decision or behavior occurs significantly before the associated consequences. It may be possible, in some cases, to address this question by simply discounting future outcomes to account for their timing, as is often done for cash flows. (If so, it should be recognized that the implicit discount rate in such decisions is often quite high.) In addition, there may be a sort of "not-on-my-watch" syndrome, where individuals are willing to take reasonable precautions against accidents that might occur in the short-term, during their tenure in the system, but are less concerned about longer-term system degradation that may cause failure after they are gone. This may be similar to the question of "inter-generational" values, which come into play when trading off current outcomes to oneself against future outcomes to others.

### Effects of Organizational Structure

While the purpose of this research is to incorporate the effects of organization and management on system performance, it has not explicitly examined the question of the structure of the organization. That is, an organization's formal and informal structure - the communication and authority relationships between individuals and groups within the organization - can affect individual and organizational performance. Several researchers (e.g., Carley and Prietula, 1994; Levitt, et al., 1993; Sah and Stiglitz, 1986) have developed simulation models of simplified organizations to examine the effects of organizational structure. To some extent, it may be possible to capture these effects within the framework developed here, though some of the issues raised by organizational structure must be resolved before the models in this framework can be applied. For instance, while the expected utility model can show how information affects the behavior of a rational actor, it may be necessary to look explicitly at the structure of the organization in order to determine what information will be available to the actor. Organizational structure can also have a significant effect on the allocation of tasks and responsibilities among individuals, such as whether individuals specialize or are generalists, and which individuals are responsible for what tasks. Structure may even affect what actions are relevant - coordination functions that are important to organizational functioning in one structure may be unnecessary in an alternative structure.

### Group Effects

In addition to the effects of organizational structure on communication and authority relationships, a significant amount of social psychological research indicates that behavior in groups may be quite different from individual, non-group behavior. For

example, Wallach, et al. (1962) identified the "risky shift," in which the decisions of a group tend to be more risk-taking than those of the group's individual members; Janis (1972) coined the term "groupthink" to describe how group pressures can interfere with intelligent decision-making. This and other psychological research suggests that group behavior may be qualitatively different from that of individuals, particularly where risk is concerned. A closely related issue is group culture, which can be a very powerful force that is not entirely under management's control. Despite this, it may be reasonable in some cases to treat a group in its entirety as a single, unitary actor, characterizing the group's actions with the models already contained in the framework, or to treat the group as a simple collection of individuals, considering each actor's behavior independently. For situations in which neither of these approaches are appropriate, further research would be necessary to develop models of action that characterize group behavior, and to determine when such models would be appropriate.

#### Learning Effects

Learning is another effect that is not explicitly considered in this research. While they have not specifically been designed to do so, the models included in this framework may be able to capture some learning effects with only minor modifications. For example, in rule-based action, learning could be modeled as changes over time in an actor's rule base. Following an accident, actors' rules may change to avoid the behavior identified as the cause of failure; such an effect may fade with time. In the expected utility and bounded rationality models, learning may improve actors' knowledge of the effects of their actions, changing the outcomes in the decision problem that the actor resolves and thus influencing the actor's decision. In the execution model, learning (e.g., through experience or practice) may cause an actor's ability to increase over time. Further research may help to capture learning effects more systematically with this methodology; some of the effects of organizational and individual learning have been examined by Lounamaa and March (1987), Herriott, et al. (1988) and Carley and Prietula (1994).

#### Irrational or Reckless Behavior

Behavior that is "irrational" or "reckless" was not mentioned in the development of this methodology, and seems at first that it would be difficult to model, because such behavior is not goal-directed. However, what may appear reckless to one observer is justifiable risk-taking to another; a seemingly irrational action makes perfect sense from a different perspective. This provides a clue for how to approach modeling such behavior: rather than characterizing it as not purposeful, it may be reasonable to think of it as goal-directed behavior with unusual, highly uncertain, or even perverse values and goals. If

so, it can be modeled within the framework developed here, and it might be possible to influence it with many of the same management mechanisms – selection, indoctrination, incentives, organizational culture, etc. While there may be greater uncertainty involved, the best recourse is to characterize such actions as well as possible, and make risk management decisions accordingly.

