CSERIAC

CREW SYSTEM ERGONOMICS INFORMATION ANALYSIS CENTER

SOAR

CSERIAC 94-01

State-of-the-Art Report

Behind Human Error: Cognitive Systems, Computers, and Hindsight

David D. Woods, Ph.D.
Leila J. Johannesen
Richard I. Cook, M.D.
Nadine B. Sarter
The Ohio State University

December 1994

CSERIAC

ARMY NAVY AIR FORCE NASA FAA NATO

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
December 1994

3. REPORT TYPE AND DATES COVERED
State-of-the-Art Report

4. TITLE AND SUBTITLE
Behind Human Error: Cognitive Systems, Computers, and Hindsight

5. FUNDING NUMBERS
DLA900-88-0393

6. AUTHOR(S)
David D. Woods, Ph.D.
Leila J. Johannesen
Richard I. Cook, M.D.
Nadine B. Sarter

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Dayton Research Institute
300 College Park
Dayton, OH 45469-0157

8. PERFORMING ORGANIZATION REPORT NUMBER
CSERIAC SOAR 94-01

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Defense Technical Information Center
DTIC/AI
Cameron Station
Alexandria, VA 22304-6145

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.
Available solely through CSERIAC for $39.00 (U.S.).

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This report goes beyond a characterization of human error as a causal factor of accidents. It discusses the larger system within which practitioners operate and shows how "blunt end" factors such as organizational processes and technology design impact the cognition and behavior of those at the "sharp end." Examples from various domains are used to illustrate deficiencies in computerized devices, which can lead to breakdowns in interaction, such as mode error. Reasons are presented for why these deficiencies as "latent failures" can exist without giving rise to accidents. Also discussed is the role of outcome knowledge in the attribution of error.

14. SUBJECT TERMS
Human error, accidents, human-computer interaction deficiencies, hindsight, latent failures

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT
UNCLASSIFIED

18. SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

19. SECURITY CLASSIFICATION OF ABSTRACT
UNCLASSIFIED

20. LIMITATION OF ABSTRACT
UNLIMITED

CSERIAC

State-of-the-Art Report

BEHIND HUMAN ERROR: COGNITIVE SYSTEMS, COMPUTERS, AND HINDSIGHT

David D. Woods, Ph.D.
Leila J. Johannesen
Richard I. Cook, M.D.
Nadine B. Sarter

The Ohio State University
Columbus, Ohio

December 1994

CSERIAC SOAR Series Editor:
Aaron W. Schopper Ph.D.

Crew Systems Ergonomics
Information Analysis Center
Wright-Patterson Air Force Base, Ohio

CSERIAC
ARMY NAVY AIR FORCE NASA FAA NATO

Additional copies of this state-of-the-art report (SOAR) are available for $ 39 U.S. (plus shipping) from:

CSERIAC Program Office
AL/CFH/CSERIAC Bldg 248
2255 H Street
Wright-Patterson AFB OH 45433-7022
(513) 225-4842

CSERIAC SOAR 94-01

ABOUT CSERIAC

The Crew System Ergonomics Information Analysis Center (CSERIAC) is the gateway to worldwide sources of up-to-date human factors information and technologies for designers, engineers, researchers, and human factors specialists. CSERIAC provides a variety of products and services to government, industry, and academia promoting the use of ergonomics in the design of human-operated equipment and manned systems.

CSERIAC's primary objective is to acquire, analyze, and disseminate timely information on ergonomics. On a cost-recovery basis, CSERIAC will:

- Distribute ergonomic technologies and publications
- Perform customized bibliographic searches and reviews
- Prepare state-of-the-art reports and critical reviews
- Conduct specialized analyses and evaluations
- Organize and/or conduct workshops and conferences

CSERIAC is a Department of Defense Information Analysis Center sponsored hy the Defense Technical Information Center. It is hosted hy the Armstrong Laboratory Human Engineering Division and operated by the University of Dayton Research Institute.

To obtain additional copies of this report or further information, contact:

CSERIAC Program Office
AL/CFH/CSERIAC Bldg 248
2255 H Street
Wright-Patterson AFB OH 45433-7022
(513) 255-4842
DSN 785-4842
FAX (513) 255-4823
DSN FAX (513) 785-4823

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

FOREWORD

A Parable

There was once a man who operated a complex system. The system required the man to enter a number when a screen flashed *A* and enter a different number when a screen flashed *B*. One day the operator entered the *A* number when the *B* cue appeared. The number told the machine to blow up instead of to shut down.

It took people a long time to figure out what had happened. When they thought they had, a lot of people concluded that the accident was caused by "operator error," by which they meant that the man who entered the wrong number had made an error, and that was all one needed to know. Some people said the man should have checked himself. Some said he should have been better trained, and some even said he had been ill-suited for the job.

But some reputedly enlightened people came along and said it wasn't fruitful to blame the operator. They spoke of such things as good design and understanding how humans solve problems, of workload and of competitions among goals. People gathered around them, for they seemed to utter truth. The enlightened people said the failures had been made by the organization, which is to say by people such as managers and designers.

Thereupon the startled management people cried, "But we didn't enter the inappropriate numbers."

"No, but you created the poor conditions for the entering of the numbers," said the enlightened people.

But the designers called out, "We followed the commandments of our profession so we are blameless!"

To them the enlightened people said, "Revise your profession."

Whereupon all the organizational and design people cried, "But we didn't know what the consequences would be!" And someone else said, "We all have deadlines and budgets to meet, you know."

"And remember you have the benefit of hindsight," said a small voice in the crowd.

To this, the enlightened people didn't know what to say, for there seemed to be some truth in these lamentations.

"What should we do differently next time?" "And how do we know that the fixes you suggest will be worth the fortunes required?" asked the organizational people.

To these questions, also, the enlightened people were hesitant in their response.

Thereupon, there arose a tremendous confusion and all the people began speaking in different languages and they could not understand one another.

This is the state we are in now.

PREFACE

An International And Cross-Disciplinary Discussion On Human Error

One of the factors that greatly heightened the visibility of the label "human error" was the Three Mile Island accident in the spring of 1979. This highly publicized accident, and others that came after, drew the attention of the engineering, psychological, social science, regulatory communities, and of the public to issues surrounding human error. The result was an intense cross-disciplinary and international consideration of the topic of the human contribution to risk over the last 15 years. One can mark the emergence of this cross-disciplinary and international consideration of error with thc "clambake" conference on human error organized by John Senders and Ann Crichton-Harris at Columbia Falls, Maine in 1980 and with the publication of Don Norman's and Jim Reason's work on slips and lapses (Norman, 1981; Reason and Mycielska, 1982).

Of course, as always, there was a great deal that led up to these events and publications, e.g., a longer tradition of concern with human error in human factors (Fitts and Jones, 1947; Singleton, 1973), in laboratory studies of decision biases (Tversky and Kahnemann, 1974), and in risk analysis (Dougherty and Fragola, 1990).

The discussions have continued in a wide variety of forums, including the Bellagio workshop on human error in 1983 (cf., Senders and Moray, 1991), the Bad Homburg workshop on new technology and human error in 1986 (Rasmussen, Duncan, and Leplat, 1987), the World Bank meetings on safety control and risk management in 1988 and 1989 (e.g., Rasmussen and Batstone, 1989), Reason's elaboration of the latent failure approach (1990), thc dcbate triggered by Dougherty's editorial in *Reliability Engineering and System Safety* (1990), Hollnagel's *Human Reliability Analysis: Context and Control* (1993) and a series of four workshops sponsored by a U.S. National Academy of Sciences panel from 1990 to 1993 that examined human error from individual, team, organizational, and design perspectives.

The cross-disciplinary and international consideration of the topic of error rc-examined common assumptions, dcveloped and extended concepts and theoretical frameworks. Various participants have used these frameworks to gather data from field experiments and to examine incidents and accidents in a new light. The result is a new look at the human contribution to safety and to risk. This "new look" is not conceptually homogenous. There is no complete consensus among the participants in these discussions, although there are some generally commonly held assumptions and interpretations of the evidence. It is not a mature body of work, but rather a road map for posing new questions and for examining unresolved issues in new ways.

Our approach has been heavily influenced by this debate. In fact, we attempt to provide a summary of the basic premises that have emerged from it, in Chapter 2. This overview is essential to provide the reader with some perspective on a set of concepts that reverberate throughout the other parts of this book.

The Diversity Of Perspectives On Human Error

Human error is a very elusive concept. Over the last 13 years we have been involved in many discussions about error with specialists having widely different perspectives. Some of the professions interested in error are operators, regulators, system developers, probability reliability assessment (PRA) specialists, experimental psychologists, accident investigators, and researchers who directly study "errors." We are continually impressed by the extraordinary diversity of notions and interpretations that have been associated with the label "human error." The parable included as a foreword tries to capture some of the kinds of interchanges that can arise among representatives of different perspectives.

The label "human error" is inextricably bound up with extra-research issues. The interest in the topic derives from the real world, from the desire to avoid disasters. The potential changes that could be made in real-world hazardous systems to address a "human error problem" inevitably involve high consequences for many stakeholders. Huge investments have been made in technological systems that cannot be easily changed because some researcher claims that the incidents relate to design flaws that encourage the possibility of human error. When a researcher claims that a disaster is due to latent organizational factors and not to the proximal events and actors, he or she is asserting a prerogative to re-design the jobs and responsibilities of hundreds of workers and managers. The factors seen as contributors to a disaster by a researcher could be drawn into legal battles concerning financial liability for the damages and losses associated with an accident. Laboratory researchers may offer results on biases found in the momentary reasoning of college students while performing artificial tasks. But how much these biases "explain" the human contribution to a disaster is questionable, particularly when the researchers have not examined the disaster, or the anatomy of disasters and near misses in detail (e.g., Klein, 1989).

One cannot pretend that research in this area can be conducted by disinterested, purely objective, detached observers. Researchers, like other people, have certain goals that influence what they see. When the label "human error" becomes the starting point for investigations, rather than a conclusion, the goal of the research must be how to produce change in organizations, in systems, and in technology to increase safety and reduce the risk of disaster. Whether researchers want to recognize it or not, we are participant observers.

Our experiences in the cross-disciplinary and international discussions convince us, first, that trying to define the term "error" is a bog that quite easily generates unproductive discussions both among researchers and between researchers and the consumers of research (such as regulators, public policy makers, practitioners, and designers). If one pays close attention to the muck in the bog of what is human error, one sees great differences of perspective and many misconceptions with respect to the evidence that has been gathered about erroneous actions and system disasters. One sees that there is a huge breadth of the human performance and human-machine system issues that can become involved in discussions under the rubric of the term "human error." As a result, one cannot get onto productive tracks about error, its relationship to technology change, prediction, modeling, and countermeasures, without directly addressing the varying perspectives, assumptions, and misconceptions of the different people interested in the topic of human error. Therefore, one of the first things that we provide is a summary of the assumptions and basic concepts that have emerged from the cross-disciplinary and international discussions and the research that they provoked.

We believe that this is important in its own right hecause our experienees in the last year or two indieate that the results of the cross-disciplinary work of the last 15 years have had remarkably little impact on industries, engineering groups that operate or develop systems, and regulatory bodies. In addition, it does not seem to have impaeted deeisions about how to manage technology change or impacted public debates over accidents and hazardous technologies. Don Norman expressed his frustration concerning the lack of impact on system designers in a commentary for the Communications of the ACM (Norman, 1990a). The newer research results have not penetrated very far, at least not into the variety of groups that we come into contact with. Discussions of error with or hy these groups exhihit a set of "folk" notions that are generally quite inconsistent with the results of the last 15 years. Not surprisingly, these folk theories are quite prevalent in design, engineering, and practitioner communities.

At the root, to us, the diversity of approaches to the topic of error is symptomatic that "human error" is not a well defined category of human performance. Attributing error to the actions of some person, team, or organization is fundamentally a social and psychological process and not an objective, teehnical one. Chapter 6 discusses some of the problems in attributing error after the fact, including the role of hindsight and outeome biases.

It is important to uncover implicit, unexamined assumptions about "human error" and the human contrihution to system failures. Making these assumptions explicit and contrasting them with other assumptions and research results can provide the impetus for a substantive theoretieal and research debate. Taking into account the range of assumptions and beliefs in different communities about "human error" and system disaster also aids communication with a broad audienee. Our goal is to capture and synthesize some of the results of the recent intense examination of the label "human error," particularly with respect to cognitive factors, the impact of computer technology, and the effect of the hindsight bias on error analysis.

ACKNOWLEDGMENTS

This volume is dedicated to those who work at the sharp end.

The ideas in this book have developed from a complex web of interdisciplinary interactions. We are indebted to John Senders, Jens Rasmussen, Jim Reason, Neville Moray, and Don Norman for their efforts leading us all down a different path. There have been many participants in the various discussions trying to make sense of human error over the last 10 years who have influenced directly or indirectly the stance towards error developed in this volume: Véronique De Keyser, Jim Easter, David Embrey, Baruch Fischhoff, Zvi Lanir, Todd Laporte, Dick Pew, Emilie Roth, John Wreathall, and many others.

A special thanks is due to Erik Hollnagel. He helped pilot a way through the bog of human error, and always reminded us that "human error" is just a label.

The end result of this volume, as in any project, depends greatly on the quality of the error detection and recovery mechanisms involved. In this regard we owe a great debt to Charles Billings. His perceptive mind and eye helped us sharpen the ideas and prose in many ways.

Many of the chapters draw heavily from various projects in which we have studied the relationship of human performance, cognitive systems, problem demands, and computerized technology in demanding fields of practice such as commercial aviation, anesthesiology, nuclear power, and space systems operations. We are always particularly grateful to the various sharp-end practitioners who have participated in and assisted these investigations.

For various parts of this book we have borrowed from material published in the papers and technical reports resulting from those projects, including:

- D.D. Woods. (1994). Some observations from studying cognitive systems in context [Keynote Address]. In *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.

- R.I. Cook and D.D. Woods. (1994). Operating at the sharp end: The complexity of human error. In M.S. Bogner (Ed.), *Human error in medicine*, Hillsdale, NJ: Erlbaum.

- D.D. Woods, R.I. Cook, and N. Sarter. *(1992). Clumsy automation, practitioner tailoring and system failure* (Tech. Rep. 92-TR-04). Columbus, OH: The Ohio State University , Department of Industrial & Systems Engineering, Cognitive Systems Engineering Laboratory.

- N. Sarter and D.D. Woods. (In press). How in the world did we get into that mode? Mode error and awareness in supervisory control. *Human Factors (Special Issue on Situation Awareness).*

- V. De Keyser and D.D. Woods. (1990). Fixation errors: Failures to revise situation assessment in dynamic and risky systems. In A.G. Colombo and A. Saiz de Bustamante (Eds.), *System reliability assessment.* Dordrechts, The Netherlands: Kluwer.

- D.D. Woods. (1990). Modeling and predicting human error. In J. Elkind, S. Card, J. Hochberg, and B. Huey (Eds.), *Human performance models for computer-aided engineering.* New York: Academic.

- D.D. Woods. (1990). Risk and human performance: measuring the potential for disaster. *Reliability engineering and system safety, 29*, 387-405.

- D.D. Woods. (In press). Towards a theoretical base for representation design in the computer medium: Ecological perception and aiding human cognition. In J. Flach, P. Hancock, J. Caird, and K. Vicente (Eds.), *An ecological approach to human machine systems I: A global perspective.* Hillsdale, NJ: Erlbaum.

ABOUT THE AUTHORS

David D. Woods, Ph.D., is co-founder and co-director of the Cognitive Systems Engineering Laboratory at The Ohio State University. He has contributed to the emergence and growth of cognitive engineering through studies of human-machine cognitive systems "in the wild" for 15 years. His specialty is studies of cognitive work in different kinds of control centers:

- commercial aviation–pilot interaction with cockpit automation,
- space vehicle management–how to make AI systems team players,
- nuclear power control rooms–operator performance in dynamic fault management, and
- critical care medicine–clumsy automation in the operating room.

Dr. Woods has also developed many new design concepts, holds 5 patents, and has designed systems for aiding human performance through computer-based visualizations and decision aids in various settings. He is a Fellow of the American Psychological Association, American Psychological Society, and the Human Factors and Ergonomics Society, and has advised US government agencies on human factors. These agencies include NASA, Nuclear Regulatory Commission, Department of Energy, and FAA.

Leila Johannesen is a cognitive engineer/psychologist who joined the Cognitive Systems Engineering Laboratory in 1990. Working primarily within the space domain, she has focused on how to design more cooperative interaction among humans and intelligent fault-management systems. She has had industry experience as a research scientist specializing in human-computer interaction.

Richard I. Cook, M.D., is a practicing anesthesiologist and cognitive systems engineer, who has been associated with the Cognitive Systems Engineering Laboratory since 1988. He has consulted for various companies on the interaction between technology and skilled human performance, and has written several papers and book chapters on the interaction between physicians and technology in the operating room. Before becoming a doctor, he worked as an engineer designing and developing operating-room monitoring systems.

Nadine Sarter has several years of experience in studying human-automation interaction both in the maritime domain and in the field of commercial aviation. In 1989, she joined the Cognitive Systems Engineering Laboratory, where she specializes in field studies on the interaction between pilots and advanced automated cockpit systems. Her major interests are the evolution of automation properties and philosophies and their impact on the coordination and cooperation between man and machine. In particular, her work focuses on pilots' mode and situation awareness, strategies of automation management, attention allocation in highly dynamic multi-display environments, and the development of new approaches to training for advanced automated cockpits.

1

INTRODUCTION

The Human Error Problem

Disasters in complex systems, such as the destruction of the reactor at Three Mile Island, the explosion onboard Apollo 13, the destruction of the space shuttle Challenger, the Bhopal chemical plant disaster, the Herald of Free Enterprise ferry capsizing, the Clapham Junction railroad disaster, the grounding of the tanker Exxon Valdez, crashes of highly computerized aircraft at Bangalore and Strasbourg, the explosion at the Chernobyl reactor, AT&T's Thomas Street outage, as well as more numerous serious incidents which have only captured localized attention, have left the technologist perplexed. From a narrow, technology-centered point of view, incidents seem more and more to involve mis-operation of engineered systems that are otherwise functional. Small problems seem to cascade into major incidents. Systems with minor problems are managed into much more severe incidents. What stands out in these cases is the human element.

Human error is over and over again cited as a major contributing factor or cause of incidents. Most people accept the term "human error" as one category of potential causes for unsatisfactory activities or outcomes. Human error as a cause of bad outcomes is used in engineering approaches to the reliability of complex systems (probabilistic risk assessment) and is widely used as a basic category in incident reporting systems in a variety of industries. For example, surveys of anesthetic incidents in the operating room have attributed between 70 and 75% of the incidents surveyed to the human element (Cooper, Newbower, and Kitz, 1984; Chopra, Bovill, Spierdijk, and Koornneef, 1992; Wright, Mackenzie, Buchan, Cairns, and Price, 1991). Similar incident surveys in aviation have attributed over 70% of incidents to crew error (Boeing, 1993). In general, incident surveys in a variety of industries attribute high percentages of critical events to the category "human error" (for example, see Hollnagel, 1993).

The result is the widespread perception of a human error problem.[1] The typical belief is that the human element is separate from the system in question and, hence, that problems reside either in the human side or in the engineered side of the equation (Woods, 1990b). Incidents attributed to human error then become indicators that the human element is unreliable. This view implies that solutions to a human error problem reside in changing the people or their role in the system. To cope with this perceived unreliability of people, the implication is that one should reduce or regiment the human role in managing the potentially hazardous system. In general, this is attempted by enforcing standard practices and work rules, by exiling culprits, by policing of practitioners, and by using automation to shift activity away from people. Note that this view assumes that the overall tasks and system remain the same regardless of the allocation of tasks to people or to machines and regardless of the pressures managers or regulators place on the practitioners.[2]

For those who accept human error as a potential cause, the answer to the question, "What is human error?" seems self evident. Human error is a specific variety of human performance that is so clearly and significantly substandard and flawed when viewed in retrospect that there is no doubt that it should have been viewed by the practitioner as substandard *at the time the act was committed or omitted.* The judgment that an outcome was due to human error is an attribution that (a) the human performance immediately preceding the incident was unam-

[1] One aviation organization concluded that to make progress on safety, we must have a better understanding of the so-called human factors which control performance simply because it is these factors which predominate in accident reports (Aviation Daily, November 6, 1992). Similar statements could be extracted from many industries.

[2] The term practitioner refers to a person engaged in the practice of a profession or occupation (Webster's, 1990).

biguously flawed and (b) the human performance led directly to the negative outcome.

But in practice, things have proved not to be this simple. The label "human error" is very controversial (e.g., Hollnagel, 1993). Attribution of error is a *judgment* about human performance. These judgments are rarely applied except when an accident or series of events have occurred that ended with a bad outcome or nearly did so. Thus, these judgments are made *ex post facto*, with the benefit of *hindsight* about the outcome or near miss. This factor makes it difficult to attribute specific incidents and outcomes to human error in a consistent way. Fundamental questions arise. When precisely does an act or omission constitute an error? How does labeling some act as a human error advance our understanding of why and how complex systems fail? How should we respond to incidents and errors to improve the performance of complex systems? These are not academic or theoretical questions. They are close to the heart of tremendous bureaucratic, professional, and legal conflicts and are tied directly to issues of safety and responsibility. Much hinges on being able to determine how complex systems have failed and on the human contribution to such outcome failures. Even more depends on judgments about what means will prove effective for increasing system reliability, improving human performance, and reducing or eliminating erroneous actions.

Studies in a variety of fields show that the label "human error" is prejudicial and unspecific. It retards rather than advances our understanding of how complex systems fail and the role of human practitioners in both successful and unsuccessful system operations. The investigation of the cognition and behavior of individuals and groups of people, not the attribution of error in itself, points to useful changes for reducing the potential for disaster in large, complex systems. Labeling actions and assessments as errors identifies a symptom, not a cause; the symptom should call forth a more in-depth investigation of how a system comprising people, organizations, and technologies both functions and malfunctions (Rasmussen et al., 1987; Reason, 1990; Hollnagel, 1991b; 1993).

Consider this episode which apparently involved a human error and which was the stimulus for one of earliest developments in the history of experimental psychology. In 1796 the astronomer Maskelyne fired his assistant Kinnebrook because the latter's observations did not match his own. This incident was one stimulus for another astronomer, Bessel, to examine empirically individual differences in astronomical observations. He found that there were wide differences across observers given the methods of the day and developed what was named the "personal equation" in an attempt to model and account for these variations (see Boring, 1950). The full history of this episode foreshadows the latest results on human error. The problem was not that one person was the source of errors. Rather, Bessel realized that the standard assumptions about inter-observer accuracies were wrong. The techniques for making observations at this time required a combination of auditory and visual judgments. These judgments were heavily shaped by the tools of the day–pendulum clocks and telescope hairlines, in relation to the demands of the task. In the end, the solution was not dismissing Kinnebrook, but rather searching for better methods for making astronomical observations, re-designing the tools that supported astronomers, and re-designing the tasks to change the demands placed on human judgment.

The results of the recent intense examination of the human contribution to safety and to system failure indicate that the story of human error is markedly complex. For example:

- the context in which incidents evolve plays a major role in human performance,
- technology can shape human performance, creating the potential for new forms of error and failure,
- the human performance in question usually involves a set of interacting people,
- the organizational context creates dilemmas and shapes trade-offs among competing goals,
- the attribution of error after-the-fact is a process of social judgment rather than an objective conclusion.

Our Approach

The goal of this book is to go behind the label "human error." It may seem simpler merely to attribute poor outcomes to human error and stop there; the swirl of factors and issues behind the label may seem very eomplex. But it is in the examination of these deeper issues that one ean learn how to improve the performance of large, complex systems.

There are three main themes that we will explore behind the label of human error:

- the role of eognitive system factors in incidents (see Chapter 4),
- how the clumsy use of computer technology can increase the potential for erroneous actions and assessments (see Chapter 5),
- the hindsight bias and how attributions of error are a social and psychologieal judgment process rather than a matter of objective fact (see Chapter 6).

The book is organized into four basic parts. The first part, Chapter 2, presents a set of basie premises or themes that reeur frequently throughout the book and that summarize many of the important ideas behind the label of human error. This chapter ean be interpreted in two ways. It provides an introduction to the later chapters by presenting basic concepts and recurring themes. This is important because many of the ideas detailed in this volume depend intimately on each other. But this chapter can also be interpreted as an overview of the results of the intense and cross-disciplinary examination of error and disaster that has been going on since about 1980. As a result, this ehapter provides a kind of summary of many of the important ideas behind the label of human error. If a reader needs an overview of developments on human error, this is the place.

One of these basic concepts is the latent failure model of complex system breakdown (Reason, 1990). This concept is fundamental to the diseussion of eognitive system faetors, how the elumsy use of eomputer technology influenees the potential for error, and the operation of the hindsight bias in the process of attributing causes to incidents. As a result, Chapter 3 provides a brief introduction and overview of the concept.

Cognitive Systems

The demands that large, complex systems operations place on human performance are mostly cognitive. In the second part of the book (Chapter 4) we have chosen to focus on cognitive factors related to the expression of expertise and error. The difference between expert and inexpert human performance is shaped, in part, by three classes of cognitive factors: knowledge factors, attentional dynamics, and strategic factors. However, these cognitive factors do not apply just to an individual, but also to teams of practitioners. In addition, the larger organization places constraints that shape how practitioners meet the demands of that field of practice.

One of the basic themes that has emerged in more recent work on error is the need to model team and organizational factors. Chapter 4 integrates individual, team, and organizational perspectives by viewing operational systems as distributed and joint human-machine cognitive systems. It also lays out the cognitive processes carried out across a distributed system that govern the expression of expertise as well as error in real systems. It explores some of the ways that these processes go off track or break down and increase the vulnerability to erroneous actions.

Computers

The third part of the book addresses the clumsy use of new technological possibilities in the design of computer-based devices and shows how these design errors can create the potential for erroneous actions and assessments. Some of the questions addressed in Chapter 5 include:

- What are these classic design errors in human-computer systems, computer-based advisors, and automated systems?
- Why do we see them so frequently in so many settings?
- How do devices with these characteristics shape practitioner cognition and behavior?
- How do practitioners cope with the complexities introduced by clumsy use of technological possibilities?
- What do these factors imply about the human contribution to risk and to safety?

We will refer frequently to mode error as an exemplar of the issues surrounding the impact of computer technology and error, especially in Chapter 5. We use this topic as an example extensively because it is an error form that exists only at the intersection of people and technology. Mode error requires a device where the same action or indication means different things in different eontexts (i.e., modes) and a person who loses track of the currcnt context. But there is a second and perhaps more important reason why we have ehosen this error form as a central exemplar. If we as a eommunity of researchers eannot get design and development organizations to acknowledge, deal with, reduce, and better cope with the proliferation of complex modes, then we fear there is no issue where we ean shift design resources and priorities to include a user-eentered point of view.

Hindsight

The fourth part of the book cxamines how the hindsight bias affeets the possibilities for error analysis. It shows how attrihutions of error are a social and psychological judgment process rather than a matter of objeetive faet.

The latent failure model points out that there are many faetors that contribute to incidents and disasters. Whieh of these many factors we focus on are the products of *human* processes of causal attribution. What we identify as eauses dcpends on who we are communicating to, on the assumed contrast cases or causal haekground for that exchange, on the purposes of the inquiry, and on knowledge of the outcome.

Hindsight bias is the tendency for people to eonsistently exaggerate what eould have been antieipated in foresigbt (Fiscbhoff, 1975). Studies have consistently shown that people have a tendency to judge the quality of a process by its outcome (Baron and Hershey, 1988). The information about outcome biases their evaluation of the process that was followed. Deeisions and actions followed hy a negative outeome will be judged more harshly than if tbe *same* decisions bad resulted in a neutral or positive outcome. Indeed this effeet is present even when those making the judgments bave been warned about the phenomenon and heen advised to guard against it.

The hindsight bias leads us to "construct . . . a map tbat shows only those forks in the road that we deeided to take, where we see the view from one side of a fork in the road, looking baek" (Lubar, 1993, p. 1168). Given knowledge of outcome, reviewers will tend to *simplify* the prohlem-solving situation that was actually faced by the practitioner. The dilemmas, the uncertainties, thc tradeoffs, thc attentional dcmands, and double binds faced by practitioners may be missed or under-emphasized when an incident is viewed in hindsight. Typically, hindsight bias makes it seem that participants failed to account for information or conditions that should havc bccn obvious or bchavcd in ways that were inconsistent with the (now known to bc) significant information. Possessing knowledge of thc outcomc, because of thc hindsight bias, trivializcs the situation confronting the practitioner and makcs the correct choice seem crystal clear.

The hindsight bias has strong implications for studying erroneous actions and assessments and for lcarning from system failures. If wc recognizc the role of hindsight and psychological proccsses of causal judgment in attributing error aftcr-thc-fact, then we can begin to devise new ways to study and learn from crror and system failure. Wc necd techniqucs to hclp us construct (1) an aerial vicw that reveals thc possible paths, those followed and those not taken, and (2) what thc vicw was like or would have been like had we stood on the road. These topics are covered in Chapter 6.

In many ways, the topics addresscd in cach chapter interact and dcpend on the concepts introduced in thc discussion of other topics from other chapters. For cxample, the ehapter on the elumsy use of computer tcchnology in some ways dcpcnds on knowledge of cognitive system factors, but in other ways it hclps to motivate thc cognitivc system framework. Therc is no requircment to movc lincarly from onc chaptcr to another. Jump around as your intcrests and goals suggest. However, rcading Chapter 2 first may hclp to provide an ovcrvicw of the basic concepts and assumptions that wcavc togcther across thc rest of thc book.

Two cavcats arc in order. First, wc primarily are intercstcd in how people form intentions to aet and how thcsc proeesses eontribute to error and expertise. This refers to how people decidc what to do as opposed to the processes involved in going from intention to action (the error forms related to the lattcr process are called slips of action).[3] Studies of actual critical incidcnts (c.g., Pew, Miller, and Feehrer, 1981;

[3] In part, slips and lapses will not be considered because extensive treatments are already available—Norman (1981), Reason and Mycielska (1982), and Baars (1992); for overviews on slips see Norman (1988) and Reason (1990).

Woods, OBrien, and Hanes, 1987; Cook, Woods, and McDonald, 1991; Reason, 1990) have shown that intention errors (sometimes termed cognitive errors) are a major contributor to the risk of disaster. Intention formation refers to the cognitive processes by which a set of agents decide on what actions are appropriate to carry out (information gathering, situation assessment, diagnosis, and response selection). Intention formation is important to risk and safety because, when an erroneous intention to act is formed, practitioners may not only omit correct acts, but they may also carry out other acts that would be appropriate given the perceived situation, but are, in fact, incorrect given the actual situation. This means that erroneous intention leads to a kind of common mode failure.

Second, we will not be concerned with work that goes under the heading of Human Reliability Analysis (HRA), because (a) such work is summarized elsewhere (e.g., Dougherty and Fragola, 1990), and (b) HRA has been dominated by the assumptions made for risk analysis of purely technological systems, assumptions that do not apply to people and human-machine systems very well. Third, an excellent re-examination of human reliability from a cognitive perspective has recently emerged (cf., Hollnagel, 1993).

2

BASIC PREMISES FOR RESEARCH ON HUMAN ERROR

Designing human error out of systems was one of the earliest activities of human factors (e.g., Fitts and Jones, 1947). Error counts have been used as a measure of performance in laboratory studies since the beginning of experimental psychology. In fact an episode involving a human error was the stimulus for one of the earliest developments in experimental psychology.[4] While error has a long history in human factors and experimental psychology, the decade of the 1980s marked the beginning of an especially energetic period for researchers exploring issues surrounding the label human error. This international and cross-disciplinary debate on the nature of erroneous actions and assessments has led to a new paradigm about what is error, how to study error, and what kinds of countermeasures will enhance safety. This chapter is an overview of these results. It also serves as an introduction to the later chapters by presenting basic concepts that recur frequently throughout the book.

Fourteen Premises

Traditionally, error has been seen as a thing in itself, a kind of cause of incidents, a meaningful category that can be used to aggregate specific instánces. As a thing, different instances of error can be lumped together and counted, as in laboratory studies of human performance

[4] The personal equation (see Boring, 1950).

or as in risk analyses. Different *kinds* of errors could be ignored safely and treated as a homogenous category.

For example, in the experimental psychology laboratory, errors are counted as a basic unit of measurement for comparing performance across various factors. However, this use of error assumes that all types of errors can be combined in a homogenous category, that all specific errors can be treated as equivalent occurrences. This may be true when one has reduced a task to a minimum of content and context as is traditional in laboratory tasks. But real-world, complex tasks carried out by domain practitioners embedded in a larger temporal and organizational context are diverse. The activities and the psychological and behavioral concepts that are involved in these tasks and activities are correspondingly diverse. Hence, the resulting observable erroneous actions and assessments are diverse. In other words, in real fields of practice (where real hazards exist),

errors are heterogeneous.

One case may involve diagnosis; another may involve perceptual motor skills. One may involve system *X* and another system *Y*. One may occur during maintenance, another during operations. One may occur when there are many people interacting; another may occur when only one or a few people are present.

Noting the heterogeneity of errors was one of the fundamental contributions made by John Senders, to begin the new and intensive look at human error in 1980. An understanding of erroneous actions and assessments in the real world means that we cannot toss them into a neat causal category labeled "human error." It is fundamental to see that

erroneous actions and assessments should be taken as the ***starting point*** *for an investigation, not an ending.*

This premise is the cornerstone of the paradigm shift for understanding error (Rasmussen, 1986), and much of the material in this book should help to indicate why this premise is so fundamental.

It is common practice for investigators to see errors simply as a specific and flawed piece of human behavior within some particular task. Consider a simple example. Let us assume that practitioners repeatedly confuse two switches, *A* and *B*, and inadvertently actuate the wrong one in some circumstances. Then it seems obvious to describe the behavior as a human error where a specific person confused these two switches. This type of interpretation of errors is stuck in describing the episode in terms of the external mode of appearance or the surface manifestation (these two switches were confused), rather than also searching for descriptions in terms of deeper and more general categorizations and underlying mechanisms. For example, this confusion may be an example of a more abstract category such as a slip of action (see Norman, 1981 or Reason and Mycielska, 1982) or a mode error (see Sarter and Woods, in press, or Chapter 5).

Hollnagel (1991a; 1993) calls this the difference between the phenotype (the surface appearance) and the genotype of errors (also see the taxonomy of error taxonomies on p. 26, in Rasmussen et al., 1987). Typically, the explicit or implicit typologies of erroneous actions and assessments, such as those used in formal reporting systems, categorize errors only on the basis of phenotypes. They do not go beyond the surface characteristics and local context of the particular episode.

As early as Fitts and Jones (1947), researchers were trying to find deeper patterns that cut across the particular. The work of the 1980s has expanded greatly on the repertoire of genotypes that are related to erroneous actions and assessments. In other words, the research has been searching to expand the conceptual and theoretical basis that explains data on system breakdowns involving people. We will lay out several of these in later chapters: ones that are related to cognitive system factors that influence the formation of intentions to act, and ones that are influenced by skillful or clumsy use of computer technology. If we can learn about or discover these underlying patterns, we gain leverage on how to change human-machine systems and about how to anticipate problems prior to a disaster in particular settings.

Thus, in a great deal of the recent work on error, erroneous actions and assessments are treated as the starting point for an investigation, rather than a conclusion to an investigation (Rasmussen, 1986). The label "error" should be the starting point for investigation of the dynamic interplay of larger system and contextual factors that shaped the evolution of the incident. The attribution of human error is no longer adequate as an explanation for a poor outcome; the label "human error" is not an adequate stopping rule. It is the investigation of factors that influence the cognition and bchavior of groups of pcoplc, not the attribution of error in itself, that helps us find useful ways to change systems in order to reduce the potential for disaster and to develop higher reliability human-machine systems. In other words, it is more useful from a systcm dcsign point of view to sec that

erroneous actions and assessments are a symptom, not a cause.

There is a great diversity of notions about what "human crror" mcans. The term is problematic, in part, because it is oftcn uscd in a way that suggests that a mcaningful cause has been identified, namely thc human. To shed this causal connotation, Hollnagel (1993, p. 29) has proposed the term "crroneous action," which means an action that fails to producc the expected result and/or which produces an unwanted conscqucnce. We prefer this term for thc samc rcason.

Another contributor to thc divcrsity of intcrpretations about human error is a confusion between outcomc and process. To talk to each other about error wc must bc very clcar about whether we arc referring to bad outcomes or a defect in a process for carrying out some activity. Wc will emphasize the differencc between outcome (or performance) failures and dcfccts in thc problcm-solving proccss.

Outcomc (or pcrformance) failures are defined in terms of a categorical shift in conscquenccs on some pcrformance dimension. They are dcfincd in terms of some potentially observable standard and in terms of the language of the particular field of activity. If wc consider military aviation, somc examples of outcomc failures might include an unfulfilled mission goal, a failurc to prevent or mitigate the consequences of some system failure on the aircraft or a failure to survivc the mission. Typically, an outcomc failure (or a ncar miss) provides the impctus for an accident investigation.

Process defects are departures from some standard about how problems should be solvcd. Gcncrally, thc process dcfect, instantaneously or over time, lcads to or increases the risk of somc typc of outcome failure. Proccss defects can be defined in terms of a particular field of activity (e.g., failing to verify that all safcty systcms came on as demanded following a rcactor trip in a nuclcar power plant) or cognitively in terms of deficiencies in some cognitive or information processing function (e.g., as slips of action, Norman, 1981; fixations or cognitive lockup, De Keyser and Woods, 1990; or vagabonding, Dorner, 1983). The distinction between outcome and process is important because the relationship between them is not fixed. In other words,

there is a loose coupling between process and outcome.

This premise is implicit in Abraham Lincoln's (1864) vivid statement about process and outcome, "If the end brings me out all right what is said against me won't amount to anything. If the end brings me out wrong, ten angels swearing I was right would make no difference."

Today's students of decision making echo Lincoln, by warning us not to judge the quality of a decision by its outcome. To do so is to assume that decision makers can think of all contingencies and predict the consequences of their actions with certainty. Good decisions may be followed by bad outcomes (Fischhoff, 1982; Edwards, 1984). For example, in critical care medicine it is possible that the physician's assessments, plans, and therapeutic responses are correct for a trauma victim, and yet the patient outcome may be less than desirable—the patient's injuries may have been too severe or extensive.

Similarly, not all process defects are associated with bad outcomes. Less than expert performance may be insufficient to create a bad outcome by itself; the operation of other factors may be required as well. This may be, in part, the result of successful engineering (such as defenses in depth) and multiple opportunities for detection and recovery may occur as the incident evolves. Thus, the label "error" alone is ambiguous, in part, because it is not clear whether it refers to outcome or process.

The loose coupling of process and outcome occurs because incidents evolve along a course that is not preset. Further along there may be opportunities to direct the evolution towards successful outcomes, or other events or actions may occur that direct the incident towards negative consequences.

Consider a pilot who makes a mode error which, if nothing is done about it, would lead to disaster within some minutes. It may happen that the pilot notices certain unexpected indications and responds to the situation, thereby diverting the incident evolution back onto a benign course. The fact that process defects do not always, or even frequently, lead to bad outcomes makes it very difficult for people or organizations to understand the nature of error, its detection, and recovery.

As a result of the loose coupling between process and outcome, we are left with a nagging problem. Defining human error as a form of process defect implies that there exists some criterion or standard against which the performance has been measured and deemed inadequate. However, what standard should be used? We do not think that there will be a single and simple answer to this question. However, if we are ambiguous about the particular standard adopted to define error in particular studies or incidents, then we greatly retard our ability to engage in a constructive and empirically grounded debate about error. All claims about when an action or assessment is erroneous in a process sense must be accompanied by an explicit statement of the standard used for defining departures from good process.

One kind of standard that can be invoked is a normative model of task performance. For many fields of activity where bad outcomes can mean dire consequences, there are no normative models or there are great questions surrounding how to transfer normative models developed for much simpler situations to a more complex field of activity. For example, laboratory-based normative models may ignore the role of time or may assume that cognitive processing is resource-unlimited.

Another possible kind of standard is standard operating practices (e.g. written policies and procedures). However, work analysis has shown that formal practices and policies often depart substantially from the dilemmas, constraints, and tradeoffs present in the actual workplace (e.g., Hirschhorn, 1993).

For realistically complex problems there is often no one best method; rather, there is an envelope containing multiple paths each of which can lead to a satisfactory outcome. This suggests the possibility of a third approach for a standard of comparison. One could use an empirical standard that asks: What would other similar practitioners have thought or done in this situation? De Keyser and Woods (1990) called these empirically based comparisons neutral observer criteria. A simple example occurred in regard to the Strasbourg aircraft crash (Monnier, 1992). Mode error in pilot interaction with cockpit automation seems to have been a contributor to this accident. Following the accident, several people in the aviation industry noted a few precursor incidents or dress rehearsals for the crash where similar mode errors had occurred, although the incidents did not evolve as far towards negative consequences. (At least one of these mode errors resulted in an unexpected rapid descent, and the ground proximity warning system alarm alerted the crew, who then executed a go-around). Issues about standards used to define process defects, especially neutral observer criteria, will be explored more in Chapter 6.

Whatever kind of standard is adopted for a particular study,

knowledge of outcome (hindsight) biases judgments about process.

People have a tendency to judge the quality of a process by its outcome. The information about outcome biases their evaluation of the process that was followed (Baron and Hershey, 1988). The loose coupling between process and outcome makes it problematic to use outcome information as an indicator for error in a process. (Chapter 6 explains the outcome bias and related hindsight bias and discusses their implications for the study of error.)

Studies of disasters have revealed an important common characteristic:

incidents evolve through the conjunction of several failures/factors.

Actual accidents develop or evolve through a conjunction of several small failures, both machine and human (Pew et al., 1981; Perrow, 1984; Wagenaar and Groeneweg, 1987; Reason, 1990). This pattern is seen in virtually all of the significant nuclear power plant incidents, including Three Mile Island, Chernobyl, the Brown's Ferry fire, the incidents examined in Pew et al. (1981), the steam generator tube rupture at the Ginna station (Woods, 1982), and others. In the near miss at the Davis-Besse nuclear station (U.S. NRC, NUREG-1154, 1985), there were about ten machine failures and several erroneous actions that initiated the loss-of-feedwater accident and determined how it evolved.

In the evolution of an incident, there are a series of interactions between the human-machine system and the hazardous process. One acts and the other responds, which, in turn, generates a response from the first and so forth. Incident evolution points out that there is some initiating event in some human and technical system context, but there is no one clearly identifiable cause of the accident (Rasmussen, 1986; Senders and Moray, 1991). However, after the fact, several points during the accident evolution can be identified where the evolution can be stopped or redirected away from undesirable outcomes.

Gaba, Maxwell, and DeAnda (1987) applied this idea to critical incidents in anesthesia, and Cook, Woods, and McDonald (1991), also working in anesthesia, identified several different patterns of incident evolution. For example, "acute" incidents present themselves all at once, while in "going sour" incidents, there is a slow degradation of the monitored process.

One kind of "going sour" incident, which they called decompensation incidents, occurs when an automatic system's responses mask the diagnostic signature produced by a fault (cf. Woods, in press-a). As the abnormal influences produced by a fault persist or grow over time, the capacity of automatic systems to counterbalance or compensate becomes exhausted. At some point they fail to counteract and the system collapses or decompensates. The result is a two-phase signature. In phase 1 there is a gradual falling off from desired states over a period of time. Eventually, if the practitioner does not intervene in appropriate and timely ways, phase 2 occurs—a relatively rapid collapse when the capacity of the automatic systems is exceeded or exhausted. During the first phase of a decompensation incident, the gradual nature of the symptoms can make it difficult to distinguish a major challenge, partially compensated for, from a minor disturbance (see National Transportation Safety Board, 1986a). This can lead to a great surprise when the second phase occurs (e.g., some practitioners who miss the signs associated with the first phase may think that the event began with the collapse; cf. Cook, Woods, and McDonald, 1991). The critical difference between a major challenge and a minor disruption is not the symptoms, per se, but rather the force with which they must be resisted. This case illustrates how incidents evolve as a function of the interaction between the nature of the trouble itself and the responses taken to compensate for that trouble.

Some of the contributing factors to incidents are ***latent*** *in the system.*

Some of the factors that combine to produce a disaster are latent in the sense that they were present before the incident began. Turner (1978) diseusses the incubation of factors prior to the ineident itself, and Reason (1990) refers to potential destruetive forces that build up in a system in an explicit analogy to resident pathogens in the body. Thus, latent failures refer to problems in a system that produee a negative effeet but whose eonsequenees are not revealed or activated until some other enabling eondition is met. Examples include failures that make safety systems unable to funetion properly if ealled on, such as the error during maintenanee that resulted in the emergency feedwater system being unavailable during the Three Mile Island incident (The Kemeny Commission, 1979). Latent failures require a trigger, i.e., an initiating or enabling event, that aetivates its effects or consequences. For example in the Spaee Shuttle Challenger disaster, the deeision to launch in cold weather was the initiating event that aetivated the eonsequenees of the latent failure—a highly vulnerable booster roeket seal design. This generalization means that assessment of the potential for disaster should include a seareh for evidenee about latent failures hidden in the system (Reason, 1990).

When error is seen as the starting point for study, when the heterogeneity of errors (their external mode of appearanee) is appreeiated, and the difference between outeome and proeess is kept in mind, then it beeomes clear that one cannot separate the study of error from the study of normal human behavior. We quickly find that we are not studying error, but rather, human behavior itself, embedded in meaningful eontexts. As Rasmussen (1985) states:

> It. . . [is] important to realize that the seientifie basis for human reliability eonsiderations will not be the study of human error as a separate topic, but the study of normal human behavior in real work situations and the meehanisms involved in adaptation and learnin (p. 1194).

The point is that

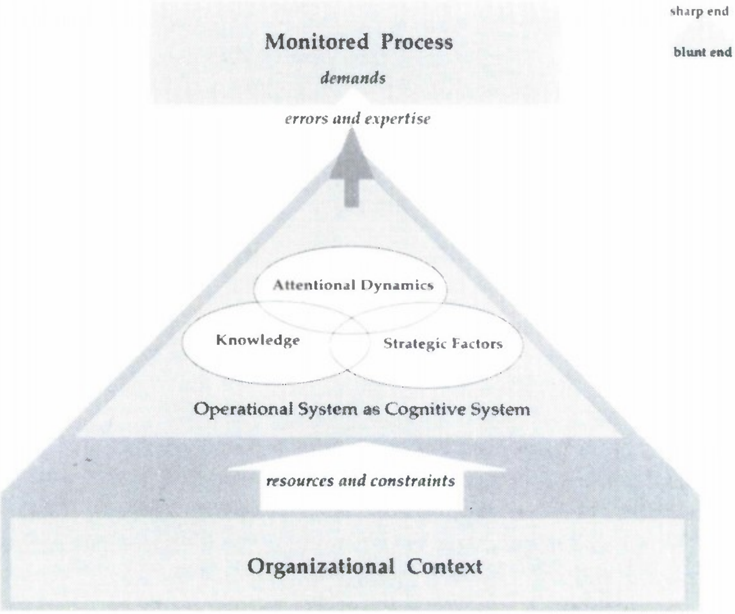
the same factors govern the expression of expertise and of error.