On the other hand, there are a number of psychological theories that describe behavior as being driven by emotions, social pressures to conformity, repressed sexual urges, etc., rather than being "rationalistic" or goal-directed. While it might be possible to capture some of these effects in terms of the models in this framework (e.g., social pressure as a dimension of outcome in an expected utility or bounded rationality model), generally, these phenomena are too poorly understood to be useful for prediction, and are not considered by this framework. Of course, if one of these models could be quantified to characterize behavior according to the requirements of the framework, it should be possible to include it along with the existing models.

#### Pathological Behavior

Another type of behavior not considered here is pathological behavior – actions such as sabotage or terrorist attack by individuals within or outside the system. Fortunately, such actions seem to be relatively rare, and account for a small share of failures in complex systems. While there are some management approaches that can reduce the likelihood of such actions (e.g., good employee relations may prevent sabotage by a disgruntled employee; security can reduce the chance of outside sabotage), this framework is not particularly well-suited to addressing these types of questions.

### 7.3 Conclusion

The methodology developed here clearly has some limitations. Data to support the modeling of specific actions may be difficult to obtain. While limited historical or statistical data can be supplemented with expert judgment (often a very useful source of information), the use of experts does raise some additional problems, such as biases in judgments, resolving disagreement between experts, etc. Beyond the problems of acquiring data, the models used here are, at best, crude abstractions of the complexities of human behavior, and in order to keep the models tractable, it will often be necessary to greatly simplify the system model and limit the effects that are examined in detail. As a result, the output of a risk model developed with this framework should be considered a

rough quantification of management effects, appropriate for identifying the direction and approximate magnitude of effects, but not for making precise measurements. And as discussed in the previous section, there are a number of effects that this framework does not handle well, or at all.

However, while modeling management effects is an inexact endeavor, any risk analysis methodology that does not consider such effects can offer only illusory precision, and may in fact be quite inaccurate. Despite the limitations of this methodology, there is often wide agreement, particularly in the wake of a major disaster, that technological systems are not managed as well as they should be. The research presented here is an attempt to do as well as possible (or at least, to do better) with the admittedly limited information, knowledge, and resources that are available. In spite of the complexity of the models developed in this framework, in some ways it is a simplistic solution to a very difficult problem. But at this point, our understanding of human behavior is so limited that a simplistic solution may be the best that is available, and it is certainly a promising step in the right direction.

A final issue to be mentioned is the potential for other applications of the methods developed here. While this research focuses on factors that affect risk, the techniques used here to model management influence on action may be generalizable to behavior that affects other dimensions of organizational performance as well. That is, it may be possible to develop a similar framework to model management effects on productivity, cost, or any of a number of other measures of performance. In fact, the only part of this framework that is unique to the analysis of risk is the final piece - the risk model of the physical system that characterizes the components and relationships of the physical system to calculate overall failure probability. The largest part of the framework - the models of how management factors affect action - are just as relevant to actions that affect system outcomes other than risk. So it should be possible to use these models to study how management factors influence actions that affect, for example, productivity or the reliability of manufactured products. The probabilistic techniques that are used here to generate and utilize probabilistic predictions of action would also be important for a framework that characterizes other system outcomes, because human behavior is so unpredictable that deterministic predictions are not useful. This framework could turn out to be the basis of a quantitative management analysis methodology that describes the influence of management on many different types of action. This is an exceptionally ambitious aim, but such a quantitative management methodology would have application far beyond the questions addressed here.

## References

- Apostolakis, G.; Bickel, J.; and Kaplan, S. (1989) Editorial on Probabilistic Risk Assessment in the Nuclear Power Utility Industry, *Reliability Engineering and System Safety*, **24**, 91-94.
- Argyris, C. (1964) *Integrating the Individual and the Organization*, Wiley, New York.
- Arrow, K. (1971) "Insurance, Risk, and Resource Allocation," *Essays in the Theory of Risk Bearing*, Markham, Chicago.
- Asch, S. (1952) *Social Psychology*, Prentice-Hall, Englewood Cliffs, NJ.
- Atomic Energy Commission (1957) *Theoretical possibilities and consequences of major accidents in large nuclear power plants*, USAEC report WASH-740, Washington, D.C.
- Bannister, J.E. (1988) "The human factor in risk analysis," *Chartered Mechanical Engineer*, February, 24-6.
- Baron, M.M. (1994) "Managing the tradeoff between productivity and safety in critical engineering systems," Working paper, Stanford University.
- Bea, R. and Moore, W. (1991) "Management of Human and Organizational Error in Operational Reliability of Marine Structures," Second SNAME Offshore Symposium, April.
- Bell, D.E. (1982) "Regret in Decision Making Under Uncertainty," *Operations Research*, **30**, 961-81.
- Berkun, M.M. (1964) "Performance decrement under psychological stress," *Human Factors*, **6**, 21-30.
- Bley, D.; Kaplan, S.; and Johnson, D. (1992) "The strengths and limitations of PSA: where we stand," *Reliability Engineering and System Safety*, **38**, 3-26.
- Bueno de Mesquita, B. (1981) *The War Trap*, Yale University Press, New Haven.
- Bueno de Mesquita, B. (1980) "An Expected Utility Theory of International Conflict," *American Political Science Review*, **74**, 917-31.
- Carley, K.M. and Prietula, M.J. (1994) "ACTS theory: extending the model of bounded rationality," in K.M. Carley and M.J. Prietula (eds.), *Computational Organization Theory*, Lawrence Erlbaum Associates, Hillsdale, NJ.
- Clancey, W.J. (1985) "Heuristic classification," *Artificial Intelligence*, **27**, 289-350.
- Cohen, M.D. (1991) "Individual learning and organizational routine: emerging connections," *Organizational Science*, **2**, 135-9.
- Coombs, C. (1975) "Portfolio theory and the measurement of risk," in M. Kaplan and S. Schwartz, (eds.) *Human Judgment and Decision Processes*, Academic Press, New York.
- Cooper, J.B.; Newbower, R.S.; and Kitz, R.J. (1984) "An Analysis of Major Errors and Equipment Failures in Anesthesia Management: Considerations for Prevention and Detection," *Anesthesiology*, **60**, 34-42.
- Cooper, J.B.; Newbower, R.S.; Long, C.D.; and McPeck, B. (1978) "Preventable Anesthesia Mishaps: a study of human factors," *Anesthesiology*, **49**, 399-406.