Jens Rasmussen frequently quotes Ernst Mach (1905, p. 84) to reinforce this point: "Knowledge and error flow from the same mental source; only success can tell one from the other."

Furthermore, to study error in real-world situations necessitates studying groups of individuals embedded in a larger system that provides resources and constraints, rather than simply studying private, individual cognition. To study error is to study the function of the system in which practitioners are embedded. Chapter 4 covers a variety of cognitive system factors that govern the expression of error and expertise. It also explores some of the demand factors in complex domains and the organizational constraints that also play an important role in the expression of error and expertise (see Figure 1, p. 21).

Generally, the human referred to when an incident is ascribed to human error is some individual or team of practitioners who work at what James Reason calls the "sharp end" of the system (Reason, 1990; see Figure 1, p. 21). Practitioners at the sharp end actually interact with the hazardous process in their roles as pilots, physicians, space controllers, or power plant operators. In medicine, these practitioners are anesthesiologists, surgeons, nurses, and some technicians who are physically and temporally close to the patient. Those at the "blunt end" of the system, to continue Reason's analogy, affect safety through their effect on the constraints and resources acting on the practitioners at the sharp end. The blunt end includes the managers, system architects, designers, and suppliers of technology. In medicine the blunt end includes government regulators, hospital administrators, nursing managers, and insurance companies. To understand the sources of expertise and error at the sharp end, one must also examine this larger system to see how resources and constraints at the blunt end shape the cognition and behavior of sharp end practitioners (Reason, 1990).

Note that there is a theme that underlies all of the above points about the study of error:



©1993 Woods, Johannesen

Figure 1. The sharp and blunt ends of a large complex system. The interplay ớf problem demands and the resources of practitioners at the sharp end govern the expression of expertise and error. The resources available to meet problem demands are shaped and constrained in large part by the organizational context at the blunt end of the system.

lawful factors govern the types of erroneous actions or assessments to be expected.

Errors are not some mysterious product of the fallibility or unpredictability of people; rather errors are regular and predictable consequences of a variety of factors. In some cases we understand a great deal about the factors involved, while in others we currently know very little. This premise is not only useful in improving a particular system, but also assists in defining general patterns that cut across particular circumstances. Finding these regularities requires examination of the contextual factors surrounding the specific behavior that is judged faulty or erroneous. In other words,

erroneous actions and assessments are context-conditioned.

Many kinds of contextual factors are important to human cognition and behavior (see Figure 1, p. 21). The demands imposed by the kinds of problems that can occur are one such factor. The constraints and resources imposed by organizational factors are another. The temporal context defined by how an incident evolves is yet another (e.g., from a practitioner's perspective, a small leak that gradually grows into a break is very different from an incident where the break occurs quite quickly). Chapter 4 discusses these and many other cognitive factors that affect the expression of expertise and error.

Variability in behavior and performance turns out to be crucial for learning and adaptation. In some domains, such as control theory, an error signal, as a difference from a target, is informative because it provides feedback about goal achievement and indicates when adjustments should be made. Error, as part of a continuing feedback and improvement process, is information to shape future behavior. However, in certain contexts this variability can have negative consequences. As Rasmussen (1986) puts it, in "unkind work environments" variability becomes an "unsuccessful experiment with unacceptable consequences." This view emphasizes the following important notion:

error tolerance, error detection, and error recovery are as important as error prevention.

Again, according to Rasmussen (1985),

> . . .The ultimate error frequency largely depends upon the features of the work interface which support immediate error recovery, which in turn depends on the observability and reversibility of the emerging unacceptable effects. The feature of reversibility largely depends upon the dynamics and linearity of the system properties, whereas observability depends on the properties of the task interface which will be dramatically influenced by the modern information technology. (p. 1188)

Figure 2 (p. 24) illustrates the relationship between recovery from error and the negative consequences of error (outcome failures). An erroneous action or assessment occurs in some hypothetical system. It is followed by a recovery interval, i.e., a period of time during which actions can be taken to reverse the effects of the erroneous action or during which no consequences result from the erroneous assessment. If error detection occurs, the assessment is updated or the previous actions are corrected or compensated for before any negative consequences accrue. If not, then an outcome failure has occurred. There may be further recovery intervals during which other outcome consequences (of a more severe nature) may be avoided if detection and recovery actions occur.

A field of activity is tolerant of erroneous actions and assessments to the degree that such errors do not immediately or irreversibly lead to negative consequences. An error-tolerant system has a relatively long recovery interval, i.e., there are extensive opportunities for reversibility of actions. Error recovery depends on the *observability* of the monitored process which is in large part a property of the human-computer interface for computerized systems. For example, is it easy to see if there is a mismatch between expected state and the actual state of the system? Several studies show that many human-computer interfaces provide limited observability, i.e., they do not provide effective visualization of events, change and anomalies in the monitored process (e.g., Moll van Charante, Cook, Woods, Yue, and Howie, 1993 for automated operating room devices; Woods, Potter, Johannesen, and Holloway, 1991 for intelligent systems for fault management of space vehicle

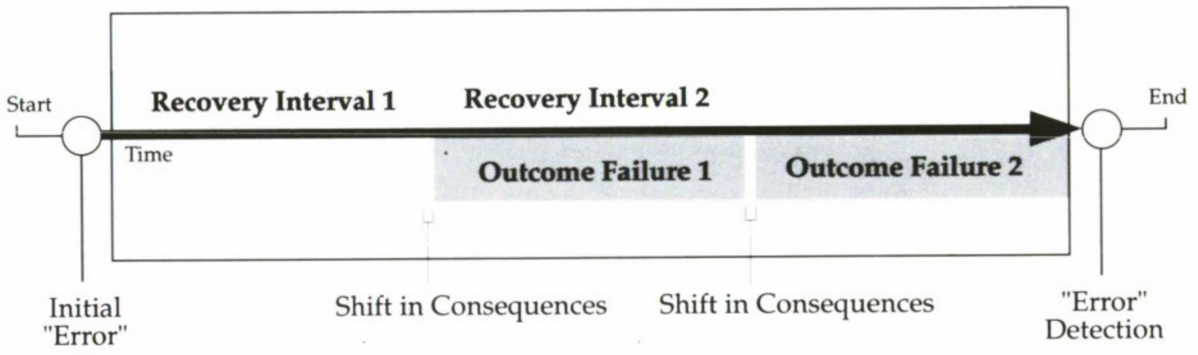


Figure 2. The relationship between error recovery and outcome failure. Outcome failures of various types (usually of increasing severity) may be averted if recovery occurs within a particular time span, the length of which depends on the system characteristics.

systems; Sarter and Woods, 1994 for cockpit automation). The opaque nature of the interfaces associated with new technology is particularly troubling because it degrades error recovery. Moll van Charante, et al. (1993) and Cook, Woods, and Howie (1992) contain data directly linking low observability through the computer interface to critical incidents in the case of one automated operating room device. Sarter and Woods (in press) link low observability through the interface to problems in mode awareness for cockpit automation (cf. also, the Therac-25 accidents, in which a radiation therapy machine delivered massive doses of radiation, for another example where low observability through the computer interface to an automatic system blocked error or failure detection and recovery; Leveson and Turner, 1992). Chapter 5 discusses these issues in greater depth.

While design to minimize or prevent erroneous actions is good practice, one cannot eliminate the possibility for error. It seems that the path to high-reliability systems critically depends on design to enhance error recovery prior to negative consequences (Lewis and Norman, 1986; Rasmussen, 1986; Reason, 1990). Rasmussen (1985) points out that reported frequencies of "human error" in incident reports are actually counts of errors that were not detected and recovered from, prior to some negative consequence or some criterion for cataloging incidents. Opportunities for the detection and correction of error, and hence tools that support people in doing so, are critical influences on how incidents will evolve (see Seifert and Hutchins, 1992 for just one example).

Enhancing error tolerance and error recovery is a common prescription for designing systems (e.g., Norman, 1988). Some methods include:

- design to prevent an erroneous action, e.g., forcing functions which constrain a sequence of user actions along particular paths.
- design to increase the tolerance of the underlying process to erroneous actions, and
- design to enhance recovery from errors and failures through effective feedback and visualizations of system function—enhanced observability of the monitored process (e.g., Potter, Woods, Hill, Boyer, and Morris, 1992; Yue, Woods, and Cook, 1992; Woods, in press-b).

Let us pause and summarize a few important points: failures involve *multiple* contributing factors. The label "error" is often used in a way that simply restates the fact that the outcome was undesirable. Error is a symptom indicating the need to investigate the larger operational system and the organizational context in which it functions. In other words,

systems fail.

If we examine actual accidents, we will typically find that several groups of people were involved. For example, in the Dallas windshear aircraft crash (National Transportation Safety Board, 1986b), the incident evolution involved the crew of the aircraft in question, what other planes were doing, air traffic controllers, the weather service, company dispatch, company and ·industry pressures about schedule delays.

Failures involve multiple groups and people, even at the sharp end. One also finds in complex domains that error detection and recovery are inherently distributed over multiple people and groups and over human and machine agents. This is the case in aircraft carrier flight operations (Roehlin, La Porte, and Roberts, 1987), maritime navigation (Hutchins, 1990; in press), power plant startup (Roth and Woods, 1988) and many others. Woods et al. (1987) synthesized results across several studies of simulated and actual nuclear power plant emergencies and found that detection and correction of erroneous state assessments came primarily from other crew members who brought a fresh point of view into the situation. Miscommunications between air traffic control and commercial airline flight decks occur frequently, but the air transport system has evolved robust cross-people mechanisms to detect and recover from communication breakdowns, e.g., crew cross-checks and read backs, although miscommunications still can play a role in accidents (National Transportation Safety Board, 1991). Systems for cross-checking occur in pilots' coordination with cockpit automation. For example, pilots develop and are taught cross-check strategies to detect and correct errors that might occur in giving instructions to the flight computers and automation. There is evidence, though, that the current systems are only partially successful and that there is great need to improve the coordination between people and automated agents in error or failure detection (e.g., Sarter and Woods, in press).

Systems are always made up of people in various roles and relationships. The systems exist for human purposes. So when systems fail, of course human failure can be found in the rubble. But progress towards safety can be made by understanding the system of people and the resources that they have evolved and their adaptations to the demands of the environment. Thus, when we start at "human error" and begin to investigate the factors that lead to behavior that is so labeled, we quickly progress to studying systems of people embedded in a larger organizational context (Reason, 1990). In this book we will tend to focus on the sharp-end system, i.e., the set of practitioners operating near the process and hazards, the demands they confront, and the resources and constraints imposed by organizational factors (see Chapter 4).

The perception that there is a "human error problem" is one force that leads to computerization and increased automation in operational systems. As new information and automation technology is introduced into a field of practice what happens to "human error"? The way in which technological possibilities are used in a field of practice affects the potential for different kinds of erroneous actions and assessments. It can reduce the chances for some kinds of erroneous actions or assessments, but it may create or increase the potential for others. In other words,

the design of artifacts affects the potential for erroneous actions and paths towards disaster.

Artifacts are simply human-made objects. In this context we are interested particularly in computer-based artifacts from individual microprocessor-based devices such as infusion pumps for use in medicine to the suite of automated systems and associated human-computer interfaces present in advanced cockpits on commercial jets. One goal for this book is to focus on the role of design of computer-based artifacts in human error.

Properties of specific computer-based devices or aspects of more general "vectors" of technology change influence the cognition and activities of those people who use them. As a result, technology change can have profound repercussions on system operation, particularly in terms of the types of "errors" that occur and the potential for failure. It is important to understand how technology change shapes human cognition and action in order to see how design can create latent failures which may contribute, given the presence of other factors, to disaster. For example, a particular technology change may increase the coupling in a system (Perrow, 1984). Increased coupling increases the cognitive demands on practitioners. If the computer-based artifacts used by practitioners exhibit "classic" flaws such as weak feedback about system state (what we will term low observability), the combination can function as a latent failure awaiting the right circumstances and triggering events to lead the system close to disaster (see Moll van Charante et al., 1993 for one example of just this sequence of events).

One particular type of technology change, namely increased automation, is assumed by many to be the prescription of choice to cure an organization's "human error problem."[5] If incidents are the result of "human error," then it seems justified to respond by retreating further into the philosophy that "just a little more technology will be enough" (Woods, 1990b; Billings, 1991). Such a technology-centered approach is more likely to increase the machine's role in the cognitive system in ways that will squeeze the human's role (creating a vicious cycle as evidence of system problems will pop up as more human error; Cook and Woods, in press). As S. S. Stevens noted (1946, p. 390):

> . . . the faster the engineers and the inventors served up their 'automatic' gadgets to eliminate the human factor the tighter the squeeze became on the powers of the operator . . . .

And as Norbert Wiener noted some years later (1964, p. 63):

> The gadget-minded people often have the illusion that a highly automatized world will make smaller claims on human ingenuity than does the present one . . . . This is palpably false.

[5] One recent example of this attitude comes from a commentary about cockpit developments envisioned for a new military aircraft in Europe: "The sensing, processing and presentation of such unprecedented quantities of data to inform and protect one man requires new levels of . . . system integration. When proved in military service, these automation advances will read directly across to civil aerospace safety. They will also assist the industrial and transport communities' efforts to eliminate 'man-machine interface' disasters like King's Cross, Herald of Free Enterprise, Clapham Junction and Chernobyl." *Aerospace*, November, 1992, p. 10.

Failures to understand the reverberations of technological change on the operational system hinder the understanding of important issues such as what makes problems difficult, how breakdowns occur, and why experts perform well.

Our strategy is to focus on how technology change can increase or decrease the potential for different types of erroneous actions and assessments. In Chapter 5 we will lay out a broad framework that establishes three inter-related linkages: the effect of technology on the cognitive activities of practitioners; how this, in turn, is linked to the potential for erroneous actions and assessments; and how these can contribute to the potential for disaster.

The concept that the design of the human-machine system, defined very broadly, affects or "modulates" the potential for erroneous actions and assessments, was present at the origins of Human Factors when the presence of repeated "human errors" was treated as a signal pointing to context-specific flaws in the design of human-machine systems (e.g., cockpit control layout). This idea has been reinforced more recently when researchers have identified *kinds* of design problems in computer-based systems that cut across specific contexts. In general, "clumsy" use of technological powers can create additional mental burdens or other constraints on human cognition and behavior that create opportunities for erroneous actions and assessments by people, especially in high-criticality, high-workload, high-tempo operations (Wiener, 1989; Sarter and Woods, in press).

Computer-based devices, as typically designed, tend to exhibit classic human-computer cooperation flaws such as lack of feedback on device state and behavior (e.g., Norman, 1990b; Woods, Cook, and Sarter, 1992). Furthermore, these HCI flaws increase the potential for erroneous actions and for erroneous assessments of device state and behavior. The low observability supported by these interfaces and the associated potential for erroneous state assessment is especially troublesome because it impairs the user's ability to detect and recover from failures, repair communication breakdowns, and detect erroneous actions.

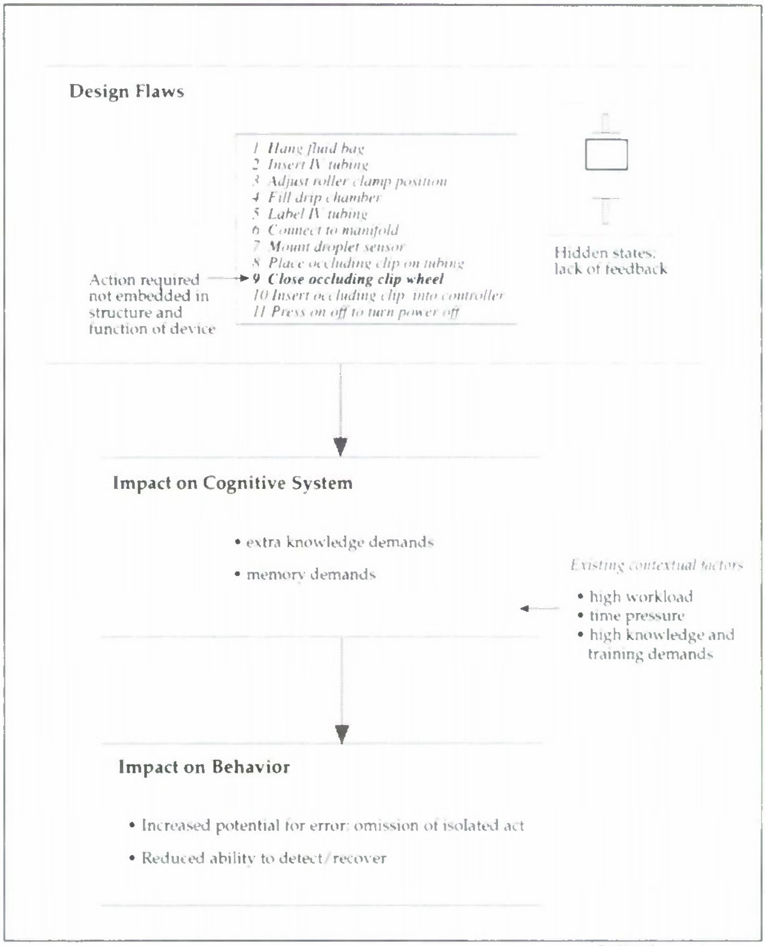
These data, along with critical incident studies, directly implicate the increased potential for erroneous actions and the decreased ability to detect errors and failures as one kind of important contributor to actual incidents. The increased potential for error that emanates from poor human-computer cooperation is one type of latent failure that can be activated and progress towards disaster given the presence of other potential factors.

Our goals are to expose various design "errors" in human-computer systems that create latent failures, show how devices with these characteristics shape practitioner cognition and behavior, and how these characteristics can create new possibilities for error and new paths to disaster. In addition, we will examine data on how practitioners cope with the complexities introduced by the clumsy use of technological possibilities and how this adaptation process can obscure the role of design and cognitive system factors in incident evolution (Woods et al., 1992; Cook and Woods, 1994). This information should help developers detect, anticipate, and recover from designer errors in the development of computerized devices.

An Example

Figures 3 and 4 (pp. 31 and 33) summarize an example of error as a predictable consequence of task and other factors (taken from Yue et al., 1992; Moll van Charante et al., 1993). The setup of a new microprocessor automated controller for use in the operating room involves a series of steps. Physical and functional relationships that are apparent in the device components themselves provide some constraints so that the steps are performed successfully. However, for one step of the formal procedure specified in the manual, the action required is not related to the structure or function of the device in any sense that a user can see. It stands out as an isolated act from the rest of the sequence. Observations of device setup in context revealed that this step was frequently omitted. This omission is erroneous relative to the standard of the formal procedure for device setup (the step in question is specified on a single page of a 40-page device manual).

This omission is not particularly surprising—the physicians had to know that this was even a formal step in the procedure (some did not), and they had to remember this step (since there were no cues in the device or the sequence of activities to act as a reminders). The work context is one of very high workload with many demands including


Design Flaws
1 Hang fluid bag
2 Insert IV tubing
3 Adjust roller clamp position
4 Fill drip chamber
5 Label IV tubing
6 Connect to manifold
7 Mount droplet sensor
8 Place occluding clip on tubing
9 Close occluding clip wheel
10 Insert occluding clip into controller
11 Press on off to turn power off
Action required not embedded in structure and function of device
Hidden states: lack of feedback
Impact on Cognitive System
• extra knowledge demands
• memory demands
Existing contextual factors
• high workload
• time pressure
• high knowledge and training demands
Impact on Behavior
• Increased potential for error: omission of isolated act
• Reduced ability to detect/recover

© 1992 Woods, Johannesen

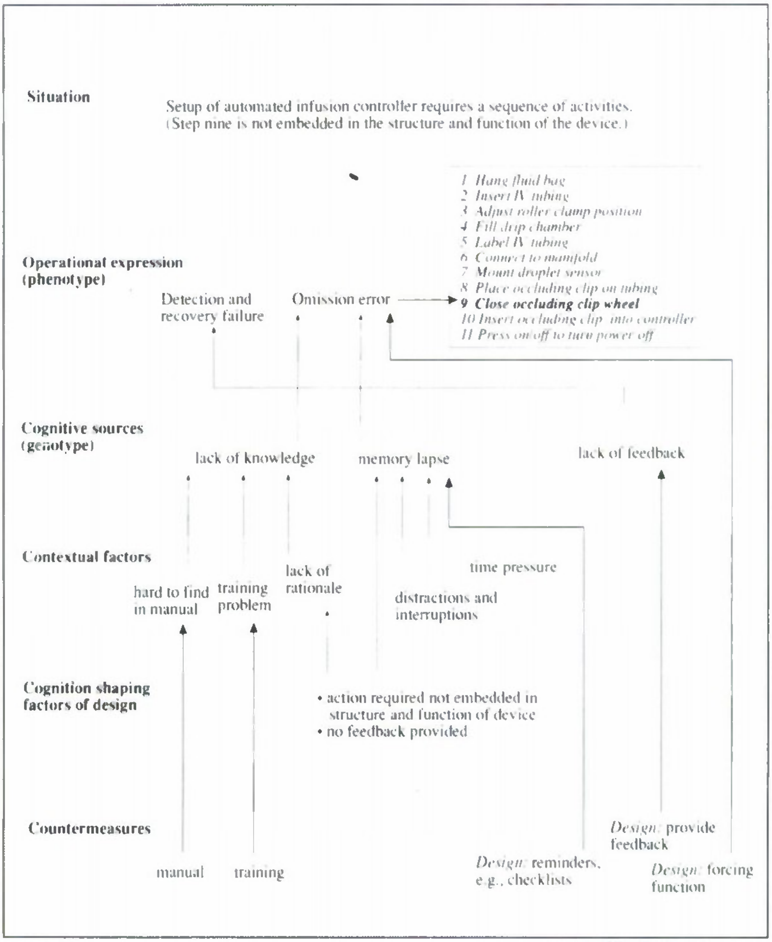
Figure 3. An example of how design flaws (in an operating room automated infusion controller) impact the cognitive system, which, in turn, impacts behavior.

time pressure. This device is just one of dozens that need to be set up prior to cardiac surgery and, interestingly enough, it is one that is supposed to off-load the practitioner. Furthermore, if the step is omitted there is no visible feedback about whether the device is assembled correctly (to check this step and for most other potential misassemblies one would have to disconnect the entire assembly). Even after an incident occurred where the omission of this step played a role in a device failure (the critical event was probably a software bug), observation showed that it was still quite easy for practitioners to forget this step (Cook et al., 1992).

The scenario is a classic case of an isolated act, which is likely to lead to an error, namely an omission. The omission is the external manifestation of the error, in other words, the *phenotype* (Hollnagel, 1991-a). Figure 4 (p. 33) charts how we can go further by exploring the *genotype* (underlying cognitive mechanisms) of the erroneous action. We would get at this by asking questions that reveal the knowledge, memory and other cognitive demands faced by practitioners *in situ*. Do the practitioners have the relevant knowledge? Given the design of the manual and other contextual factors (little formal training on each device in this environment; many different devices to be set up and operated), the relevant knowledge may not be there to be activated when needed. The design also creates new memory demands. Given the context (i.e., high workload and high likelihood of distractions and interruptions) and the absence of any external memory cues or aids, it is easy for a memory lapse to occur. Finally, what about error detection and recovery? Lack of feedback on the state of the device just about eliminates any possibility of detecting a problem prior to device use (ironically, the device's purpose is to help offload the physician at the highest workload and most critical period of cardiac surgery).

Figure 4 (p. 33) also charts the various countermeasures that could be brought to bear. For example, the knowledge problem could possibly be handled by redesigning the manual or the training. The memory problem could be attacked through external memory aids.

The feedback problem could be attacked by providing information about device state and redesigning the device to eliminate the need to remember or to perform this isolated step. In the latter case, a forcing function could be used. (Just how to do any one of these strategies


Situation
Setup of automated infusion controller requires a sequence of activities.
(Step nine is not embedded in the structure and function of the device.)
1 Hang fluid bag
2 Insert IV tubing
3 Adjust roller clamp position
4 Fill drip chamber
5 Label IV tubing
6 Connect to manifold
7 Mount droplet sensor
8 Place occluding clip on tubing
9 Close occluding clip wheel
10 Insert occluding clip into controller
11 Press on/off to turn power off
Operational expression (phenotype)
Detection and recovery failure
Omission error
Cognitive sources (genotype)
lack of knowledge
memory lapse
lack of feedback
Contextual factors
hard to find in manual
training problem
lack of rationale
time pressure
distractions and interruptions
Cognition shaping factors of design
• action required not embedded in structure and function of device
• no feedback provided
Countermeasures
manual
training
Design: reminders, e.g., checklists
Design: provide feedback
Design: forcing function

© 1992 Woods, Johannesen

Figure 4. Underlying cognitive factors behind an omission error (occurring in the setup of an operating room automated controller) and possible countermeasures.

effectively and in detail is another matter.)

Note how in this case, the discussion is shifted away from the external appearance of the error (its phenotype) and towards typologies that express regularities about task or psychological or human-machine system factors that shape the possibilities for erroneous actions or assessments (its genotype). The research community's knowledge of these regularities or types of error forms is limited (though it is far from an empty set), and our ability to predict the timing and statistical properties of error distributions is very limited. However, we can make predictions about the forms errors will take when they do occur.

3

COMPLEX SYSTEM BREAKDOWN: THE LATENT FAILURE MODEL

The Anatomy of Disaster

To study accidents, it is important to understand the dynamics and evolution of the conditions that give rise to system breakdowns. Various stakeholders often imagine that the typical path to disaster is a single and major failure of a system component, either a machine or a human component. Studies of the anatomy of disasters in highly technological systems, however, show a different pattern—one that James Reason has called the latent failure model of complex system breakdown (Reason, 1990, chapter 7).

Highly technological systems such as aviation, air traffic control, telecommunications, nuclear power, space missions, and medicine include potentially disastrous failure modes. Significantly, these systems usually have multiple redundant mechanisms, safety systems, and elaborate policies and procedures to keep them from failing in ways that produce bad outcomes. The results of combined operational and engineering measures make these systems relatively safe from single point failures; that is, they are protected against the failure of a single component or procedure directly leading to a bad outcome.

The need to make these systems reliable in large part also makes them very complex. They are large systems, semantically complex (it generally takes a great deal of time to master the relevant domain knowledge), with tight couplings between various parts, and operations are often carried out under time pressure or other resource constraints. The scale and coupling of these systems create a different pattern for disaster where incidents develop or *evolve* through a *conjunction* of several small failures, hoth machine and human (e.g., Turner, 1978; Pew et al., 1981; Perrow, 1984; Wagenaar and Groeneweg, 1987; Reason, 1990). This pattern can be seen in disasters or events in a variety of different industries, and despite the fact that each critical incident is unique in many respects.

These incidents evolve through a series of interactions hetween the people responsible for system integrity and the behavior of the technical systems themselves (the engineered or physiological processes under control). One acts, the other responds, which generates a response from the first and so forth. The incident evolution can be stopped or redirected away from undesirable outcomes at various points.

Incidents that evolve to–or near to–disaster seem to share several common characteristics.

1. Disasters are characterized hy a concatenation of several small failures and contributing events rather than a single large failure (e.g., Pew et al., 1981; Reason, 1990). The multiple contributors are all necessary hut individually insufficient for the disaster to have occurred. If any of the contrihuting factors were missing, the disaster would have heen avoided. Similarly, a contributing failure can occur without producing negative outcomes if other potential factors are not present.

For example, the comhination of multiple contributing events is seen in virtually all of the significant nuclear power plant incidents, including Three Mile Island, Chernohyl, the Browns Ferry fire, the incidents examined in Pew et al. (1981), the steam generator tube rupture at the Ginna station (Woods, 1982) and others. In the near miss at the Davis-Besse nuclear station (NUREG-1154), there were about ten machine failures and several erroneous human actions that initiated the loss-of-feedwater accident and determined how it evolved.

2. Some of the factors that combine to produce a disaster are latent in the sense that they were present before the incident began. Turner (1978) discusses the incuhation of factors prior to the incident itself, and Reason (1990) refers to hidden pathogens that huild in a system in an explicit analogy to viral processes in medicine.

Reason (1990) uses the term *latent failures* to refer to conditions resident in a system that can produce a negative effect but whose consequences are not revealed or activated until some other enabling condition is met. These conditions are latent or hidden because their consequences are not manifest until the enabling conditions occur. A typical example is a condition that makes safety systems unable to function properly if called on, such as the maintenance problem that resulted in the emergency feedwater system being unavailable during the Three Mile Island incident (The Kemeny Commission, 1979). Latent failures require a trigger, i.e., an initiating or enabling event, that activates its effects or consequences. For example in the Space Shuttle Challenger disaster, the decision to launch in cold weather was the initiating event that activated the consequences of the latent failure in booster seal design (Rogers et al., 1986).[6]

3. The concatenation of factors in past disasters includes both human and machine elements intertwined as part of the multiple factors that contribute elements, but only as part of the dynamics of a human-machine operational system that has adapted to the demands of the field of activity and to the resources and constraints provided by the larger organizational context (Rasmussen, 1986; see Figure 1, p. 21, for a graphic rendering of this point).

Reason's Latent Failure Model

Reason's (1990) latent failure model distinguishes between active and latent failures. Active failures are "unsafe acts" whose negative consequences are immediately or almost immediately apparent. These are associated with the people at the "sharp end," that is, the operational personnel who directly see and influence the process in question. Latent failures are decisions "whose adverse consequences may lie dormant within the system for a long time, only becoming evident when they combine with other factors to breach the system's defenses" (Reason, 1990). Some of the factors that serve as "triggers" may be active failures, technical faults, or atypical system states. Latent fail-

[6] Strictly speaking, latent failures are not outcome failures but conditions that can lead to outcome failures. We will use the label "latent failures" because that is the term originally employed by Reason.

ures are associated with managers, designers, maintainers, or regulators—people who are generally far removed in time and space from handling incidents and accidents.

According to Reason (1990), one should think of accident potential in terms of organizational processes, task and environmental conditions, individual unsafe acts, and failed defenses (see Figure 5, p. 39, a slight adaptation of a figure from Reason, 1990). The organizational plane involves such processes as goal setting, organizing, communicating, managing, designing, building, operating, and maintaining. The latent failures that occur here are fallible decisions, which can result in incompatible goals, organizational deficiencies, inadequate communications, poor planning and scheduling, inadequate control and monitoring, design failures, unsuitable materials, poor procedures (both in operations and maintenance), deficient training, and inadequate maintenance management (see Reason, 1993, p. 230-1).

Chapter 4, particularly the sections on Strategic Factors, shows how blunt-end factors can shape practitioner cognition and create the potential for erroneous actions and assessments. Chapter 5 shows how the clumsy use of technology is one type of latent failure. This type of latent failure arises in the design organization. It predictably leads to certain kinds of unsafe acts on the part of practitioners at the sharp end and contributes to the evolution of incidents towards disaster. Task and environmental conditions are typically thought of as "performance-shaping factors." The unsafe acts are the active failures; according to Reason these consist of both errors and violations.

Violations are deviations from some code of practice or procedure (but see the sections on Practitioner Tailoring and on Rule Following in Chapter 5, to see how violations are but one point on a dimension of adaptation). Defenses are measures that protect against hazards or lessen the consequences of malfunctions or erroneous actions. Some examples include safety systems or forcing functions such as interlocks. According to Reason (1990), the "best chance of minimizing accidents is by identifying and correcting these delayed action failures [latent failures] before they combine with local triggers to breach or circumvent the system's defenses."

The latent failure model broadens the story of error. It is not enough to stop with the attribution that some individual at the sharp end erred.

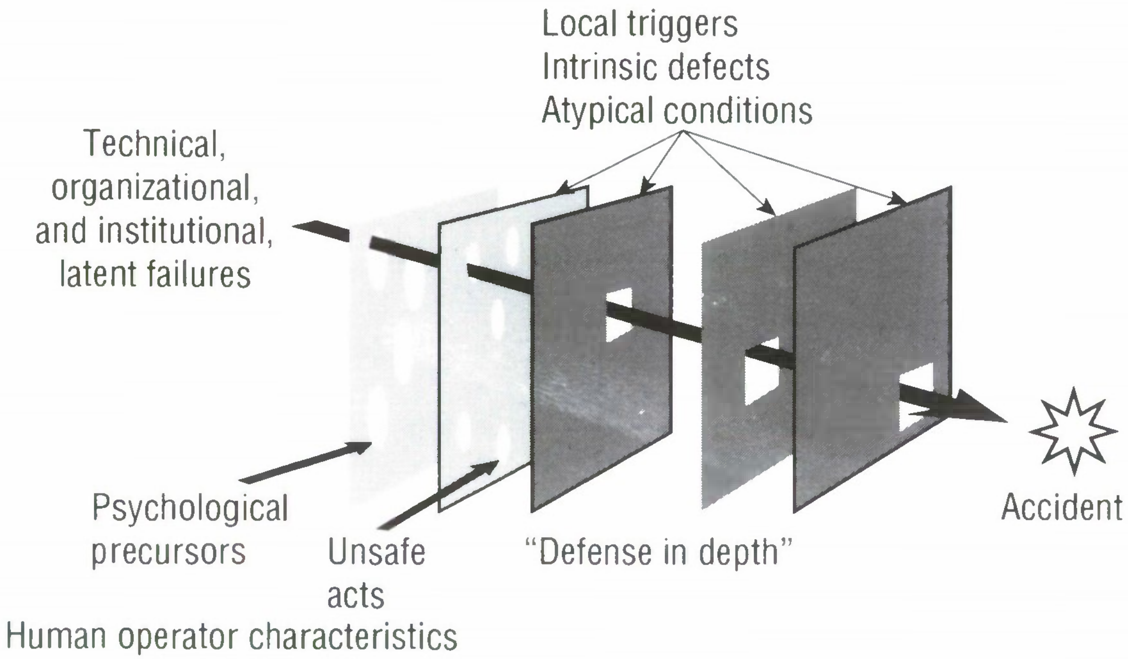


Figure 5. Reason's Latent Failure Model (slightly modified from Reason, J. [1990]. *Human error*. Cambridge, England: Cambridge University Press. Reprinted with the permission of Cambridge University Press).

The concept of latent failures highlights the importance of organizational factors. It shows how practitioners at the sharp end can be constrained or trapped by larger factors.

The latent failure model has profound implications for the story of error. This concept will be referred to frequently throughout the book. If the reader wishes to understand morc about the latcnt failure model, see Reason (1990) especially Chapter 7. To illustrate the concepts we will describe a concrete example of how latent failures can contribute to incident evolution in the next section.

An Example of Latent Failures: The Missing O-Rings in the Eastern L1011 Incident

The following case illustrates how multiple factors come together to result in accidents. The case is that of an Eastern L1011 flying from Miami to Nassau in May of 1983. Thc aircraft lost oil pressure in all three of its engines in mid-flight. Two of the engines stopped, and the third gave out at about the time the crew safely landed the aircraft. The proximal event was that O-rings, which normally should be attached to an engine part, were missing from all three engines.[7] A synopsis of relevant events leading up to the incident is given below, based on the National Transportation Safety Board report (NTSB, 1984) and on Norman's commentary on this incident (Norman, 1992).

One of the tasks of mechanics is to replace an engine part, called a master chip detector, at scheduled intervals. The master chip detector fits into the engine and is used to detect engine wear. O-rings are used to prevent oil leakage when the part is inserted. The two mechanics for the flight in question had always gotten replacement master chip detectors from their foreman's cabinct. Thesc chip detectors were all ready to go, with new O-rings installed. The mechanics' work cards specified that new O-rings should be installed with a space next to this instruction for their initials when the task was completed. However, their usual work situation meant that this step was unnecessary, be-

[7] It is interesting to note that from the perspective of the pilot, it seemed impossible that all three should go out at once. There must have been a common mode failure—but what was it? The only thing they could think of was that it must be an electrical system problem. In actuality, it was a common mode failure, though a different one than they hypothesized.

cause someone else (apparently their supervisor) was already installing new O-rings on the chip detectors.

The night before the incident, an unusual event occurred. When the mechanics were ready to replace master chip detectors, they found there were no chip detectors in the foreman's cabinet. The mechanics had to get the parts from the stockroom. The chip detectors were wrapped in a "semi-transparent sealed plastic package with a serviceable parts tag." The mechanics took the packages to the aircraft and replaced the detectors in low light conditions. It turned out the chip detectors did not have O-rings attached. The mechanics had not checked for them, before installing them. There was a check procedure against improper seals: motoring the engines to see if oil leaked. The technicians did this, but apparently not for a long enough time to detect oil leaks.

One might argue that the technicians *should* have checked the O-rings on the part, especially since they initialed this item on the work card. But consider that they did not work strictly from the work card—the work card said that they should install a new seal. But they never needed to; someone else always took care of this, so they simply checked off on it. Also, they could not work strictly from procedure; for example, the work card read "monitor engine and check chip detector for leaks" but it didn't specify how long. The mechanics had to fill in the gap, and it turned out the time they routinely used was too short to detect leaks (a breakdown in the system for error detection).

Even without these particular technicians, the system held the potential for breakdown. Several problems or latent failures existed. The unusual event (having to get the part from supply) served as a trigger. (These latent failures are points where a difference might have prevented this particular incident.) Some of these were:

- The fact that someone other than the technicians normally put the O-rings on the chip detectors left in the cabinet and yet did not initial the workcard (effectively leaving no one in charge of O-ring verification).[8]
- The fact that the chip detectors from supply were not packed with O-rings.

[8] There would have been no place to initial since the task of using a new seal was a subtask of the larger step which included replacing the chip detector.

- Personnel did not know what was a sufficient length of time to run the engincs to sec if thcir tasks had been carried out successfully.

Other factors that may have played a role include:

- Low lighting conditions and the necessity of working by feel when inserting the part made it unlikely that the lack of O-rings would have been detected without explicitly checking for them.
- Special training procedures concerning the importance of checking O-rings on the chip detectors were posted on bulletin boards and kept in a binder on the general foreman's desk. Theoretically, the foremen were supposed to ensure that their workers followed the guidance, but there was no follow-up to ensure that each mechanic had read these.
- The variation from a routine way of doing something (opening up the potential for slips of action).

The latent factors involved multiple people in different jobs and the procedures and conditions established for the tasks at the sharp end. Notice how easy it is to miss or rationalize the role of latent factors in the absence of outcome data (see Chapter 6 for more on this point). In this case, the airline had previous O-ring problems, but these were attributed to the mechanics. According to the NTSB report, the propulsion engineering director of the airline, after conferring with his counterparts, said that all the airlines were essentially using the same maintenance procedure but were not experiencing the same in-flight shutdown problems. Hence, it was concluded that the procedures used were valid, and that the problems in installation were due to personnel errors. Also, in reference to the eight incidents that occurred in which O-rings were defective or master chip detectors were improperly installed (prior to this case), the "FAA concluded that the individual mechanic and not Eastern Air Lines maintenance proccdures was at fault" (National Transportation Safety Board, 1984, p. 32).

As Norman (1992) points out, these are problems in the system. These latent failures are not easy to spot; one needs a systems view (i.e., view of the different levels and their interactions) as well as knowledge of how they hold the potential for error. Because of how difficult it is to see these, and how much easier it is to focus on the individual and the actions or omissions that directly impacted the event, the tendency is to attribute the problem to the person at the sharp end. But behind the label "human error" is another story that points to many system-oriented deficiencies that made it possible for the faulty installation to occur and to go undetected.

The best chance of minimizing accidents is by learning how to detect and appreciate the significance of latent failures *before* they combine with other contributors to produce disaster (Reason, 1990).

4

COGNITIVE SYSTEM FACTORS

Distributed Cognitive Systems

We normally think that the canonical case of cognition is an individual rapt in thought. Since we also recognize that an individual's activities occur with some relation to other people, we layer on top of individuals the perspective of a group made up of interacting individuals. And then on top of these two layers, we can point to the role of organizational factors that affect different groups composed of individuals. But this way to parse human-machine systems may he an artifact of how we primarily have studied cognition—individuals alone in tasks removed from any larger context.

If we look at cognition in the "wild," as Ed Hutchins (in press) likes to phrase it, if we look at flightdecks of commercial jet airliners, or control centers that manage space missions, or surgical operating rooms, or control rooms that manage chemical or energy processes, or control centers that monitor telecommunication networks, or many other fields of human activity, what do we see?

First, we do not see cognitive activity isolated in a single individual, but rather cognitive activity going on distributed across multiple agents (Resnick, Levine, and Teasley, 1991; Hutchins, in press). Second, we do not see cognitive activity separated in a thoughtful individual, but rather as a part of a stream of activity (Klein, Orasanu, and Calderwood, 1993). Third, we see these sets of active agents emhedded in a larger group, professional, organizational, or institutional context which constrains their activities, sets up rewards and punishments, defines goals which are not always consistent, and provides resources (e.g., Hutchins, 1990; Thordsen and Klein, 1989; Perkins and Salomon, 1989). Even the moments of individual cognition are set up and conditioned by the larger system and communities of practice in which that individual is embedded.

Fourth, we see phases of activity with transitions and evolutions. Cognitive and physical activity ebbs and flows, with periods of lower activity and more self-paced tasks interspersed with busy, externally paced operations where task performance is more critical. These higher-tempo situations create greater need for cognitive work and at the same time often create greater constraints on cognitive activity (e.g., time pressure, uncertainty, exceptional circumstances, failures, and their associated hazards). We see that there are consequences at stake for the individuals, groups, and organizations involved in the field of activity or affected by that field of activity––such as economic, personal, safety goals.

Fifth, even a casual glance at these domains reveals that tools of all types are everywhere. Almost all activity is aided by something or someone beyond the unit of the individual cognitive agent. More in-depth observation reveals that the technology is often not well adapted to the needs of the practitioner—that much of the technology is clumsy in that it makes new demands on the practitioner, demands that tend to congregate at the higher tempo or higher criticality periods (Woods, 1990b). Close observation reveals that people and systems of people (operators, designers, regulators, etc.) adapt their tools and their activities continuously to respond to indications of trouble or to meet new demands. Furthermore, new machines are not used as the designers intended, but are shaped by practitioners to the contingencies of the field of activity in a locally pragmatic way (Woods et al., 1992).

Looking at cognition in the "wild," maybe it is better to see, as the canonical case, cognition as public and shared, distributed across agents, distributed between external artifacts and internal strategies, embedded in a larger context that partially governs the meanings that are made out of events. Understanding cognition then depends as much on studying the context in which cognition is embedded and the larger distributed system of artifacts and multiple agents, as on studying what goes on between the ears.

The idea suggested hy Hollnagel and Woods (1983) and Hutchins (1991) among others, is that one can look at operational systems as a single-but-distrihuted cognitive system. This operational system cum cognitive system includes the individual people, the communities of practitioners, the organization hoth formal and informal, the high technology artifacts (AI, automation, computer-based visualizations, and intelligent tutors), and the low-technology artifacts (displays, alarms, procedures, paper notes, and training systems) intended to support human practitioners (cf., Hutchins, 1990; Hutchins, 1991 for examples of cognitive system analyses of operational systems).

Operational systems can be thought of as joint or distributed human-machine cognitive systems in that:

- one can describe and study these systems in terms of cognitive concepts such as information flow, knowledge activation, control of attention, etc.,
- cognitive systems are distrihuted over multiple agents, hoth multiple people and mixtures of people and agent-like machines,
- external artifacts modify the activities of agents within a cognitive system and are shaped to function as cognitive tools,[9]
- cognitive systems adapt to the demands of the field of practice and the constraints of the organizational context in which they function.

Hughes, Randall, and Shapiro (1992, p. 5) illustrate the cognitive system viewpoint in their studies of the UK air traffic control system and the reliability of this system.

> If one looks to see what constitutes this reliahility, it cannot he found in any single element of the system. It is certainly not to he found in the equipment . . . for a period of several months during our field work it was failing regularly. . . . Nor is it to he found in the rules and procedures, which are a resource for safe operation but which can never cover every circumstance and condition. Nor is it to be found in the personnel who, though very highly skilled,

[9] There is a reciprocal relationship or mutual shaping between properties of external artifacts and representations of aspects of the field of activity and the cognitive activities distributed over the cognitive system. Properties of these artifacts and representations shape practitioner cognitive strategies and in turn these artifacts are shaped by practitioners to function as tools within a field of activity.

> motivated, and dedicated, are as prone as people everywhere to human error. Rather we believe it is to be found in the cooperative activities of controllers across the 'totality' of the system, and in particular in the way that it enforces the active engagement of controllers, chiefs and assistants with the material they are using and with each other.

The canonical tradition where cognition is a private process of individuals leaches over into discussions of error. One common view is to attribute erroneous actions or assessments to individuals. But in several senses, the proper unit of analysis is not the individual. Erroneous actions that lead to bad consequences involve multiple people embedded in larger systems. It is this operational *system* that fails. When this system fails, there is a breakdown in cognitive activities which are distributed across multiple agents and influenced by the artifacts used by those agents. This is perhaps best illustrated in processes of error detection and recovery which are inherently distributed and play a key role in determining system reliability in practice (e.g., Rochlin et al., 1987).

Cognitive Factors, Problem Demands, Organizational Resources, and Constraints

What factors affect the performance of practitioners in complex settings like medicine, aviation, telecommunications, process plants, and space mission control? Figure 1 (p. 21) provides a schematic overview. For practitioners at the sharp end of the system, there are three classes of cognitive factors that govern how people form intentions to act:

- *Knowledge factors*—factors related to the knowledge that can be drawn on in solving problems in context.
- *Attentional dynamics*—factors that govern the control of attention and the management of workload as situations evolve over time.
- *Strategic factors*—the tradeoffs among different goals that conflict, especially when the people embedded in the situation must act under uncertainty, risk, and the pressure of limited resources (e.g., time pressure, opportunity costs).

They are depicted as interlocking rings at the sharp end of the operational system to point out that these functions overlap and that an effective system depends on their smooth integration across teams of practitioners. Also we do not show a single individual in the figure because these functions rarely are assigned to individuals in a one-to-one fashion. Rather, they are distributed and coordinated across multiple people and across the artifacts they use. This is the basis for thinking about an operational system as a distributed cognitive system.

The above cognitive factors govern the expression of both expertise and error in real systems in conjunction with two other classes of factors. One is the *demands* placed on practitioners by characteristics of the incidents and problems that occur (depicted at the top of Figure 1). These problem demands vary in type and degree—one incident may present itself as a textbook version of a well practiced plan while another may occur accompanied by several complicating factors which together create a more substantive cognitive challenge to practitioners (e.g., Woods, Pople, and Roth, 1990).

One example of a characteristic of a field of activity that affects the kinds of problems that arise is the degree of coupling in the monitored process (Perrow, 1984). Highly coupled processes create or exacerbate a variety of demands on cognitive functions (Woods, 1988). For example, increased coupling creates:

- new knowledge demands, e.g., knowing how different parts of the system interact physically or functionally;
- new attentional demands, e.g., deciding whether or not to interrupt ongoing activities and lines of reasoning as new signals occur;
- new strategic tradeoffs, e.g., one must balance dynamically between the need to diagnose the source of the disturbances and the simultaneous need to cope with the consequences of the disturbances for safety goals.