- Cyert, R.M. and March, J. (1963) *A Behavioral Theory of the Firm*, Prentice-Hall, Englewood Cliffs, NJ.
- Davoudian, K.; Wu, J.; and Apostolakis, G. (in press, a) "Incorporating organizational factors into risk assessment through the analysis of work processes," *Reliability Engineering and System Safety*.
- Davoudian, K.; Wu, J.; and Apostolakis, G. (in press, b) "The work process analysis model (WPAM)," *Reliability Engineering and System Safety*.
- Deci, E.L. (1975) *Intrinsic Motivation*, Plenum, New York.
- Duffy, E. (1962) *Activation and Behavior*, Wiley, New York.
- Eisenhardt, K. (1985) "Control: Organizational and Economic Approaches," *Management Science*, **31**.
- Embrey, D.E. (1992) "Incorporating management and organizational factors into probabilistic safety assessment" *Reliability Engineering and System Safety*, **38**, 199-208.
- Festinger, L. (1957) *A Theory of Cognitive Dissonance*, Harper & Row, New York.
- Freud, S. (1966) *Psychopathology of Everyday Life*, Norton, New York.
- Freudenburg, W. (1992) "Nothing Recedes Like Success? Risk Analysis and the Organizational Amplification of Risks," *Risk – Issues in Health and Safety*, Winter.
- Freudenburg, W. (1988) "Perceived Risk, Real Risk: Social Science and the Art of Probabilistic Risk Assessment," *Science*, **242**, 44-49.
- Gaba, D.M. (1991) "Human performance issues in anesthesia patient safety," *Problems in Anesthesia*, **5**, 329-350.
- Greenhalgh, G. (1990) Focus shifts to organizational issues, *Nuclear Engineering International*, March.
- Grinker, R.R. and Spiegel, J.P. (1963) *Men Under Stress*, McGraw-Hill, New York (reprinted from 1945).
- Haber, S.B.; O'Brien, J.N.; Metlay, D.S.; Crouch, D.A. (1991) "Influence of organizational factors on performance reliability, Volume 1: Overview and detailed methodological development," Report to the U.S. Nuclear Regulatory Commission, NUREG/CR-5538, BNL-NUREG-52301.
- Haber, S.B.; O'Brien, J.N.; and Ryan, T.G. (1988) "Model Development for the Determination of the Influence of Management on Plant Risk," presented at the IEEE Fourth Conference on Human Factors and Power Plants, Monterey, CA, June 5-9.
- Heimer, C.A. (1988) "Social structure, psychology, and the estimation of risk," *Annual Review of Sociology*, **14**, 491-519.
- Henley, E.J. and Kumamoto, H. (1981) *Reliability Engineering and Risk Assessment*, Prentice-Hall Englewood Cliffs, NJ.
- Herriott, S.R., Levinthal, D., and March, J.G. (1988) "Learning from experience in organizations," in J.G. March (ed.), *Decisions and Organizations*, Basil Blackwell, Cambridge.
- Hogarth, R. (1980) *Judgment and Choice: the Psychology of Decision*, Wiley, Chichester.



- International Atomic Energy Association (1991) *Safety Culture: a report by the International Nuclear Safety Advisory Group*, Vienna.
- Jacobs, R. and Haber, S.B. (in press) "Organizational Processes and Nuclear Power Plant Safety," *Reliability Engineering and System Safety*.
- Janis, I. (1972) *Victims of Groupthink*, Houghton-Mifflin, Boston.
- Janis, I. and Mann, L. (1977) *Decision Making: A Psychological Analysis of Conflict, Choice, and Commitment*, Free Press, New York.
- Kahneman, D. and Tversky, A. (1979) "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, **47**, 263-91.
- Karmarkar, U. (1978) "Subjectively Weighted Utility: A Descriptive Extension of the Expected Utility Model," *Organizational Behavior and Human Performance*, **21**, 61-72.
- Kletz, T.A. (1985) *An engineer's view of human error*. The Institute of Chemical Engineers, Warwickshire, England.
- La Porte, T.R. and Consolini, P.M. (1991) "Working in practice but not in theory: theoretical challenges of 'high-reliability organizations'," *Journal of Public Administration Research and Theory*, **1**, 19-47.
- Levitt, R.E. (1975) "The effect of top management on safety in construction," Doctoral dissertation, Stanford University.
- Levitt, R.E., Cohen, G.P., Kunz, J.C., Nass, C.I., Christiansen, T., Jin, Y. (1994) "The virtual design team: simulating how organizational structure and information processing tools affect team performance," in K.M. Carley and M.J. Prietula (eds.), *Computational Organization Theory*, Lawrence Erlbaum Associates, Hillsdale, NJ.
- Lounamaa, P.H. and March, J.G. (1987) "Adaptive coordination of a learning team," *Management Science*, **33**, 107-23.
- Luce, R.D. and Raiffa, H. (1956) *Games and Decisions*, Wiley, New York.
- March, J. (1988) *Decisions and Organizations*, Blackwell, Cambridge.
- March, J. and Simon, H. (1958) *Organizations*, Wiley, New York.
- Morone, J.G. and Woodhouse, E.J. (1986) *Averting Catastrophe: Strategies for Regulating Risky Technologies*, University of California Press, Berkeley, CA.
- Moore, W.H. (1994) "The grounding of Exxon Valdez: an examination of the human and organizational factors," *Marine Technology* **31**, 41-51.
- Nagel, D.C. (1988) "Human error in aviation operations," in E.L. Wiener and D.C. Nagel, (eds.), *Human factors in aviation*, Academic Press, New York.
- Neiss, R. (1988) "Reconceptualizing Arousal: Psychobiological States in Motor Performance," *Psychological Bulletin*, **103**, 345-66.
- Newell, A.; Rosenbloom, P.S.; and Laird, J.E. (1987) "SOAR: an architecture for general intelligence," *Artificial Intelligence*, **33**, 1-64.
- Norman, D.A. (1988) *The Design of Everyday Things*, Doubleday, New York.
- Norman, D.A. (1981) "Categorization of Action Slips," *Psychological Review*, **88**, 1-15.

- Nuclear Regulatory Commission (1975) *Reactor safety study : an assessment of accident risks in U.S. commercial nuclear power plants*, USNRC report WASH-1400, National Technical Information Service, Washington, D.C.
- Okrent, D. and Arueti, S. (1990) "Can We Bring Quality of Management into PRAs?" in *New Risks*, L. Cox, and P. Ricci, (eds.), Plenum Press, New York.
- Ouchi, W. (1979) "A conceptual framework for the design of organization control mechanisms," *Management Science*, **25**, 833-48.
- Paté-Cornell, M.E. (1990) "Organizational Aspects of Engineering System Safety: The Case of Offshore Platforms," *Science*, **250**, 1210-1217.
- Paté-Cornell, M.E. (1989) "Organizational Extensions of PRA Models and NASA Application," Proceedings of PSA '89 International Topical Meetings on Probability, Reliability, and Safety Assessment, April, Pittsburgh, Pennsylvania, 218-225.
- Paté-Cornell, M.E. and Bea, R. (1992) "Management Errors and System Reliability: A Probabilistic Approach and Application to Offshore Platforms," *Risk Analysis*, **12**, 1-18.
- Paté-Cornell, M.E. and Fischbeck, P. (1990) "Safety of the Thermal Protection System of the Space Shuttle Orbiter: Quantitative Analysis and Organizational Factors," Report to the National Aeronautics and Space Administration.
- Paté-Cornell, M.E.; Lakats, L.; and Murphy, D. (1994a, under review) "Anesthesia Patient Risk: a Quantitative Approach to Organizational Factors and Risk Management Options."
- Paté-Cornell, M.E.; Murphy, D.; and Lakats, L. (1994b) "Anesthesia Patient Safety: Probabilistic Risk Analysis and Benefits of Organizational Improvements," report to the Anesthesia Patient Safety Foundation, October.
- Paté-Cornell, M.E.; Murphy, D.; Lakats, L.; and Gaba, D. (1994c, under review) "Patient Risk in Anesthesia: Probabilistic Risk Analysis, Management Effects and Improvements."
- Payne, J. (1973) "Alternative approaches to decision making under risk: moments versus risk dimensions," *Psychological Bulletin*, **80**, 439-53.
- Perrow, C. (1984) *Normal Accidents*, Basic Books.
- Perrow, C. (1983) "The Organizational Context of Human Factors Engineering," *Administrative Science Quarterly*, **28**, 521-41.
- Perrow, C. (1981) "Normal Accident at Three Mile Island," *Society*, July/August, 17-26.
- Rasmussen, J. (1990) "The role of error in organizing behaviour," *Ergonomics*, **33**, 1185-99.
- Rasmussen, J. (1983) "Skills, Rules, Knowledge: Signals, Signs and Symbols and Other Distinctions in Human Performance Models," *IEEE Transactions: Systems, Man and Cybernetics*, SMC-13, 257-267.
- Rasmussen, J. (1982) "Human errors: A taxonomy for describing human malfunction in industrial installations," *Journal of Occupational Accidents*, **4**, 311-35.
- Reason, J. (1990a) *Human Error*, Cambridge University Press, Cambridge.
- Reason, J. (1990b) "The age of the organizational accident," *Nuclear Engineering International*, July, 18-19.