Problem demands shape the cognitive activities of any agent or agents who might confront that incident. The expression of expertise and error is governed by the interplay of problem demands inherent in the field of activity and the resources of the distributed cognitive system. Figure 1 (p. 21) depicts this relationship through a balance motif at the sharp end. It is at this balance point between demands and resources that failures typically are observed.

Cognitive systems fail from problems in the coordination of these cognitive functions across the distributed operational system, relative to the demands imposed by the field of activity. In terms of *knowledge factors*, some of the possible problems are buggy knowledge (e.g., incorrect model of device function), oversimplifications (Spiro, Coulson, Feltovich, and Anderson, 1988) and inert knowledge. Disruptions in *attentional dynamics* include problems in situation awareness, fixations, and thematic vagabonding. Situation awareness is about the timely perception of critical elements of the situation, about information integration and management, and about anticipating future situations (Sarter and Woods, 1991). Fixations refer to a failure to revise an erroneous situation assessment or course of action despite opportunities to revise. Thematic vagabonding refers to one form of loss of coherence where multiple interacting themes are treated superficially and independently so that the person or team jumps incoherently from one theme to the next (Dorner, 1983). Failures very often can be traced back to *strategic dilemmas and tradeoffs* that arise from multiple interacting and sometimes conflicting goals. Practitioners by the very nature of their role at the sharp end of systems must implicitly or explicitly resolve these conflicts and dilemmas as they are expressed in particular situations (Cook and Woods, 1994).

The final class of factors that we need to consider is the resources and constraints imposed by the *organizational context* in which the practitioners function. The shape of the unitizer (the central shaded region) for the operational system in Figure 1 (p. 21) visually represents "sharp" and "blunt" ends of the system. Recent work on human error has recognized the importance of organizational factors in system failures, e.g., Reason's latent failure model (Reason, 1990, chapter 7). For example, the organizational context influences the knowledge that is available through investments in training and through opportunities to practice rare but high-consequence scenarios. Organizational context also influences the implicit system that affects how more knowledge and more specialist knowledge are brought to bear as an incident evolves and escalates. This occurs through the technology and organizational structures used to access knowledge stored in different systems, places, or people. Organizational context has a particularly important influence on the strategic dilemmas practitioners face. Organizational pressures can exacerbate conflict between different goals and affect the criteria adopted by practitioners in making tradeoffs between goals.

Human Performance at the "Sharp End": Knowledge, Attention, and Goals

The next three sections from Cook and Woods (1994) explore in more detail how various knowledge factors, attentional dynamics, and strategic factors govern the expression of expertise and error in distributed cognitive systems. To accomplish this, we will introduce each section with an actual incident that we have investigated ourselves taken from the field of anesthesiology. Note that one could just as easily substitute incidents from nuclear power operations, aviation, or other domains to illustrate the same concepts.

Each incident was chosen to highlight one of the classes of cognitive factors that are important in human performance as indicated in Table 1 (p. 52). Each could be judged to contain one or more human errors. This judgment is usually the stopping rule for investigators. The incident then can be tabulated in the category "human error" in an incident reporting scheme. But here we take the analysis much further, revealing the complex interplay of the multiple factors sketched in Figure 1 (p. 21) that contributed to the evolution of each incident.

We then return to consider the other two classes of factors represented in Figure 1. The interplay of demands and resources is examined in more detail in terms of the concept of bounded or local rationality. Finally, we will re-examine the relationship of sharp end and blunt end factors.

Knowledge Factors

Knowledge factors refer to what knowledge cognitive agents possess about the system or process in question, how this knowledge is organized so that it can be used flexibly in different contexts, and the processes involved in calling to mind the knowledge relevant to the situation at hand. In other words, they are concerned with the process of bringing knowledge to bear effectively in problem solving.

Category	Exemplar Incident	Cognitive Issues	Examples of Conflicts Present
Knowledge Factors	*Mayocardial infarction in a vasculary surgery patient*	❐ Buggy knowledge ❐ Mental models ❐ Knowledge calibration ❐ Inert knowledge ❐ Simplifications & heuristics ❐ Imprecise knowledge	Imperfect, contradictory, incomplete domain knowledge
Attentional Dynamics	*Hypotension during cardiac surgery*	❐ Situation awareness ❐ Fixations	Limited attentional resource demanded by multiple attractors
Strategic Factors	*Busy weekend operating schedule*	❐ Goal tradeoffs and · decision choice ❐ Risk assessments	Goal tradeoffs Procedural rules that do not apply to all cases Organizational double binds

Table 1. Categories of cognitive factors.

Incident #1: Myocardial Infarction

An elderly patient presented with a painful, pulseless, blue arm indicating a blood clot in one of the major arteries that threatened loss of that limb. The patient had a complex medical and surgical history with high blood pressure, diabetes requiring regular insulin treatment, a prior heart attack and previous coronary artery bypass surgery. The patient also had evidence of recently worsening congestive heart failure, i.e., shortness of breath, dyspnea on exertion and leg swelling (pedal edema). Electrocardiogram (ECG) changes included inverted T waves. Chest x-ray suggested pulmonary edema. The arterial blood gas (ABG) showed markedly low oxygen in the arterial blood (P_aO_2 of 56 on unknown F_iO_2). The blood glucose was high, 800. The patient received furosemide (a diuretic) and 12 units of insulin in the emergency room. The patient was taken to the operating room for removal of the clot under local anesthesia with sedation provided by the anesthetist. In the operating room the patient's blood pressure was high, 210/120; a nitroglycerine drip was started and in an effort to reduce the blood pressure. The arterial oxygen saturation (S_aO_2) was 88% on nasal cannula and did not improve with a rebreathing mask, but rose to the high 90s when the anesthesia machine circuit was used to supply 100% oxygen by mask. The patient did not complain of chest pain but did complain of epigastric pain and received morphine for pain. Urine output was high in the operating room. The blood pressure continued about 200/100. Nifedipine was given sublingually and the pressure fell over ten minutes to 90 systolic. The nitroglycerine was decreased and the pressure rose to 140. The embolectomy was successful. Postoperative cardiac enzyme studies showed a peak about 12 hours after the surgical procedure, indicating that the patient had suffered a heart attack sometime in the period including the time in the emergency room and the operating room. The patient survived.[10]

In this incident the anesthetist confronted several different conditions. The patient's poor cardiac state was one factor that led the anes-

[10] This incident comes from Cook, Woods, and McDonald, 1991 which examined a corpus of cases in anesthesiology and associated human performance issues.

thetist to use local rather than general anesthesia. The arterial blood gas showed markedly low oxygcn in the arterial blood which required several stages of rcsponse to bring it up to an accptablc value. In the operating room the blood pressure was high, but then, after treatment, quite low. To deal with each of thcsc issues thc practitioner was employing a great deal of knowledge (in fact, the description of just a few of the relevant aspects of domain knowlcdge important to the incident would occupy several pages). But these issues also interacted in several ways important to the ovcrall statc of the cardiovascular system. The high glucose value indicated diabctes out of control. This in combination with urine output and the earlier administration of a diuretic in the emergency room indicates that the patient's intravascular volume was low. This probably increased thc demands on a heart that was already starved for oxygen (the prcviously grafted arteries probably were working poorly, a conclusion supportcd by the cvidencc of congestive heart failure, shortness of breath, dyspnea on exertion, leg swelling, and the time since the coronary artery bypass surgery).

In this incident there is evidence that the practitioner was missing or misunderstanding important features of the evolving situation. It scems (and seemed to peer experts who evaluated the incident shortly thereafter; cf., Cook, Woods, and McDonald, 1991) that the practitioncr misunderstood the naturc of thc patient's intravascular volume, believing the volume was high rather than low. The prescnce of high urine output, the previous use of a diuretic (furosemide) in the emergency room, and the high serum glucosc together are indications that a paticnt should be treated differently than was the casc here. The high glucose levels indicated a separate problcm that scemed to be unapprcciated by the practitioner on the scene. In retrospect, other practitioners argued that the patient probably should have receivcd morc intravenous fluid and should have been monitored using more invasive monitoring to determine when enough fluid had been given (e.g., via a catheter that goes through the heart and into the pulmonary artcry).

It is also apparent that many of the practitioner's actions were appropriate in the context of the case as it cvolved. For example, the level of oxygen in the blood was low and the anesthetist pursued sevcral different means of increasing the blood oxygen level. Similarly the blood pressure was high and this, too, was treated, first with nitroglycerin (which may lower the blood pressure but also can protect the heart by increasing its blood flow) and then with nifedipine. The fact that the blood pressure fell much further than intended was probably the result of depleted intravascular volume which was, in turn, the result of the high urinary output provoked by the diuretic and the high serum glucose level. It is this last point that appears to have been unappreciated, at first by the physicians who first saw the patient and then by the anesthetist (note that multiple people were involved in the evolution of the incident). In the opinion of anesthesiologist reviewers of this incident shortly after it occurred, the circumstances of this case should have brought to mind a series of questions about the nature of the patient's intravascular volume. Those questions would then have prompted the use of particular monitoring techniques before and during the surgical procedure.

This incident raises a host of issues regarding how knowledge factors affect the expression of expertise and error. Bringing knowledge to bear effectively in problem solving is a process that involves:

- content (what knowledge)—is the right knowledge there? is it incomplete or erroneous (i.e., "buggy");
- organization—how knowledge is organized so that relevant knowledge can be activated and used effectively; and
- activation—is relevant knowledge "called to mind" in different contexts.

Note that research in this area has emphasized that mere possession of knowledge is not enough for expertise. It is also critical for knowledge to be organized so that it can be activated and used in different contexts (Bransford, Sherwood, Vye, and Rieser, 1986). Thus, Feltovich, Spiro, and Coulson (1989) and others emphasize that one component of human expertise is the *flexible* application of knowledge in *new* situations.

There are at least four lines of overlapping research related to the activation of knowledge in context use by humans performing in complex systems. These include:

- the role of mental models and of knowledge flaws (sometimes called "buggy" knowledge),
- the issue of knowledge calibration,
- the problem of inert knowledge, and
- the use of heuristics, simplifications, and approximations.

Going behind the label "human error" involves investigating how knowledge was or could have been brought to bear in the evolving incident. Any of the above factors could contribute to the activation of knowledge in context—for example, did the participants have incomplete or erroneous knowledge? Were otherwise useful simplifications applied in circumstances that demanded consideration of a deeper model of the factors at work in the case? How knowledge is organized is important to the ability to use it effectively, especially in non-routine circumstances; otherwise, relevant knowledge can remain inert. We will briefly sample a few of the issues in this area.

Mental Models and "Buggy" Knowledge

Knowledge of the world and its operation may be complete or incomplete and accurate or inaccurate. Practitioners may act based on inaccurate knowledge or on incomplete knowledge about some aspect of the complex system or its operation. When the mental model that practitioners hold of such systems is inaccurate or incomplete, their actions may well be inappropriate. These mental models are sometimes described as "buggy" (see Gentner and Stevens, 1983; Rouse and Morris, 1986; Chi, Glaser, and Farr, 1988 for some of the basic results on mental models). The study of practitioners' mental models has examined the models that people use for understanding technological, physical, and physiological processes.

For example, Sarter and Woods (1992, 1994) found that buggy mental models contributed to the problems pilots experienced in using cockpit automation. Airplane cockpit automation has various modes of automatic flight control, ranging between the extremes of automatic and manual. The modes interact with each other in different flight contexts. Having a detailed and complete understanding of how the various modes of automation interact and the consequences of transitions between modes in various flight contexts is a demanding new knowledge requirement for the pilot in highly automated cockpits. They also found that buggy mental models played a role in automation surprises, cases where pilots are "surprised" by the automation's behavior. The buggy knowledge contributed to difficulties in monitoring and understanding automatic system behavior (what is it doing? why did it do that?) and to projecting or anticipating future states (what will it do next?). This is a common finding in complex systems and has also been described in anesthesiologists using microcomputer-based devices (Cook, Potter, Woods, and McDonald, 1991).

It is possible to design experiments that reveal specific bugs or gaps in practitioners' mental models. By forcing pilots to deal with various non-normal situations in simulator studies, it was possible to reveal gaps or errors in their understanding of how the automation works in various situations. Although pilots were able to make the automation work in typical flight contexts, they did not fully exploit the range of the system's capabilities. Pilots tend to adopt and stay with a small repertoire of strategies, in part, because their knowledge about the advantages and disadvantages of the various options for different flight contexts is incomplete. In unusual or novel situations, however, it may be essential to have a thorough understanding of the functional structure of the automated systems and to be able to use this knowledge in operationally effective ways.

Novel or unusual situations can reveal the presence of a "buggy" mental model, and many incidents are associated with situations that are unusual to some degree. In Incident #1 this was certainly the case as the practitioner had to confront multiple interacting issues.

Technology Change and Knowledge Factors

Technology change can have important impacts on knowledge factors in a cognitive system. First, technology change can introduce substantial new knowledge requirements. This is much more than simply a new list of facts about how the computerized or automated device works. For the case of cockpit automation in commercial aviation, pilots must learn and know about the functions of the different automated modes, how to coordinate which mode to use when, how to switch from one mode to another smoothly. In other words, the pilots must know how the automated system works and, especially, they must develop skill at how to work the system (how to coordinate their activities with the activities of the automated systems). For example, pilots must learn about all of the available options, learn and remember how to deploy them across a variety of operational circumstances—especially rarely occurring but more difficult or more critical ones, learn and remember the interface manipulations required to invoke the different modes or features, learn and remember how to interpret or where to find the various indications about which option is active or armed and the associated target values entered for each. Pilots must do more than just possess such knowledge in principle; they must be able to call it to mind and use it effectively in actual task contexts.

The new knowledge demands created by technology change require that more attention be paid to developing and teaching knowledge and strategies for how to coordinate a system of automated resources in varying operational contexts (analogous to cooperating with other team members). In addition, for highly automated systems there is a major constraint that impacts on knowledge demands: if the automation is well engineered in a narrow sense, it will define and work well in a variety of routine situations, but it may not work as well when complicating factors that go beyond the routine occur. Meeting the knowledge demands will require investing in maintaining usable knowledge relevant to the more difficult but infrequently occurring situations. Thus in several ways technology change creates new kinds of training issues and requirements (e.g., Adler, 1986; Bereiter and Miller, 1988).

Significantly, the design of devices, particularly the interface to human practitioners, can either aid or impede the development of useful mental models by practitioners. The absence of a bug-free mental model of a device is more likely to indicate poor device design (low observability) than it is some inadequacy of the user's mental machinery (Norman, 1988). We can draw several generalizations about the interaction between human-device interface and the development of mental models based on studies (e.g., Norman, 1988; Cook, Potter, Woods, and McDonald, 1991). One, users transfer their mental models of past devices to try to explain the perceived behavior of *apparently* similar devices. However, the device's external indications to the user may mislead them about what knowledge or analogies are appropriate to transfer.

Two, users' mental models develop based on experience with the *perceived* behavior of the device. External appearance affects the perception of device structure and function. Flaws in the human-computer interface may obscure important states or events, or incidentally create the appearance of linkages between events or states that are *not* in fact linked (e.g., Cook, Potter, Woods, and McDonald, 1991). This can contribute to buggy user models of device function.

Three, users actively fill in gaps in the model or image the device presents to them. They experiment with ways of using the device that will shape the models of device function that they learn.

Four, apparent simplicity leads users to be unaware of gaps or bugs in their model of the device.

Knowledge Calibration

Results from several studies (Sarter and Woods, 1994; Cook, Potter, Woods, and McDonald, 1991; Moll van Charante et al., 1993) indicate that practitioners are often unaware of gaps or bugs in their model of a device or system due to several factors. This is the issue of knowledge calibration (e.g. Wagenaar and Keren, 1986). All of us have areas where our knowledge is more complete and accurate than in other areas. Individuals are well calibrated if they are aware of how well they know what they know. People are miscalibrated if they are overconfident and believe that they understand areas where in fact their knowledge is incomplete or buggy. Note that degree of calibration is not the same thing as expertise.

There are several factors that could contribute to miscalibration of practitioners' awareness about their knowledge of the domain and the technology with which they work. First, areas of incomplete or buggy knowledge can remain hidden from practitioners because they have the capability to work around these areas by sticking with a few well practiced and well understood methods. Second, situations that challenge practitioner mental models or force them to confront areas where their knowledge is limited and miscalibrated may arise infrequently. Third, studies of calibration have indicated that the availability of feedback, the form of feedback and the attentional demands of processing feedback, can affect knowledge calibration (e.g., Wagenaar and Keren, 1986).

Problems with knowledge calibration can be severe, especially when information technology is involved. For example, many computerized devices fail to provide users with adequate feedback to allow them to learn about (to calibrate their knowledge about) the internal relationships of the deviee. A relationship between poor feedback and miscalibrated practitioners was found in studics of pilot-automation interaction (Sarter and Woods, 1994) and of physieian-automation interaction (Cook, Woods, MeColligan, and Howic, 1991; and Cook, Potter, Woods, and McDonald, 1991). For example, some of the participants in the former study made comments in the post-scenario debriefings such as: "I never knew that I did not know this. I just never thought about this situation." Although this phenomenon is most easily demonstrated when practitioners attempt to use computerized devices, it is probably ubiquitous.

Erroneous actions and assessments ean be due, in part, to a lack of effective feedback on the state of the device or system in question and, in part, due to buggy mental models. The lack of feedback on the state and behavior of the device can in turn limit practitioners' ability to learn from experience and eorreet or elaborate their mental models of system funetion over time. It also limits their ability to learn how to figure out the state of the deviee or automation from the available indieations. All of this is further eomplieated if the situations that stress these problems occur relatively rarely in opcrations.

Knowledge miscalibration is important in several respects. Onc, it ean lead to undcr-reporting of problems with elumsy use of teehnology. Second, when combined with buggy mental models, it ean eontribute to problems in reconstructing the sequence of evcnts in accident investigation where human-maehine interaetion played a role.

Activating Relevant Knowledge in Context: The Problem of Inert Knowledge

Laek of knowledge or buggy knowledge may be one part of the puzzle, but the more eritieal question may be faetors that affeet whcther relevant knowledgc is aetivated and utilized in the aetual problem-solving eontext (e.g., Bransford et al., 1986). The question is not just does the problem solver know some particular pieee of domain knowledge, but does he or she call it to mind when it is relevant to the problem at hand and does he or she know how to utilize this knowledge in problem solving? We tend to assume that if a person ean be shown to have some particular knowledge in one situation and context, then this knowledge should be accessible under all conditions where it might be useful. In contrast, a variety of research results have revealed dissociation effects where knowledge accessed in one context remains inert in another (Gentner and Stevens, 1983; Perkins and Martin, 1986). This situation may well have been the case in the first incident: the practitioner knew about the relationships determining the urine output in the sense that he was able to explain the relationships after the incident, but this knowledge was inert because it was not summoned up during the incident.

Thus, the fact that people possess relevant knowledge does not guarantee that this knowledge will be activated when needed. The critical question is not to show that the problem solver possesses domain knowledge, but rather the more stringent criterion that situation-relevant knowledge is accessible under the conditions in which the task is performed. Knowledge that is accessed only in a restricted set of contexts is called *inert knowledge*. Inert knowledge may be related to cases that are difficult to handle, not because problem solvers do not know the individual pieces of knowledge needed to build a solution, but because they have not confronted the need to join the pieces together previously.[11] Thus, the practitioner in the first incident could be said to *know* about the relationship between blood glucose, furosemide, urine output, and intravascular volume but also to *not know* about that relationship in the sense that the knowledge was not activated at the time when it would have been useful. Studies of practitioner interaction with computerized systems show that the same pattern can occur with computer aids and automation. Sarter and Woods (1994) found that some pilots possessed knowledge in the sense of being able to recite the relevant facts in debriefing, but they were unable to apply the same knowledge successfully in an actual flight context, that is, their knowledge was inert.

Results from accident investigations often show that the people involved did not call to mind all the relevant knowledge during the incident although they "knew" and recognized the significance of the knowl-

[11] Note that inert knowledge is a concept that overlaps both knowledge and attention in that it refers to knowledge that is present in some form but not activated in the appropriate situation. The interaction of the three cognitive factors is the norm.

edge afterwards. The triggering of a knowledge item *X* may depend on subtle pattern recognition factors that are not present in every case where *X* is relevant. Alternativcly, that triggering may depend critically on having sufficient time to process all the available stimuli in order to extract the pattern. This may explain the difficulty practitioners have in "seeing" the relevant details in a certain case where the pace of activity is high and where there are multiple demands on the practitioner. These circumstances were present in Incident #1 and are typical of systems "at the edge of the performance envelope."

One implication of these results is that training experiences should conditionalize knowledge to its use in the contexts where it is likely to be needed. In other words, practitioners do not only need to know how the computerized system works; they need to know how to work the system in differing operational circumstances.

Oversimplifications

People tend to cope with complexity through simplifying heuristics. Heuristics are useful because they are usually relatively easy to apply and minimize the cognitive effort required to produce decisions. These simplifications may be useful approximations that allow limited resource practitioners to function robustly over a variety of problem demand factors (Woods, 1988) or they may be distortions or mis-conceptions that appear to work satisfactorily under some conditions but lead to error in others. Feltovich et al. (1989) call the latter "over-simplifications."

In studying the acquisition and representation of complex concepts in biomedicine, Feltovich et al. (1989) found that various oversimplifications were held by some medical students and even by some practicing physicians. They found that ". . . bits and pieces of knowledge, in themselves sometimes correct, sometimes partly wrong in aspects, or sometimes absent in critical places, interact with each other to create large-scale and robust misconceptions" (Feltovich et al., 1989, p. 162). Examples of kinds of oversimplification include (see Feltovich, Spiro, and Coulson, 1993):

- seeing different entities as more similar than they actually are,
- treating dynamic phenomena statically,

- assuming that some general principle accounts for all of a phenomenon,
- treating multidimensional phenomena as unidimensional or according to a subset of the dimensions,
- treating continuous variables as discrete,
- treating highly interconnected concepts as separable,
- treating the whole as merely the sum of its parts.

Feltovich and his colleagues' work has important implications for the teaching and training of complex material. Their studies and analyses challenge the view of instruction that presents initially simplified material in modules that decompose complex concepts into their simpler components with the belief that these will eventually "add up" for the advanced learner (Feltovich et al., 1993). Instructional analogies, while serving to convey certain aspects of a complex phenomenon, may miss some crucial ones and mislead on others. The analytic decomposition misrepresents concepts that have interactions among variables. The conventional approach may produce a false sense of understanding and inhibit pursuit of deeper understanding because learners may resist learning a more complex model once they already have an apparently useful simpler one (Spiro et al., 1988). Feltovich and his colleagues have developed the theoretical basis for a new approach to advanced knowledge acquisition in ill-structured domains.

Why do practitioners utilize simplified or oversimplified knowledge? These simplifying tendencies may occur because of the cognitive effort required in demanding circumstances.

> It is easier to think that all instances of the same nominal concept . . . are the same or bear considerable similarity. It is easier to represent continuities in terms of components and steps. It is easier to deal with a single principle from which an entire complex phenomenon 'spins out' than to deal with numerous, more localized principles and their interactions. . . (Feltovich et al., 1989, p. 131).

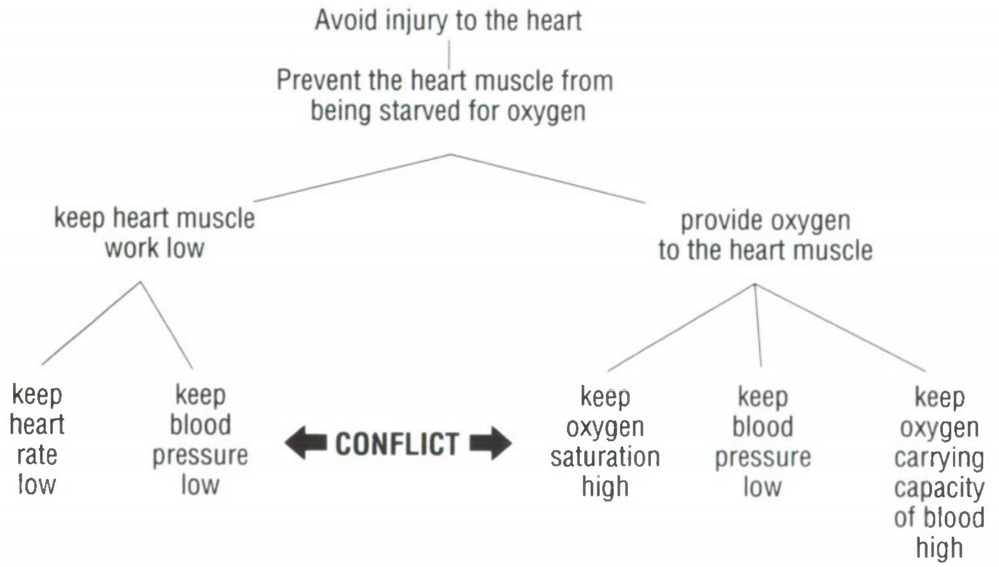
Simplifications may be adaptive, first, because the effort required to follow more "ideal" reasoning paths may be so large that it would keep practitioners from acting with the speed demanded in actual environments. This has been shown elegantly by Payne, Bettman, and Johnson (1988) and by Payne, Johnson, Bettman, and Coupey (1990) who demonstrated that simplified methods will produee a higher proportion of eorrect ehoiees betwcen multiple alternativcs under eonditions of time pressure.

Seeond, there may be uncertainties, imprccision, or conflicts that need to be resolved in each individual case by the praetitioncr. In Ineident #1, for example, there are eonfliets between the need to keep the blood pressure high and thc need to keep the blood pressurc low (Figure 6, p. 65). As is often the casc in this and similar domains, the locus of confliet may vary from ease to ease and from moment to moment. The heart depends on blood pressure for its own blood supply, but incrcasing the blood pressure also increases the work it is required to perform. The practitioner must deeide what blood pressure is aeeeptable. Many faetors enter into this dccision process. For examplc, how labilc is the blood pressure now? How will attempts to reduee blood pressure affeet other physiologieal variables? How is the pressure likcly to change without therapy? How long will the surgery last?

In summary, hcuristics represcnt effective and nccessary adaptations to the demands of real workplaces (Rasmussen, 1986). The problem may not always bc thc shortcut or simplification itself, but whether praetitioners know the limits of the shortcuts, ean rccognize situations where the simplification is no longer relevant, and have thc ability to usc more complex concepts, mcthods, or modcls (or thc ability to integratc hclp from speeialist knowledge sourees) whcn thc situation they facc dcmands it

Incident #1 and Knowledge Factors

It can be quite difficult to dctermine how buggy mental models, ovcrsimplifications, inert knowlcdge, or some eombination was involved in an ineident. The kinds of data available about how the ineident evolved, the speeific practitioners involved, thc practitioncr population in general, and their training experiencces arc neeessary to understand the role of knowledge faetors. But these data are rarely available without special effort from investigators and researchers. In Incident #1, the combination of factors prcsent in the ineidcnt was unusual, and


Avoid injury to the heart
Prevent the heart muscle from being starved for oxygen
keep heart muscle work low
provide oxygen to the heart muscle
keep heart rate low
keep blood pressure low
CONFLICT
keep oxygen saturation high
keep blood pressure low
keep oxygen carrying capacity of blood high

Copyright © 1991 R I Cook

Figure 6. Conflicting goals in anesthesiology 1. For a cardiac surgery patient the blood pressure should be kept low to minimize the work of the heart, but the blood pressure should be kept high to maximize the blood flow to heart muscle. How practitioners at the sharp end resolve this conflict depends on several factors (From Cook, Woods, and McDonald, 1991).

it is possible that the participant had a buggy mental model of the relationship between these factors (e.g., the participant did not work routinely in cardiovascular anesthesia as a subspecialty of anesthesia).

Given the complexities of the case, oversimplification strategies could be implicated. The combination of congestive heart failure with low circulating blood volume is unusual. Congestive heart failure is normally associated with too much fluid in the circulation. But in this case high blood glucose and a diuretic drug (furosemide) led to too little circulating volume. The participant seemed to be responding to each issue in isolation and missing the interconnections that would have led to a more coherent approach.

Inert knowledge may have played a role as well. The cues in this case were not the ones that are usually associated with deeper knowledge about the inter-relationships of intravascular volume, glucose level, and cardiovascular volume. The attentional demands of the patient's low oxygen saturation and other abnormal conditions could have prevented the participants from exploring their knowledge sufficiently as related to this particular situation *in situ*.

Interestingly, practitioners are acutely aware of how deficient their rules of thumb may be and how certain situations may require abandoning the cognitively easy method in favor of more cognitively demanding "deep thinking." For example, senior anesthesiologists commenting on the first incident were critical of practitioner behavior:

> . . . this man was in major sort of hyperglycemia and with popping in extra Lasix [furosemide] you have a risk of hypovolemia from that situation. I don't understand why that was quietly passed over, I mean that was a major emergency in itself . . . . This is a complete garbage amount of treatment coming in from each side, responding from the gut to each little bit of stuff [but it] adds up to no logic whatsoever . . . the thing is that this patient [had] an enormous number of medical problems going on which have been simply reported [but] haven't really been addressed . . . (Cook, Woods, and McDonald, 1991, p. 35-6).

This is a pointed remark, made directly to the participant by those with whom he worked each day. While it is not couched in the language of cognitive science, it remains a graphic reminder that practitioners recognize the importance of cognition to their success and sometimes distinguish between expert and inexpert performance by looking for evidence of cognitive processes.

Attentional Dynamics

Attentional dynamics refer to the factors that operate when cognitive systems function in dynamic, evolving situations—how to manage workload in time; how to control attention when there are multiple signals and tasks competing for a limited attentional focus. In many ways this is the least explored frontier in cognitive science and human-machine cooperation, especially with respect to error (but see Hollister, 1986; Gopher, 1991; Moray, Dessouky, Kijowski, and Adapathya, 1991; and Woods, 1992).

Incident #2: Hypotension

> *During a coronary artery bypass graft procedure an infusion controller device used to control the flow of a potent drug to the patient delivered a large volume of drug at a time when no drug should have been flowing. Five of these microprocessor-based devices were set up in the usual fashion at the beginning of the day, prior to the beginning of the case. The initial sequence of events associated with the case was unremarkable. Elevated systolic blood pressure (>160 torr) at the time of sternotomy prompted the practitioner to begin an infusion of sodium nitroprusside via one of the devices. After this device was started at a drop rate of 10/min, the device began to sound an alarm. The tubing connecting the device to the patient was checked and a stopcock (valve) was found to be closed. The operator opened the stopcock and restarted the device. Shortly after restart, the device alarmed again. The blood pressure was falling by this time, and the operator turned the device off. Over a short period, hyperten-*

> *sion gave way to hypotension (systolic pressure <60 torr). The hypotension was unresponsive to fluid challenge but did respond to repeated injections of neosynephrine and epinephrine. The patient was placed on bypass rapidly. Later, the container of nitroprusside was found to be empty; a full bag of 50 mg in 250 ml was set up before the case.*

The physicians involved in the incident were comparatively experienced device users. Reconstructing the events after thc incident led to the conclusion that the device was assembled in a way that would allow free flow of drug. Drug delivery was blocked, however, by the closed downstream stopcock. The device was started, hut the machine did not detect any flow of drug (the stopcock was closed) triggering visual and auditory alarms. When thc stopcock was opencd, free flow of fluid containing drug began. The controllcr was rcstartcd, hut the machine again detected no drip rate because flow was a continuous stream and no individual drops were being formcd. The controller alarmed again with the same message which appeared to indicate that no flow had occurred. Between opening the stopcock and the generation of the error message, sufficient drug was delivered to substantially reduce the blood pressure. The operator saw thc reduced hlood pressure, concluded that the sodium nitroprusside drip was not required and pushed the button marked "off." This powered down the device, but the flow of drug continued. The blood pressure fell even further, prompting a diagnostic search for sources of low blood pressure. The sodium nitroprusside controller was seen to be off. Treatment of the low blood pressure itself commenced and was successful. The patient suffered no sequelae.[12]

In Incident #2 the data are strong enough to support a reconstruction of some of the actual changes in focus of attention of thc participants during the incident. The free flow of the drug hegan when one of the physicians opened the stopcock, but this source of the hypotension was not identified until the hag of fluid was nearly empty. A numher of factors in the environment contrihute to the failure to observe (i.e., attend to) the unintended flow of drug via the infusion device including

[12] This case is described more fully in Cook et al., 1992, and weaknesses in the infusion device from the point of view of human-computer cooperation are covered in Moll van Charante et al., 1993.

(1) the drip chamber was ohscured hy the machine's sensor, making visual inspection difficult, (2) presence of an aluminum shield around the fluid bag, hiding its decreasing volume, (3) misleading alarm messages from the device, and (4) presence of multiple devices making it difficult to trace the tubing pathways.

There are also extra-environmental factors that contrihuted to the failure to observe the free flow. Most importantly, the practitioners reported that they turned the device off as soon as the pressure fell and the device alarmed a second time. In their view of the external world, the device was off, therefore not delivering any drug, and therefore not a plausible source of the hypotension. When they looked at the device, the displays and alarm messages indicated that the device was not delivering drug or later that it had heen turned off. The issue of whether "off" might have meant something else (e.g., that the device was powered down but a path for fluid flow remained open) might have heen revisited had the situation heen less demanding, hut the fall in hlood pressure was a critical threat to the patient and demanded the limited resource of attention. Remarkably, the practitioners intervened in precisely the right way for the condition they were facing. The choice of drugs to increase the blood pressure was ideal to counteract the large dose of sodium nitroprusside that the patient was receiving. Attention did not focus on the fluid hags on the infusion support tree until the decision was made to start an infusion of the antagonist drug and a hag for that drug was heing placed on the support tree.

.This incident is remarkable in part for the way in which it shows both the fragility and robustness of human performance. The inability to diagnose the cause of hypotension is in contrast to the ability to manage successfully the complications of the inadvertent drug delivery. There are a number of potential causes of hypotension in the cardiac surgical patient. In this case, successful diagnosis of the root cause was less important than successful treatment of the consequences of the prohlem. The practitioners were quick to correct the physiologic, systemic threat even though they were unable to diagnose its source. They shifted their focus of attention from diagnosing the source of the hypotension to responding to the immediate threat to the patient. This ability to shift from diagnosis to *disturbance management* is crucial in the operating room and in other domains to maintain the system in a stable configuration and permit later diagnosis and correction of the underlying faults (Woods, 1988; in press-a).

The control of attention is an important issue for those trying to understand human performance, especially in event-rich domains such as flightdecks, operating rooms, or control centers. Attention is a limited resource. One cannot attend to more than one thing at a time, and so shifts of attention are necessary to be able to "take in" the ways in which the world is changing. When something in the world is found that is anomalous (what is sensed in the world is not consistent with what is expected by the observer) attention focuses on that thing and a process of investigation begins that involves other shifts of attention. This process is ongoing and has been described by Neisser as the perceptual or *cognitive cycle* (Neisser, 1976; see Tenney, Jager Adams, Pew, Huggins, and Rogers, 1992 for one application of his concepts to the aviation domain). It is a crucial concept for those trying to understand human performance because it is the basis for all diagnosis and action. Nothing can be discovered in the world without attention; no intended change in the world can be effected without shifting attention to the thing being acted upon. At least two kinds of human performance problems are based on attentional dynamics. The first is a loss of situation awareness and the second is psychological fixation.

Loss of Situation Awareness

Situation awareness is a label that is often used to refer to many of the cognitive processes involved in what we have called here attentional dynamics (Endsley, 1988; Sarter and Woods, 1991; Tenney et al., 1992).[13] Just a few of the cognitive processes that may be involved when one invokes the label of situation awareness arc control of attention (Gopher, 1991), mental simulation (Klein and Crandall, in press), forming expectancies (Woods, in press-b; Johnson, Grazioli, Jamal, and Zualkernan, 1992), directed attention (Woods, 1992), and contingency

[13] There are many debates about what is situation awareness and attempts to measure it as a unitary phenomenon. For example, does situation awareness refer to a product or a process? It is not our intention here to engage in or outline a position in these debates. Here we are using the label situation awareness, since it is a commonly used expression, to point to the cognitive processes involved in the control of attention.

planning (Orasanu, 1990). Because the concept involves tracking processes in time, it can also be described as mental bookkeeping—keeping track of multiple threads of different but interacting sub-problems as well as of influences of the activities undertaken to control them (Cook, Woods, and McDonald, 1991).

Maintaining situation awareness necessarily requires shifts of attention between the various threads. It also requires more than attention alone, for the objective of the shifts of attention is to inform and modify a coherent picture or model of the system as a whole. Building and maintaining that picture require cognitive effort.

Breakdowns in these cognitive processes can lead to operational difficulties in handling the demands of dynamic, event-driven incidents. In aviation circles this is known as "falling behind the plane" and in aircraft carrier flight operations it has been described as "losing the bubble" (Roberts and Rousseau, 1989). In each case what is being lost is the operator's internal representation of the state of the world at that moment and the direction in which the forces active in the world are taking the system that the operator is trying to control. Dorner (1983) calls breakdowns in mental bookkeeping "thematic vagabonding" as the practitioner jumps from thread to thread in an uncoordinated fashion (the response in Incident #1 may have possessed an element of vagabonding).

Fischer, Orasanu, and Montvalo (1993) examined the juggling of multiple threads of a problem in a simulated aviation scenario. More effective crews were better able to coordinate their activities with multiple issues over time; less effective crews traded one problem for another. More effective crews were sensitive to the interactions between multiple threads involved in the incident; less effective crews tended to simplify the situations they faced and were less sensitive to the constraints of the particular context they faced. Less effective crews "were controlled by the task demands" and did not look ahead or prepare for what would come next. As a result, they were more likely to run out of time or encounter other cascading problems. Interestingly, there were written procedures for each of the problems the crews faced. The cognitive work associated with managing multiple threads of activity is different from the activities needed to merely follow the rules.

Obtaining a clear, empirically testable model for situation awareness is difficult. For example, Hollister (1986) presents an overview of a model of divided attention operations—tasks where attention must be divided across a number of different input channels and where the focus of attention changes as new events signal new priorities. This model then defines an approach to breakdowns in attentional dynamics (what has been called a divided attention theory of error) based on human divided attention capabilities balanced against task demands and adjusted by fatigue and other performance-shaping factors. Situation awareness is clearly most in jeopardy during periods of rapid change and where a confluence of forces makes an already complex situation critically so. This condition is extraordinarily difficult to reproduce convincingly in a laboratory setting. Practitioners are, however, particularly sensitive to the importance of situation awareness even though researchers find that a clear definition remains elusive (Sarter and Woods, 1991).

Understanding these attentional dynamics relative to task complexities, and how they are affected by computer-based systems, is a very important research issue for progress in aiding situation awareness and for safety in supervisory control systems (cf. McRuer et al., [Eds.] 1992, *National Academy of Sciences Report on Aeronautical Technologies for the Twenty-First Century*, chapter 11). To meet this research objective we will need to understand more about coordination across human and machine agents, about how to increase the observability of the state and activities of automated systems, and about what are the critical characteristics of displays that integrate multiple sources of data in mentally economical ways.

Failures to Revise Situation Assessments: Fixation or Cognitive Lockup

The results of several studies (e.g., De Keyser and Woods, 1990; Cook, McDonald, and Smalhout, 1989; Johnson et al., 1981; Johnson, Moen, and Thompson 1988; Gaba and DeAnda, 1989) strongly suggest that one source of error in dynamic domains is a *failure to revise* situation assessment as new evidence comes in. Evidence discrepant with the agent's or team's current assessment is missed or discounted or rationalized as not really being discrepant with the current assessment. The operational teams involved in several major accidents seem to have exhibited this pattern of behavior; examples include the Three Mile Island accident (Kemeny et al., 1979) and the Chernobyl accident.

Many critical real-world human problem-solving situations take place in dynamic, event-driven environments where the evidence arrives over time and situations can change rapidly. Incidents rarely spring full blown and complete; incidents *evolve*. In these situations, people must amass and integrate uncertain, incomplete, and changing evidence; there is no single well formulated diagnosis of the situation. Rather, practitioners make provisional assessments and form expectancies based upon partial and uncertain data. These assessments are incrementally updated and revised as more evidence comes in. Furthermore, situation assessment and plan formulation are not distinct sequential stages, but rather they are closely interwoven processes with partial and provisional plan development and feedback leading to revised situation assessments (Woods and Roth, 1988; Klein et al., 1993; Woods, in press-b).

In psychological fixations (also referred to as cognitive lockup and cognitive hysteresis), the initial situation assessment tends to be appropriate, in the sense of being consistent with the partial information available at that early stage of the incident. As the incident evolves, however, people fail to revise their assessments in response to new evidence, evidence that indicates an evolution away from the expected path. The practitioners become fixated on an old assessment and fail to revise their situation assessment and plans in a manner appropriate to the data now present in their world. Thus, a *fixation* occurs when practitioners fail to revise their situation assessment or course of action and maintain an inappropriate judgment or action *in the face of opportunities to revise.*

Several criteria are necessary to describe an event as a fixation. One critical feature is that there is some form of *persistence* over time in the behavior of the fixated person or team. Second, opportunities to revise are cues, available or potentially available to the practitioners, that could have started the revision process if observed and interpreted properly. In part, this feature distinguishes fixations from simple cases of inexperience, lack of knowledge, or other problems that impair error detection and recovery (Cook et al., 1989).[14] The basic defining characteristic of fixations is that the immediate problem-solving context has biased the practitioners in some direction. In naturally occurring problems, the context in which the incident occurs and the way the incident evolves activate certain kinds of knowledge as relevant to the evolving incident. This knowledge, in turn, affects how new incoming information is interpreted. After the fact or after the correct diagnosis has been pointed out, the solution seems obvious, even to the fixated person or team.

De Keyser and Woods (1990) describe several patterns of behavior that have been observed in cases of practitioner fixation. In the first one, "everything but that," the operators seem to have many hypotheses in mind, but never entertain the correct one. Their external behavior looks incoherent because they are often jumping from one action to another one without any success. The second one is the opposite: "this and nothing else." The practitioners are stuck on one strategy, one goal, and they seem unable to shift or to consider other possibilities. One can observe a great deal of persistence in their behavior in this kind of case; for example, practitioners may repeat the same action or recheck the same data channels several times. This pattern is easy to see because of the unusual level of repetitions despite an absence of results. The practitioners often detect the absence of results themselves but without any change in strategy. A third pattern is "everything is O.K." In this case, the practitioners do not react to the change in their environment. Even if there are multiple cues and evidence that something is going wrong, they do not seem to take these indicators at face value. They seem to discount or rationalize away indications that are discrepant with their model of the situation. On the other hand, one must keep in mind the demands of situation assessment in complex fields of practice (cf., Woods, in press-b). For example, some discrepant data actually may be red herrings or false alarms which should be discounted for effective diagnostic search (e.g., false or nuisance alarms can be frequent in many

[14] Of course, the interpretation problem is to define a standard to use to determine what cue or when a cue should alert the practitioners to the discrepancy between the perceived state of the world and the actual state of the world. There is a great danger of falling into the hindsight bias when evaluating after the fact whether a cue "should" have alerted the problem solvers to the discrepancy.

systems). This is essentially a strategic dilemma in diagnostic reasoning, the difficulty of which depends in part on the demands of problems and on the observability of the processes in question.

Certain types of problems may encourage fixations by mimicking other situations, in effect, leading practitioners down a *garden path* (Johnson et al., 1988; Johnson, Jamal, and Berryman, 1991; Johnson et al., 1992). In garden-path problems "early cues strongly suggest [plausible but] incorrect answers, and later, usually weaker cues suggest answers that are correct" (Johnson et al., 1988). It is important to point out that the erroneous assessments resulting from being led down the garden path are not due to knowledge factors. Rather, they seem to occur because "a problem-solving process that works most of the time is applied to a class of problems for which it is not well suited" (Johnson et al., 1988). This notion of garden path situations is important because it identifies a task genotype in which people become susceptible to fixations. The problems that occur are best attributed to the interaction of particular environmental (task) features and the heuristics people apply (local rationality given difficult problems and limited resources), rather than to any particular bias or problem in the strategies used. The way that a problem presents itself to practitioners may make it very easy to entertain plausible but in fact erroneous possibilities.

Diagnostic problems fraught with inherent uncertainties are common in complex fields of practice (Woods, in press-b). As a result, it may be necessary for practitioners to entertain and evaluate what turn out later to be erroneous assessments. Problems arise when the revision process breaks down and the practitioner becomes fixated on an erroneous assessment, missing, discounting, or re-interpreting discrepant evidence (see Johnson et al., 1988; Roth, Woods, and Pople, 1992 for analyses of performance in garden path incidents). What is important is the process of error detection and recovery which fundamentally involves searching out and evaluating discrepant evidence to keep up with a changing incident.

Several cognitive processes involved in attentional dynamics may give rise to fixation; these include:

- breakdowns in shifting or scheduling attention as the incident unfolds;

- factors of knowledge organization and access that make critical knowledge inert;
- difficulties calling to mind alternative hypotheses that could account for observed anomalies—problems in the processes underlying hypothesis generation;
- problems in strategies for situation assessment (diagnosis) given the probability of *multiple* factors, e.g., how to value parsimony (single factor assessments) versus multi-factor interpretations.

Fixation may represent the down side of normally efficient and reliable cognitive processes involved in diagnosis and disturbance management in dynamic contexts (Woods, in press-a, provides a more detailed examination of the reasoning processes involved and how they break down). Although fixation is fundamentally about problems in attentional dynamics, it may also involve inert knowledge (failing to call to mind potentially relevant knowledge such as alternative hypotheses) or strategic factors (tradeoffs about what kinds of explanations to prefer).

It is clear that in demanding situations where the state of the monitored process is changing rapidly, there is a potential conflict between the need to revise the situation assessment and the need to maintain coherence. Not every change is important; not every signal is meaningful. The practitioner whose attention is constantly shifting from one item to another may not be able to formulate a complete and coherent picture of the state of the system. For example, the practitioner in Incident #1 was criticized for failing to build a complete picture of the patient's changing physiological state. Conversely, the practitioner whose attention does not shift may miss cues and data that are critical to updating the situation assessment. This latter condition may lead to fixation. How practitioners manage this conflict is largely unstudied.

Given the kinds of cognitive processes that seem to be involved in fixation, there are a variety of techniques that, in principle, may reduce this form of breakdown. Data on successful and unsuccessful revision of erroneous situation assessments show that it usually takes a person with a fresh point of view on the situation to break a team or individual out of a fixation (Woods et al., 1987). Note that this result again reveals the multi-agent nature of cognitive activities in the wild. Thus, one can change the architecture of the distributed system to try to ensure a fresh point of view, i.e., one that is unbiased by the immediate context. Practically, this has been tried by adding a new person to the team who has a different background and viewpoint or by organizing the team so that some members develop their views of the evolving situation separately from others. Another approach is to try to develop distributed system architectures where one person or group criticizes the assessments developed by the remainder of the group (e.g., a Devil's advocate team member; Schwenk and Cosier, 1980). A third direction is predicated on the fact that poor feedback about the state and behavior of the monitored process, especially related to goal achievement, is often implicated in fixations and failures to revise. Thus, one can provide practitioners with new *kinds* of representations about what is going on in the monitored process (cf., Woods et al., 1987 for examples from nuclear power which tried this in response to the Three Mile Island accident).