- Reason, J. (1987a) "Generic Error Modeling System (GEMS): A Cognitive Framework for Locating Common Human Error Forms," in J. Rasmussen, K. Duncan, and J Leplat (eds.) *New Technology and Human Error*, Wiley, Chichester.
- Reason, J. (1987b) "A Framework for Classifying Errors," in J. Rasmussen, K. Duncan, and J Leplat (eds.) *New Technology and Human Error*, Wiley, Chichester.
- Reason, J. (1987c) "A Preliminary Classification of Mistakes," in J. Rasmussen, K. Duncan, and J Leplat (eds.) *New Technology and Human Error*, Wiley, Chichester.
- Regan, P.J. (1993) "Design and construction of normative risk management and decision systems," Doctoral dissertation, Stanford University.
- Report of the Presidential Commission on the Space Shuttle Challenger Accident (1986), Government Printing Agency, Washington, DC.
- Roberts, K.H. (1990) "Some Characteristics of One Type of High Reliability Organization," *Organization Science*, **1**, 160-76.
- Roberts, K.H. (1989) "New challenges in organizational research: high reliability organizations," *Industrial Crisis Quarterly*, **3**, 111-25.
- Rochlin, G.; La Porte, T.; and Roberts, K. (1987) "The Self-Designing High Reliability Organization: Aircraft Carrier Flight Operations at Sea," *Naval War College Review*, Autumn, 76-90.
- Runciman, W.B.; Sellen, A.; Webb, R.K.; Williamson, J.A.; Currie, M.; Morgan, C.; and Russel, W.J. (1993a) "Errors, incidents and accidents in anaesthetic practice," *Anaesthesia and Intensive Care*, **21**, 506-519.
- Runciman, W.B.; Webb, R.K.; Klepper, I.D.; Lee, R.; Williamson, J.A.; and Barker, L. (1993b) "Crisis management: validation of an algorithm by analysis of 2000 incident reports," *Anaesthesia and Intensive Care*, **21**, 506-519.
- Sagan, S.D. (1993) *The Limits of Safety : Organizations, Accidents, and Nuclear Weapons*, Princeton University Press, Princeton, NJ.
- Sah, R.K. and Stiglitz, J.E. (1986) "The architecture of economic systems: hierarchies and polyarchies," *American Economic Review*, **76**, 716-27.
- Savage, L. (1954) *The Foundations of Statistics*, John Wiley, New York.
- Schank, R.C. and Abelson, R. (1977) *Scripts, Plans, Goals, and Understanding*, Erlbaum, Hillsdale, NJ.
- Senders, J.W. and Moray, N.P. (1991) *Human Error: Cause, Prediction and Reduction*, Erlbaum, Hillsdale, NJ.
- Shanteau, J. (1975) "An Information-Integration Analysis of Risky Decision Making," in M. Kaplan and S. Schwartz, (eds.) *Human Judgment and Decision Processes*, Academic Press, New York.
- Simon, H. (1986) "Alternative visions of rationality," in H. Arkes and K Hammond (eds.) *Judgment and Decision Making*, Cambridge University Press, Cambridge.
- Simon, H. (1957) *Models of Man*, Wiley, New York.
- Simon, H. (1956) "Rational choice and the structure of the environment," *Psychological Review*, **63**, 129-38.
- Simon, H. (1955) "A behavioral model of rational choice," *Quarterly Journal of Economics*, **69**, 99-118.