Strategic Factors

Another set of factors at work in distributed cognitive systems is strategic in nature. People have to make tradeoffs between different but interacting or conflicting goals, between values or costs placed on different possible outcomes or courses of action, or between the risks of different errors. They must make these tradeoffs while facing uncertainty, risk, and the pressure of limited resources (e.g., time pressure, opportunity costs).

Incident #3: Busy Weekend Operating Schedule

> *On a weekend in a large tertiary care hospital, the anesthesiology team (consisting of four physicians, three of whom were residents in training) was called on to perform anesthetics for an* in vitro *fertilization, a perforated viscus, reconstruction of an artery of the leg, and an appendectomy in one building, and one exploratory laparotomy in another building. Each of these cases was an emergency, that is, a case that cannot be delayed for the regular daily operating room schedule. The exact sequence in which the cases were done depended on multiple factors. The situation was complicated by a demanding nurse who insisted*

> *that the exploratory laparotomy be done ahead of other cases. The nurse was only responsible for that single case; the operating room nurses and technicians for that case could not leave the hospital until the case had been completed. The surgeons complained that they were being delayed and their cases were increasing in urgency because of the passage of time. There were also some delays in preoperative preparation of some of the patients for surgery. In the primary operating room suites, the staff of nurses and technicians were only able to run two operating rooms simultaneously. The anesthesiologist in charge was under pressure to attempt to overlap portions of procedures by starting one case as another was finishing so as to use the available resources maximally. The hospital also served as a major trauma center which means that the team needed to be able to start a large emergency case with minimal (less than ten minutes) notice. In committing all of the residents to doing the waiting cases, the anesthesiologist in charge produced a situation in which there were no anesthetists available to start a major trauma case. There were no trauma cases, and all the surgeries were accomplished. Remarkably, the situation was so common in the institution that it was regarded by many as typical rather than exceptional.*

This incident is remarkahle in part hecause it is regarded as unremarkable by the participants. These kinds of scheduling issues recur and are considered by many to be simply part of the job. In the institution where the incident occurred, the role of being anesthetist in charge during evening and weekend duty is to determine which cases will start and which ones will wait. Being in charge also entails handling a variety of emergent situations in the hospital including calls to intuhate patients on the floors, requests for pain control, and emergency room trauma cases. The person in charge also serves as a backup resource for the operations in progress. In this incident, the anesthetist in charge committed all of her availahle resources, including herself, to doing anesthesia. This effectively eliminated the in-charge-person's ahility to act as a huffer or extra resource for handling an additional trauma case or a request from the floor. There were strong incentives to commit the resources, but also a simultaneous incentive to avoid that commitment. Trauma severe enough to demand immediate surgery occurs in this institution once or twice a week.

Factors that played a role in the anesthetist's decision to commit all available resources included the relatively high urgency of the cases, the absence of a trauma alert (indication that a trauma patient was in route to the hospital), the time of day (fairly early; most trauma is seen in the late evening or early morning hours), and pressure from surgeons and nurses. Another seemingly paradoxical reason for committing the resources was the desire to free up the resources by getting the cases completed before the late evening when trauma operations were more likely. These factors are not severe or even unusual. Rather, they represent the normal functioning of a large urban hospital as well as the nature of the conflicts and double binds that occur as part of the normal playing field of the specialty.

The conflicts and the tradeoffs between highly unlikely but highly undesirable events and highly likely but less catastrophic ones that occurred in Incident #3 are examples of strategic factors. People have to make tradeoffs between different but interacting or conflicting goals. One may think of these tradeoffs in terms of simplistic global examples like safety versus economy. Tradeoffs also occur on other kinds of dimensions. In dynamic fault management, for example, there is a tradeoff with respect to when to commit to a course of action. Practitioners have to decide whether to take corrective action early in the course of an incident with limited information or to delay the response to wait for more data to come in, to search for additional findings, or to ponder additional alternative hypotheses.

A salient example of this process occurred during the Apollo 13 mission following what turned out to be an explosion in the cryogenics systems which led to the loss of many critical systems and a serious threat to the ability of the spacecraft to return safely to earth (see Murray and Cox, 1989, p. 409).

> Lunney [the Flight Director] was persistent because the next step they were contemplating was shutting off the reactant valve in Fuel Cell 1, as they had done already in Fuel Cell 3. If they shut it off and then came up with a . . . solution that suddenly got the O_2

> pressures back up, the door would still be closed on two-thirds of the C.S.M's power supply. It was like shooting a lame horse if you were stranded in the middle of a desert. It might be the smart thing to do, but it was awfully final. Lunney, like Kranz before him, had no way of knowing that the explosion had instantaneously closed the reactant valves on both fuel cells 1 and 3. At ten minutes into his shift, seventy-nine minutes after the explosion, Lunney was close to exhausting the alternatives.
>
> "You're ready for that now, sure, absolutely, EECOM [the abbreviation for one of the flight controller positions]?"
>
> "That's it, Flight."
>
> "It [the oxygen pressure] is still going down and it's not possible that the thing is sorta bottoming out, is it?"
>
> "Well, the rate is slower, but we have less pressure too, so we would expect it to be a bit slower."
>
> "You are sure then, you want to close it?"
>
> "Seems to me we have no choice, Flight."
>
> "Well . . ."
>
> Burton, under this onslaught, polled his back room one last time. They all agreed.
>
> "We're go on that, Flight."
>
> "Okay, that's your best judgment, we think we ought to close that off, huh?"
>
> "That's affirmative."
>
> Lunney finally acquiesced. "Okay. Fuel Cell 1 reactants coming off."
>
> It was uncharacteristic behavior by Lunney—"stalling," he would later call it. "Just to be sure. Because it was clear that we were at the ragged edge of being able to get this thing back. . . . That whole night, I had a sense of containing events as best we could so as not to make a serious mistake and let it get worse."

Practitioners also trade off between following operational rules or taking action based on reasoning about the case itself (cf., Woods et al., 1987). Do the standard rules apply to this particular situation when some additional factor is present that complicates the textbook scenario? Should we adapt the standard plans or should we stick with them regardless of the special circumstances? Strategic tradeoffs can also involve coordination among agents in the distributed human-machine cognitive system (Roth, Bennett, and Woods, 1987). A machine expert recommends a particular diagnosis or action, but what if your own evaluation is different? What is enough evidence that the machine is wrong to justify disregarding the machine expert's evaluation and proceeding on your own evaluation of the situation? For example, the pulse oximeter used in the operating room may provide an unreliable reading under some circumstances (e.g., low perfusion). How does one know whether the current reading of 80% is indicative of an artifact or is an accurate representation of the patient's oxygen saturation?

Criterion setting on these different tradeoffs may not be a conscious process or a decision made by individuals. It may be much more likely that they are emergent properties of systems of people, either small groups or larger organizations. The criteria may be fairly labile and susceptible to influence, or they may be relatively stable and difficult to change. The tradeoffs may create explicit choice points for practitioners embedded in an evolving situation, or they may cast a shadow of influence over the attentional dynamics relating intertwined events, tasks, and lines of reasoning.

In hindsight, practitioners' choices or actions can often seem to be simple blunders. Indeed, most of the media reports of human error in aviation, transportation, medicine, etc. are tailored to emphasize the extreme nature of the participants' behavior. But a more careful assessment of the distributed system may reveal strategic factors at work. Behavior in the specific incident derives from how the practitioners set their tradeoff criteria across different kinds of risks from different kinds of incidents that could occur. Because incidents usually are evaluated as isolated events, such tradeoffs can appear in hindsight to be unwise or even bizarre. This is because the individual incident is used as the basis for examining the larger system (see the discussion of hindsight bias in Chapter 6).

When strategic factors are involved in an incident, changing the behavior of the operational system requires a larger analysis of how one should make the tradeoff. It also involves meaningfully and consistently communicating this policy to the operational system so that practitioners adopt it as their criterion. This may implicitly or explicitly involve the commitment of a different system (an organization's management, an entire industry, a regulatory process). Lanir, Fischhoff, and Johnson (1988) provide an excellent example through their formal analysis of criteria setting for risk-taking within a distributed cognitive system. The danger in missing the role of strategic tradeoffs in producing the observed behavior of operational systems is that the changes made or the messages received by the practitioners exacerbate the dilemma.

Many strategic factors can be elaborated; two forms are discussed here. The first is the presence of goal conflicts, and the other is the responsibility-authority double bind.

Goal Conflicts

Multiple goals are simultaneously relevant in actual fields of practice. Depending on the particular circumstances in operation in a particular situation, the means to influence these multiple goals will interact, potentially producing conflicts between different goals. To perform an adequate analysis of the human performance in an evolving incident requires an explicit description of the strategic factors acting in the incident including the interacting goals, the tradeoffs being made, and the pressures present that shift the operating points for these tradeoffs.

The impact of potential conflicts may be quite difficult to assess. Consider the anesthesiologist. Practitioners' highest level goal (and the one most often explicitly acknowledged) is to protect patient safety. But that is not the only goal. There are other goals, some of which are less explicitly articulated. These goals include reducing costs, avoiding actions that would increase the likelihood of being sued, maintaining good relations with the surgical service, maintaining resource elasticity to allow for handling unexpected emergencies, and others (Figure 7, p. 83).

In a given circumstance, the relationships between these goals can produce conflicts. In the daily routine, for example, maximizing patient safety and avoiding lawsuits create the need to maximize information about the patient through pre-operative workup. The anesthetist may find some hint of a potentially problematic condition and con-

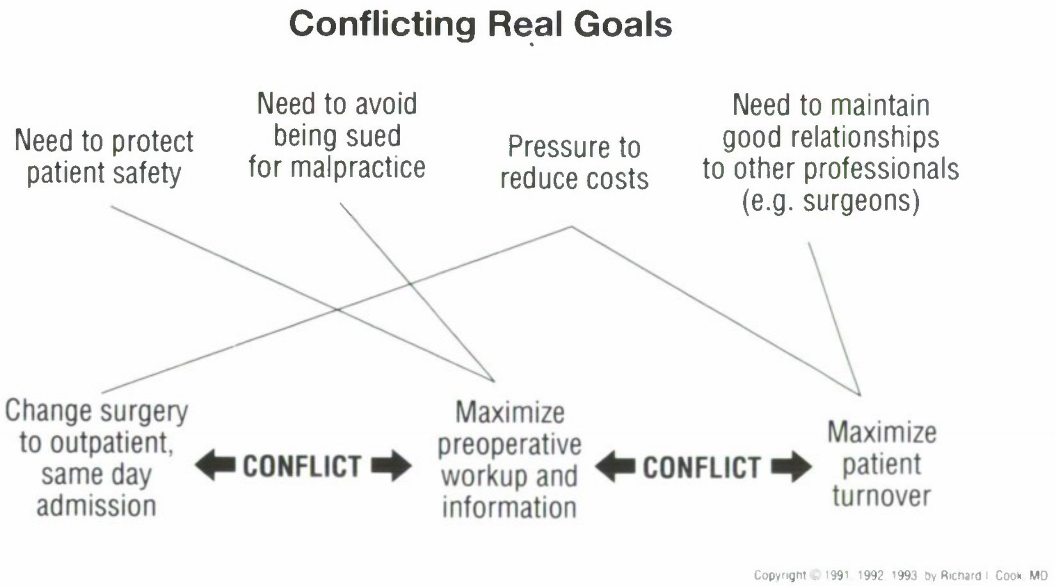

Conflicting Real Goals
Need to protect patient safety
Need to avoid being sued for malpractice
Pressure to reduce costs
Need to maintain good relationships to other professionals (e.g. surgeons)
Change surgery to outpatient, same day admission
CONFLICT
Maximize preoperative workup and information
CONFLICT
Maximize patient turnover
Copyright © 1991 1992 1993 by Richard I. Cook MD

Figure 7. Conflicting goals in anesthesiology II. Maximizing patient safety and avoiding lawsuits create the need to maximize information about the patient through pre-operative workup. The cost reduction goal provides an incentive for the use of same-day surgery and limits preoperative workup. The anesthetist may be squeezed in this conflict (From Cook, Woods, and McDonald, 1991). Compare this conflict with the one shown in Figure 6, p. 65.

sider further tests that may incur costs, risks to the patient, and a delay of surgery. The cost-reduction goal provides an incentive for a minimal preoperative workup and the use of same-day surgery. This conflicts with the other goals. The anesthetist may be squeezed in this conflict—gathering the additional information, which in the end may not reveal anything important, will cause a delay of surgery and decrease throughput. The delay will affect the day's surgical schedule, the hospital and the surgeon's economic goals, and the anesthesiologists' relationship with the surgeons. The external pressures for highly efficient performance are strongly and increasingly in favor of limiting the preoperative workup of patients and omitting tests that are unlikely to yield important findings. But failing to acquire the information may reduce the ill-defined margin of safety that exists for this patient and contribute to the evolution toward disaster if other factors are present.

For another example, consider the task of enroute flight planning in commercial aviation. Pilots sometimes need to modify their flight plans enroute when conditions change (e.g., weather). Some of the goals that need to be considered in pilot decision making are avoiding passenger discomfort (i.e., avoiding turbulence), minimizing fuel expenditure, and minimizing the difference between the target arrival time and actual arrival time. Depending on the particulars of the actual situation where the crew and dispatchers have to consider modifying the plan, these goals can interact requiring prioritization and tradeoffs. Layton, Smith, and McCoy (1994) created simulated flight situations where goal conflicts arose and studied how the distributed system of dispatchers, pilots, and computer-based advisors attempted to handle these situations.

In another aviation example, an aircraft is de-iced and then enters the queue for takeoff. After the aircraft has been de-iced, the effectiveness of the de-icing agent degrades with time. Delays in the queue may raise the risk of ice accumulation. However, leaving the queue to go back to an area where the plane can be de-iced again will cause additional delays, plus the aircraft will have to re-enter the takeoff queue again. Thus, the organization of activities (where de-icing occurs relative to where queuing occurs in the system) can create conflicts that the practitioners must resolve because they are at the sharp end of the system. The dilemmas may be resolved through conscious effort by specific teams to find ways to balance the competing demands, or practitioners may simply apply standard routines without deliberating on the nature of the conflict. In either case, they may follow strategies that are robust (but still do not guarantee a successful outcome), strategies that are brittle (work well under some conditions but are vulnerable given other eircumstances), or strategies that are very vulnerable to breakdown. Analyses of past disasters frequently find that goal conflicts played a role in the accident evolution. For example, there have been several crashes where, in hindsight, crews accepted delays of too great a duration and ice did contribute to a failed takeoff (Moshansky, 1992; National Transportation Safety Board, 1993).

Goal conflicts can arise from intrinsic characteristies of the field of aetivity (e.g., the Davis-Besse ineident in the nuelear power domain; see NUREG-1154 or the eognitive analysis in Woods and Roth, 1986). An example from anesthesiology is the conflict between the desirability of a high blood pressure to improve cardiac perfusion (oxygen supply to the heart muscle) and a low one to reduce cardiac work (Figure 6, p. 65). Specific actions will depend on details of the context. The appropriate blood pressure target adopted by the anesthetist depends in part on the praetitioner's strategy, the nature of the patient, the kind of surgical procedure, the eireumstanees within the case that may change (e.g., the risk of major bleeding), and the negotiations between different people in the operating room team (e.g., the surgeon who would like the blood pressure kept low to limit the blood loss at the surgical site).

Constraints imposed by organizational or soeial eontext represent another souree of goal competition. Some of the organizational factors producing goals include management polieies, legal liability, regulatory guidelines, and eeonomie factors. Competition between goals generated at the organizational level was an important factor in the breakdown of safety barriers in the system for transporting oil through Prince William Sound that preceded the *Exxon Valdez* disaster (NTSB, 1990). Finally, some of the goals that play a role in practitioner decision making relate to the personal or professional interests of the people in the operational system (e.g., eareer advancement, avoiding eonflicts with other groups).

It should not be thought that the organizational goals are necessarily simply the written policics and procedures of the institution. Indeed, the messages received by practitioners about thc nature of the institution's goals may be quite differcnt from those that managcment acknowledges. Many goals are indirect and implicit. Some of the organizational influences on how practitioners will negotiatc their way through conflicting goals may not bc explicitly statcd or written anywhere. These covcrt factors are espccially insidious bccause they affect behavior and yet are unacknowledgcd. For example, thc Navy sent an implicit but very clear message to its commandcrs by the differential treatment it accorded to the commander of the Stark following that incident (U.S. House of Representativcs Committce on Armed Services, 1987) as opposed to the Vincennes following that incidcnt (U.S. Department of Defense, 1988; Rochlin, 1991).

Expertise consists, in part, of being able to negotiate among interacting goals by selecting or constructing the means to satisfy all sufficiently. But practitioners may fail to deal with goal conflicts adcquately. Some medical practitioners will not follow up hints about somc aspect of the patient's history because to do so would impact the usual practiccs relative to throughput and economic goals. In a specific case, that omission may turn out to be important to the evolution of the incident. Other practitioners will adopt a defensive stancc and ordcr tests for minor indications, even though the yield is low, to bc on the safe sidc. This gencrates increased costs and incurs the wrath of thcir surgical collcagues for the delays thus generated. In either casc, thc naturc of the goals and pressures on the practitioner are seldom made cxplicit and rarcly cxamincd critically.

If those practitioner actions that are shaped by the goal conflict contribute to a bad outcomc in a specific casc, then it is easy for post-incident evaluations to say that a human error occurred—the practition-ers should havc delayed thc surgical procedurc to investigate the hint. The role of thc goal conflict may ncvcr bc noted. Conventional human factors task analyses do not pick up such tradeoffs—task analyses operate at too microscopic a grain of analysis, and how to resolve these conflicts is rarely part of formal job dcscriptions. The strategic dilemmas may not arisc as an explicit conscious decision by an individual so that knowledge acquisition scssions with an expert may not reveal its prcsencc.

To evaluate the behavior of the practitioners involved in an incident, it is important to elucidate the relevant goals, the interactions among these goals, and the factors that influenced criterion setting on how to make tradeoffs in particular situations. The role of these factors is often missed in evaluations of the behavior of practitioners. As a result, it is easy for organizations to produce what appear to be solutions that in fact exacerbate conflict between goals rather than help practitioners handle goal conflicts in context. In part, this occurs because it is difficult for many organizations (particularly in regulated industries) to admit that goal conflicts and tradeoff decisions arise. However distasteful to admit or whatever public relations problems it creates, denying the existence of goal interactions does not make such conflicts disappear and is likely to make them even tougher to handle when they are relevant to a particular incident. As Feynman remarked regarding the Challenger disaster, "For a successful technology, reality must take precedence over public relations, for nature cannot be fooled" (Rogers et al., 1986, Appendix F, p. 5). The difference is that, in human-machine systems, one can sweep the consequences of attempting to fool nature under the rug by labeling the outcome as the consequence of "human error."

Responsibility-Authority Double Binds

Another strategic factor that plays a role in incidents and especially in distributed cognition is responsibility-authority double binds. These are situations in which practitioners have the responsibility for the outcome but lack the authority to take the actions they see as necessary. Regardless of how the practitioners resolve the tradeoff, from hindsight they are vulnerable to charges of and penalties for error. In particular, control at a distance via regimentation ("just follow the procedures") or the introduction of machine cognitive agents who automatically diagnose and plan what they think are the best responses, can undermine the effective authority of the practitioners on the scene. However, these same people may still be responsible (i.e., held accountable both formally and informally) for the bad outcomes. The results on the role of responsibility and authority in distributed cognitive systems are limited but consistent—splitting authority and responsibility appears to have poor consequences for the ability of operational systems to handle variability and surprises that go beyond pre-planned routines (Roth et al., 1987; Hirschhorn, 1993). People tend to pass authority with responsibility together in advisory interactions. Billings (1991) uses this idea as the fundamental premise of his approach to develop a human-centered automation philosophy—*"if people are to remain responsible for safe operation, then they must retain effective authority."* Automation that supplants rather than assists practitioners violates this fundamental premise.

We will summarize two investigations of the effects of responsibility-authority double binds. In one (Hirschhorn, 1993), the study examined the organization's attempts to balance the need to adapt on line to complicating factors (relative to throughput and other goals) with the goal of adhering absolutely strictly to written procedures.

After the Three Mile Island accident, utility managers were encouraged by the Nuclear Regulatory Commission to develop detailed and comprehensive work procedures. The management at a particular nuclear power plant instituted a policy of verbatim compliance with all written procedures. This development occurred in a regulatory climate which believes that absolute adherence to procedures is the means to achieve safe operations and avoid "human error."

However, for the people at the sharp end of the system who actually did things, strictly following the procedures posed great difficulties because (a) the procedures were inevitably incomplete, and sometimes contradictory, and (b) novel circumstances arose that were not anticipated in the work procedures. The policy created a "double bind" because the people would be wrong if they violated a procedure even though it could turn out to be an inadequate procedure, and they would be wrong if they followed a procedure that turned out to be inadequate.

In some situations, if they followed the standard procedures strictly the job would not be accomplished adequately; if they always waited for formal permission to deviate from standard procedures, throughput and productivity would be degraded substantially. If they deviated and it later turned out that there was a problem with what they did (e.g., they did not adapt adequately), it could create re-work or safety or economic problems. The double bind arises because the workers are held responsible for the outcome (the poor job, the lost productivity, or the erroneous adaptation); yet they did not have authority for the work practices because they were expected to comply exactly with the written proeedures. As Hirsehhorn (1993) says,

> Operators, mechanics, and technicians have a good deal of responsibility. As licensed professionals, they can be personally fined for errors but are uncertain of their authority. What freedom of action do they have? What are the responsible for? This gap between the responsibility and authority means that operators and their supervisors feel accountable for events and actions they can neither influence nor control (p. 140).

Workers eoped with the double bind by developing a "covert work system" that involved, as one worker put it, "doing what the boss wanted, not what he said" (Hirschhorn, 1993). There were channels for requesting changes to problems in the procedures, but the process was cumbersome and time-eonsuming. This is not surprising since, if modifications are easy and liberally granted, then it may be seen as undermining the policy of strict procedure-following. Notice how the description of this ease may fit many different domains (e.g., the evolving nature of medical practice).

The design of computer-based systems from a cooperative point of view has also been shown to be a factor that can create authority-responsibility double binds (Woods, 1986; Roth et al., 1987). Consider a traditional artificial intelligence based expert system that solves problems on its own, eommunieating with the operator via a question-and-answer dialogue. In this approach to assistance, the machine is in control of the problem; the system is built on the premise that the expert system can solve the problem on its own if given the correct data. The human's role is to serve as the system's interface to the environment by providing it with the data to solve the problem. If the human practitioners are to do any problem solving, it is carried out in parallel, independent of the interaction with the intelligent system. Results indicate that this prosthesis form of interaction between human and intelligent system is very brittle in the face of complicating factors (Roth et al., 1987). Again, the need to cope with novel situations, adapt to special conditions or contexts, recover from errors in following the instructions, or cope with bugs in the intelligcnt systcm itself requires a robust cognitive system that can detect and recover from error.

The crux of the problem in this form of cooperation is that the practitioner has responsibility for thc outcome of the diagnosis, but the machine expert has takcn over effective authority through control of the problem-solving process. Note the double bind that practitioners are left in, even if thc machine's solution is disguised as only "advice" (Woods, 1986; Roth et al., 1987; Woods et al., 1991). In hindsight, practitioners would be wrong if thcy failed to follow the machine's solution and it turned out to be corrcct, cven though a machine can err in some cases. Thcy would be wrong if they followed the machine's "advice" in those cases where it turned out the machinc's solution was inadequate. They also would be wrong if they were correctly suspicious of the machine's proposed solution, but failed to handle the situation successfully through their own diagnosis or planning efforts (see Chapter 6 on how knowledge of outcome biascs evaluation of process). The practitioners in the evolving problem do not have the advantage of knowledge of eventual outcome; they must evaluate the data at hand including thc unccrtainties and risks.