- Slovic, P. and Lichtenstein, S. (1968) "Relative importance of probabilities and payoffs in risk taking," *Journal of Experimental Psychology*, **78**.
- Sommer, L. (1954) "Exposition of a New Theory on the Measurement of Risk," *Econometrica*, **22**, 23-26. English translation of Bernoulli, D. (1738) "Specimen Theoriae Novae de Mensura Sortis," *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, **5**, 175-92.
- Starbuck, W.H. and Milliken, F.J. (1988) "Challenger: fine-tuning the odds until something breaks," *Journal of Management Studies*, Oxford, England, **25**, 319-40.
- Staw, B.M.; Sandelands, L.E.; and Dutton, J.E. (1981) "Threat rigidity effects in organizational behavior: a multilevel analysis," *Administrative Science Quarterly*, **26**, 501-24.
- Swain, A.D. and Guttmann, H.E. (1983) *Handbook of Human Reliability Analysis with Emphasis on NPP Applications*. NUREG/CR-1278, USNRC.
- Tamuz, M. (1988) "Monitoring dangers in the air: Studies in ambiguity and information," Doctoral dissertation, Stanford University.
- Thompson, J. (1967) *Organizations in Action*, McGraw-Hill, New York.
- Tversky, A. and Kahneman, D. (1974) "Judgment under Uncertainty: Heuristics and Biases," *Science*, **185**, 1124-31.
- von Neumann, J. and Morgenstern, O. (1947) *Theory of Games and Economic Behavior*, Second edition, Princeton University Press, Princeton.
- Wagenaar, W.A. and Groeneweg, J. (1987) "Accidents at sea: Multiple causes and impossible consequences," *International Journal of Man-Machine Studies*, **27**, 587-98.
- Wahlstrom, B. and Swaton, E. (1991) "Influence of Organization and Management on Industrial Safety," in *The Influence of Organization and Management on the Safety of NPPs and Other Complex Industrial Systems*, Working paper WP-91-28, ILASA, July.
- Wallach, M.A.; Kogan, N.; and Bem, D. (1962) "Group influence on individual risk taking," *Journal of Abnormal and Social Psychology*, **65**, 75-86.
- Weick, K.E. (1990) "The Vulnerable System: An Analysis of the Tenerife Air Disaster," *Journal of Management*, **16**, 571-93.
- Weick, K. (1987) "Organizational Culture as a Source of High Reliability," *California Management Review*, **29**, 112-27.
- Williamson, A.M. and Feyer, A. (1990) "Behavioural epidemiology as a tool for accident research," *Journal of Occupational Accidents*, **12**, 207-22.
- Williamson, J.A.; Webb, R.K.; Sellen, A.; Runciman, W.B.; and Van der Walt, J.H. (1993) "Human failure: an analysis of 2000 incident reports," *Anaesthesia and Intensive Care*, **21**, 678-83.
- Wright, C. (1986) "Routine deaths: Fatal accidents in the oil industry," *Sociological Review*, **34**, 265-289.
- Wu, J.; Apostolakis, G.; and Okrent, D. (1991) "On the Inclusion of Organizational and Managerial Influences in Probabilistic Safety Assessments of Nuclear Power Plants," in *The Analysis, Communication, and Perception of Risk*, B. Garrick and W. Gekler, (eds.), Plenum Press, New York.

- Yerkes, R.M. and Dodson, J.D. (1908) "The relation of strength of stimulus to rapidity of habit formation," *Journal of Comparative Neurology and Psychology*, **18**, 459-82.
- Zajonc, R.B. (1965) "Social Facilitation," *Science*, **149**, 269-74.
- Zakay, D. (1986) "An evaluation of the utilization of multi-attribute utility (MAU) models in managerial decision making," in O. Brown, Jr. and H.W. Hendrick (eds.), *Human Factors in Organizational Design and Management – II*, North-Holland, Amsterdam.