Instructions, howcver claboratc, rcgardlcss of medium (paper- or computer-based), and regardlcss of whcther thc guidance is complctcly pre-packaged or partially gcneratcd "on-the-fly" by an cxpert systcm, are inherently *brittle* whcn followcd rotely. Brittleness means that it is difficult to build in mechanisms that cope with novel situations, adapt to special conditions or contcxts, or recover from errors in following the instructions or bugs in thc instructions thcmselves (e.g., Brown, Moran, and Williams, 1982; Woods et al., 1987; Herry, 1987). As Suchman (1987) has put it, "plans are [only] rcsources for action."

When people usc guidancc to solve problcms, erroneous actions fall into one of two gencral catcgories (Woods et al., 1987):

- rote rule following pcrsists in thc facc of changing circumstances that demand adaptation,
- the people correctly recognize that standard responses are inadequatc to mcet opcrational goals givcn thc actual circumstances, but fail to adapt thc pre-planned guidance effcctively (e.g., missing a side effect).

For example, studies of nuclear power plant operators responding to simulated and to actual accident conditions with paper-based instructions found that operator performance problems fell into one or the other of the above categories (Woods et al., 1987). If practitioners (those who must do something) are held accountable for both kinds of "error"—those where they continue to rotely follow the rules in situations that demand adaptation and those where they erroneously adapt—then the practitioners are trapped in a double bind.

Following instructions requires actively filling in gaps based on an understanding of the goals to be achieved and the structural and functional relationships between objects referred to in the instructions. For example, Smith and Goodman (1984) found that more execution errors arose in assembling an electrical circuit when the instructions consisted exclusively of a linear sequence of steps to be executed, than when explanatory material related the instruction steps to the structure and function of the device. Successful problem solving requires more than rote instruction following; it requires understanding how the various instructions work together to produce intended effects in the evolving problem context.

While some of the problems in instruction following can be eliminated by more carefully worded, detailed, and explicit descriptions of requests, this approach has limitations. Even if, in principle, it were possible to identify all sources of ambiguity and craft detailed wording to avoid them, in practice the resources required for such extensive fine tuning are rarely available. Furthermore, the kinds of literal elaborate statements that would need to be developed to deal with exceptional situations are likely to obstruct the comprehension and execution of instructions in the more typical and straightforward cases (for example, in a recent aviation incident the crew used about 26 different procedures; see Chapter 6 for more on this incident).

Attempts to eliminate all sources of ambiguity are fundamentally misguided. Examination of language use in human-human communication reveals that language is inherently underspecified; it requires the listener (or reader) to fill in gaps based on world knowledge, and to assess and act on the speaker's (writer's) intended goals rather than his literal requests (Suchman, 1987). Second, a fundamental competency in human-human communication is the detection and repair of communication breakdowns (Suchman, 1987). Again, error recovery is a key process. In part, this occurs because people build up a shared frame of reference about the state of the world and about what are meaningful activities for the current context.

Whenever organizational change or technology change occurs, it is important to recognize that these changes can sharpen or lessen the strategic dilemmas that arisc in operations and change how practitioners negotiate tradeoffs in context. In designing high-reliability systems for fields of activity with high inherent variability, one cannot rely just on rotely followed pre-planned routines (even with a tremendous investment in the system for producing and changing the routines). Nor can one rely just on the adaptive intelligence of people (even with a tremendous investment in the people in the system). Distributed cognitive system design should instead focus on how to coordinate preplanned routines with the demands for adaptation inherent in complex fields of activity (Woods, 1990a). The history of mission control during the Apollo project is a good illustration of the coordination of these two types of activity in pacc with the varying rhythms of the field of practice (e.g., Murray and Cox, 1989).

Local Rationality

Human (and real machine) problem-solvers possess finite capabilities. They cannot anticipate and consider all the possible alternatives and information that may be relevant in complex problems. Simon codified this concept in his 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 (Simon, 1957, p. 198).

People's behavior is consistent with Newell's principle of rationality—that is, they use knowledge to pursue thcir goals (Newell, 1982). But there are bounds to the data that they pay attention to, the knowledge that they possess, the knowledge that they activate in a particular context, and there may be multiple goals which conflict. In other words, people's behavior is rational, though possibly erroneous, when viewed from the locality of their knowledge, attentional focus, and strategic tradeoffs. For the context of error, we will refer to the concept that human rationality is limited or bounded as "local" rationality (cf. also, Reason, 1990).

A consequence of the perspective of local rationality is that people construct simplified but useful models; they develop and adopt simplified but useful techniques, that is, people "satisfice" (Simon, 1969). The decision procedures that humans construct are sensible given the constraints that they necessarily operate under, though these might not be sensible if the constraints are removed (March, 1978). In some situations these decision procedures may lead to erroneous assessments or actions. This points to the notion of error as a mismatch between problem demands and the human's resources (see Rasmussen, 1986), as Figure 1 (p. 21) tries to illustrate in part.

The notion of local rationality does not imply that humans are poor problem solvers or decision makers that need to be replaced by automation based on "optimal" models. The point is that, in actuality, all cognitive systems—human, machine, or distributed—are limited or constrained. For'machine cognitive systems this idea has been carried forward under the label of computational complexity. Computational processes require resources such as memory capacity and operations performed per unit of time. Some processes are computationally intractable, that is, they require exponentially increasing resources as problem size increases. For example, Oaksford and Chater (1992) point out that Bayesian inference may make exponentially increasing demands on computational resources even when problems involve moderate amounts of information. Since all cognitive systems are limited resource processors, the processes involved in risky and time-pressured decision making cannot be based upon resource unconstrained procedures, however optimal they appear on other grounds (Klein et al., 1993). This means that the only rationality to which we can aspire, as individual or organizational decision makers, "is one bounded by our limited computational resources" (Oaksford and Chater, 1992). This rationality is also local in the sense that it is context bound, that is, it is exercised relative to the complexity of the environment in which the particular cognitive system functions.

The important point here is that it takes effort (which consumes limited computational resources) to seek out evidence, to interpret it (as relevant), and to assimilate it with other evidence. Evidence may come in over time, over many noisy channels. The process may yield information only in response to diagnostic interventions. Time pressure, which compels action (or the de facto decision not to act), makes it impossible to wait for all evidence to accrue. Multiple goals may be relevant, not all of which are consistent. It may not be clear which goals are the most important ones to focus on at any one particular moment in time. Human problem solvers cannot handle all the potentially relevant information, cannot activate and hold in mind all of the relevant knowledge, and cannot entertain all potentially relevant trains of thought. Hence, rationality must be local—attending to only a subset of the possible evidence or knowledge that could be, in principle, relevant to the problem.

The Implications of Local Rationality for Studying Error

One implication of local rationality is that normative procedures based on an ideal or perfect rationality do not make sense in evaluating cognitive systems. Rather, we need to find out what are robust, effective strategies given the resources of the problem solvers (i.e., their strategies, the nature of their working memory and attention, long-term memory organization, retrieval processes, etc.), and the demands of the problem-solving situation (time pressure, conflicting goals, uncertainty, etc.). Error analyses should be based on investigating demand-resource relationships and mismatches (Rasmussen, 1986).[15]

Human decision makers generally choose strategies that are relatively efficient in terms of effort and accuracy as task and context demands are varied (Payne et al., 1988; 1990). Procedures that seem "normative" for one situation (non-time constrained) may be severely limited in another problem context (time constrained). In developing standards by which to judge what are effective cognitive processes, one must understand problem solving in context, not in "the abstract."

[15] As Simon (1969) points out, "It is wrong, in short, in ignoring the principle of bounded rationality, in seeking to erect a theory of human choice on the unrealistic assumptions of virtual omniscience and unlimited computational power" (p. 202).

For example, if one were designing a decision aid that incorporated Bayesian inference, one would need to understand the context in which the joint human-machine system functions including such factors as noisy data or time pressure. Fischhoff and Beyth-Marom (1983) point out that applying Bayesian inference in actuality (as opposed to theory) has the following error possibilities: formulation of wrong hypotheses, not correctly eliciting the beliefs and values that need to be incorporated into the decision analysis, estimating or observing prior probabilities and likelihood functions incorrectly, using a wrong aggregation rule or applying the right one incorrectly.

In other words, cognitive strategies represent tradeoffs across a variety of dimensions including accuracy, effort, robustness, risks of different bad outcomes, or the chances for gain from different possible good outcomes. Effective problem-solving strategies are situation specific to some extent; what works well in one case will not necessarily be successful in another. Furthermore, appropriate strategies may change as an incident evolves, e.g., effective monitoring strategies to detect the initial occurrence of a fault (given normal operations as a background) may be very different from search strategies during a diagnostic phase (Moray, 1984). In understanding these tradeoffs relative to problem demands we can begin to see the idea that expertise and error spring from the same sources.

The assumption of local rationality—people are doing reasonable things given their knowledge, their objectives, their point of view and limited resources, e.g., time or workload—points towards a form of error analysis that consists of tracing the problem-solving process to identify points where limited knowledge and limited processing lead to breakdowns. This perspective implies that one must consider what features of domain incidents and situations increase problem demands.

Exploring Demand-Resource Mismatches

The local-rationality assumption and the demand-resource mismatch view of erroneous actions suggest a strategy to predict how people can develop erroneous intentions to act. One can model a cognitive system in a particular task context by tracing the problem-solving process to identify points where limited knowledge and processing resources can lead to breakdowns, given the demands of the problem (Woods, 1990a).

A cognitive simulation can be an excellent tool for exploring different concepts about limits on cognitive processing (e.g., attentional bottlenecks or limited knowledge activation) in relation to the demands imposed by different kinds of problems that can occur in the field of practice (Woods, 1990a; Woods and Roth, in press). Cognitive simulation is a technique invented by Newell and Simon (Newell and Simon, 1963; Simon, 1969; Newell and Simon, 1972) in which information-processing concepts about human cognitive activities are expressed as an executable computer program, usually through symbolic processing techniques (see Johnson et al., 1988; Roth et al., 1992; or Johnson et al., 1992, for examples using symbolic processing techniques; cf., also Axelrod, 1984, or Payne et al., 1990, for examples using conventional programming techniques).

The cognitive simulation can be constructed to allow the investigator to vary the knowledge resources and processing characteristics of a limited resource computer problem-solver and observe the behavior of the computer problem-solver in different simulated domain scenarios. This strategy depends on mapping the cognitive demands imposed by the domain in question that any intelligent but *limited-resource* problem-solving agent or set of agents would have to deal with. The demands include the nature of domain incidents, how they are manifested through observable data to the operational staff, and how they evolve over time. Then, one can embody this model of the problem-solving environment as a limited-resource, symbolic-processing, problem-solving system.

When stimulated with input from a scenario (a temporal stream of the data about the state of the monitored process that is, or could be, available during an unfolding incident), the computer simulation can be made so that it carries out cognitive functions such as monitoring changes in process state, or diagnosis of underlying faults. For example, one of these cognitive simulations (Roth et al., 1992) performs some of the cognitive functions involved in dynamic fault management: it monitors and tracks changes in process state, forms expectancies based on an assessment of what influences are currently acting on the monitored process, identifies abnormal and unexpected process behaviors, builds and revises its situation assessment about influence patterns, formulates hypotheses to account for unexplained process behavior, and formulates intentions to act based on its situation assessment. The simulation is a representation or realization of a set of concepts; it is a way to formalize the concepts so that one can explore and investigate the explanatory power of the concepts in a wide range of circumstances.

A successful cognitive simulation provides a compelling demonstration of the cognitive work required to operate successfully in the problem-solving environment. Using the simulation can help reveal how locally rational processes govern the expression of both expertise and error. A variety of cognitive simulations are under development to try to explore the complexities of human-machine systems solving complex and dynamic problems (e.g., Corker, Davis, Papazian, and Pew, 1986; Cacciabue, Decortis, Drozdowicz, Masson, and Nordvik, 1992; Roth et al., 1992; Johnson et al., 1992).

In effect, with this technique one is measuring the difficulty or complexity posed by a domain incident, given some set of resources, by running the incident through the cognitive simulation (Kieras and Polson, 1985; Woods et al., 1990). In other words, the cognitive simulation supports a translation from the language of the individual field of practice to the language of cognitive activities. What data needs to be gathered and integrated, what knowledge is required to be used, and how is it activated and brought to bear in the cognitive activities involved in solving dynamic problems? In effect, the cognitive simulation yields a description of the information flow and knowledge activation required to handle domain incidents. One can investigate how changes in the incident (e.g., obscuring evidence, introducing another failure) affect the difficulty of the problem for a given set of knowledge resources. Conversely, one can investigate how changes in the knowledge resources (e.g., improved mental models of device function) or information available (e.g., integrated information displays) can affect performance.

For example, consider a textbook nuclear power incident in a pressurized light water reactor—a steam generator tube rupture where primary system cooling water flows through a break in the heat exchanger into the secondary side of the steam generator. Now let us consider a variant on this incident where the radiation monitors on the secondary side of the plant are all disabled or unavailable in some way (e.g., a loss of electric power just prior to the start of the break will, among other things, cut off the flow of water or air that would carry radiation to the sensing devices). This combination of circumstances results in no indications of the presence of radiation in the secondary part of the plant. The question is how difficult are the problems posed by these incidents for practitioners (Woods et al., 1990)?

The base incident is a *textbook* case in that there is a highly certain and highly salient cue that indicates the presence of a tube rupture condition (radiation in the secondary side of the plant). This cue strongly evokes the sole hypothesis of a tube rupture (except for the possibility of sensor failure). The diagnostic search activities that follow the initial hypothesis will reveal plant behaviors consistent with this hypothesis. Thus, incident diagnosis should occur highly reliably and early in the sequence of events.

Now consider what happens in the variant where the radiation indications do not occur (one kind of complicating factor). From a problem-solving point of view the incident is a "loss of leading indicator" incident—a bighly certain indicator of a diagnostic category is missing. Given the absence of secondary radiation signals, there is a much larger set of hypotheses that is consistent with the initial set of abnormal plant behaviors (low level, low pressure), and which should be explored during diagnosis. The results of the initial diagnostic search will eliminate some possibilities. In particular, the evidence will be consistent with a break, but which type will not be conclusively established (although the strongest candidate is the loss of primary coolant category). The question then is how sensitive is the crew to the remaining evidence which signals that a tube rupture is present, i.e., abnormally high water level in one steam generator. Since it takes some time for this evidence to be detectable by any agent given the natural evolution of the incident and the current displays of information, the diagnosis of a steam generator tube rupture will take much longer than in the textbook case. Furthermore, high workload, or some knowledge (or processing) bugs may lead the human problem solvers to miss or misinterpret the evidence when it is observable.

Data from both actual steam generator tube rupture accidents and from simulated ones run with experienced crews (Woods, Wise, and Hanes, 1982) show exactly this pattern of results. Furthermore, Woods et al. (1990) show how a cognitive simulation computer program can be used to determine the same results analytically.

Cognitive simulations provide one vehicle to explore the temporal dynamics of cognitive systems in relation to the temporal characteristics of incidents. In dynamic environments, data come in over time, change, or become obscured. Faults propagate chains of disturbances that evolve and spread through the system. Counteracting influences are injected by automated systems and by practitioners to preserve system integrity, to generate diagnostic information, and to correct faults. Information is based on change, events (behavior over time), and the response to interventions. *Static models are incapable of expressing the complexity of cognitive functioning in dynamic environments*—the interaction of data-driven and knowledge-driven reasoning, the role of interrupts in the control of attentional focus, the scheduling of cognitive activities as workload bottlenecks emerge, and the interaction of intervention and feedback on process response.

It is very difficult to appreciate the complexities of the situation faced by practitioners and the set of cognitive functions that is required to handle domain events without some mechanism to explore the dynamic interplay of problem evolution and cognitive processing. In the development of one cognitive simulation (Woods et al., 1990) it became clear that to follow and control dynamic events, it was necessary to use a computer program with elaborate mechanisms (e.g., qualitative reasoning):

- for tracking interactions among multiple influences acting on the monitored process over time;
- for tracking when automation would or should activate or inactivate various control systems;
- for projecting the impact of a state change on future process behavior to create temporal expectations or reminders to check whether the expected behavior is observed, or, more importantly, not observed.

Interestingly, in one study using this specific tool (Roth et al., 1992), the factors that made the class of incidents difficult could be found only through an analysis of the dynamics of the incident in relation to the dynamics of the joint cognitive system.

Did The Practitioners Commit Errors?

Given the discussion of cognitive factors (knowledge, attentional dynamics, and stratcgic dilcmmas) and of local rationality, let us go back to the three exemplar incidents dcscribed earlier in this chapter and re-examine them from the perspective of the question: What is human error?

These three incidents are not remarkable or unusual in their own field of activity (urban, tertiary care hospitals) or in other complex domains. In each incident, human performance is closely tied to system performance and to eventual outcomc, although the performance of the practitioners is not the sole determinant of outcome.[16] The incidents and the analysis of human performance that they prompt (including the role of latent failures in incidents) may make us change our notion of what constitutes a human error.

Arguably, the performance in each cxemplar incident is flawed. In retrospect, things can be identified that might have been done differently and which would havc forestalled or minimized the incident or its effect. In the myocardial infarction incidcnt (#1), intravascular volume was misassessed and treatment for scveral simultaneous problems was poorly coordinated. In the hypotension incident (#2), the device setup by practitioners contributcd to the initial fault. The practitioners were also unable to diagnose the fault until well after its effects had cascaded into a near crisis. In the scheduling incident (#3), a practitioner violated policy. She chose one path to meet certain demands, but simultaneously exposed the larger systcm to a rare but important variety of failure. In some sense, each of thc exemplar incidents constitutes an example of human error. Note, however, that each incident also demonstrates the complexity of the situations confronting practitioners and the way in which practitioners adjust their behavior to adapt to the unusual, difficult, and novel aspects of individual situations.

The hypotension incident (#2) particularly demonstrates the resiliency of human performance in an evolving incident. During this inci-

[16] For example, the myocardial infarction following lhe events of incident #1 may well have happened irrespective of any actions taken by practilioners. That patient was likely lo have an infarction, and il is not possible lo say if lhe aneslhelisl's actions caused the infarction.

dent the physicians engaged successfully in disturbance management (see Woods, in press-b) to cope with the consequences of a fault. The physicians were unable to identify the exact source of the incident until after the consequences of the fault had ended. However, they were able to characterize the kind of disturbance present and to respond constructively in the face of time pressure. They successfully treated the consequences of the fault to preserve the patient's life. They were able to avoid becoming fixated on pursuing what was the "cause" of the trouble. In contrast, another study of anesthesiologist cognitive activities, this time in simulated difficult cases (Schwid and O'Donnell, 1992), found problems in disturbance management where about one-third of the physicians undertreated a significant disturbance in patient physiology (hypotension) while they over-focused on diagnostic search for the source of the disturbance.

The practitioner was also busy during the myocardial infarction incident, although in this instance the focus was primarily on producing better oxygenation of the blood and control of the blood pressure and not on correcting the intravascular volume. These efforts were significant and, in part, successful. In both incidents #1 and #2, attention is drawn to the practitioner performance by the outcome.

In retrospect some would describe aspects of these incidents as human error. The high urine output with high blood glucose and prior administration of furosemide *should* have prompted the consideration of low (rather than high) intravascular volume. The infusion devices *should* have been set up correctly, despite the complicated set of steps involved. The diagnosis of hypotension *should* have included a closer examination of the infusion devices and their associated bags of fluid, despite the extremely poor device feedback. Each of these conclusions, however, depends on knowledge of the outcome; each conclusion suffers from hindsight bias. To say that something *should have been obvious*, when it manifestly was not, may reveal more about our ignorance of the demands and activities of this complex world than it does about the performance of its practitioners. It is possible to generate lists of "shoulds" for practitioners in large systems but these lists quickly become unwieldy and, in any case, will tend to focus only on the most salient failures from the most recent accident.

The scheduling incident (#3) is somewhat different. In that incident it is clear how knowledge of the outcome biases evaluations of the practitioner performance. Is there a human error in Incident #3? If a trauma case had occurred in this interval where all the resources had been committed to other cases, would her decision then be considered an error? On the other hand, if she had delayed the start of some other case to be prepared for a possible trauma case that never happened and the delay contributed to some complication for that patient, would her decision then be considered an error?

Uncovering what is behind each of these incidents reveals the label "human error" as a judgment made in hindsight. As these incidents suggest, human performance is as complex and varied as the domain in which it is exercised. *Credible evaluations of human performance must be able to account for all of the complexity that confronts practitioners and the strategies they adopt to cope with that complexity.* The term "human error" should not represent the concluding point but rather the starting point for studies of accident evolution in large systems.

The N-Tuple Bind

The three incidents described in this chapter are exemplars for the different cognitive demands encountered by practitioners who work at the sharp end of large, complex systems, including anesthetists, aircraft pilots, nuclear power plant operators, and others. Each category of cognitive issue (knowledge factors, attentional dynamics, strategic factors, and local rationality) plays a role in the conduct of practitioners and hence plays a role in the genesis and evolution of incidents. The division of cognitive issues into these categories provides a tool for analysis of human performance in complex domains. The categories are united, however, in their emphasis on the conflicts present in the domain. The conflicts exist at different levels and have different implications, but the analysis of incidents depends in large part on developing an explicit description of the conflicts and the way in which the practitioners deal with them. (See Table 1, p. 52)

Together the conflicts produce a situation for the practitioner that appears to be a maze of potential pitfalls. This combination of pressures and goals in the work environment is what we call *the n-tuple bind*[17] (Cook and Woods,1994). The practitioner is confronted with the need to follow a single course of action from myriad possible courses. How to proceed is constrained by both the technical characteristics of the domain and the need to satisfy the "correct" set of goals at a given moment chosen from the many potentially relevant ones. This is an example of an over-constrained problem, one in which it is impossible to maximize the function or work product on all dimensions simultaneously. Unlike simple laboratory worlds with a best choice, real complex systems intrinsically contain conflicts that must be resolved by the practitioners at the sharp end. Retrospective critiques of the choices made in system operation will always be informed by hindsight (see Chapter 6). For example, if the choice is between obtaining more information about cardiac function or proceeding directly to surgery with a patient who has soft signs of cardiac disease, the outcome will be a potent determinant of the "correctness" of the decision. Proceeding with undetected cardiac disease may lead to a bad outcome (although this is by no means certain), but obtaining the data may yield normal results, cost money, "waste" time, and incur the ire of the surgeon. Possessing knowledge of the outcome trivializes the situation confronting the practitioner and makes the "correct" choice seem crystal clear.

This *n-tuple bind* is most easily seen in Incident #3 where strategic factors dominate. The practitioner has limited resources and multiple demands for them. There are many sources of uncertainty. How long will the *in vitro* fertilization take? It should be a short case but it may not be. The exploratory laparotomy may be either simple or complex. With anesthetists of different skill levels, whom should she send to the remote location where that case will take place? Arterial reconstruction patients usually have associated heart disease, and the case can be demanding. Should she commit the most senior anesthetist to that case? Such cases are also usually long and committing the most experienced anesthetist will tie up that resource for a long time. What is the likelihood that a trauma case will come during the time when all

[17] This term derives from the mathematical concept of a series of numbers required to define an arbitrary point in an *n*-dimensional space. The metaphor here is one of a collection of factors that occur simultaneously within a large range of dimensions, i.e., an extension of the notion of a *double bind*.

the cases will be going on simultaneously (about an hour)? There are demands from several surgeons for their case to be the next to start. Which case is the most medically important one? The general rule is that an anesthetist has to be available for a trauma; she is herself an anesthetist and could step in but this would leave no qualified individual to go to cardiac arrests in the hospital or to the emergency room. Is it desirable to commit all the resources now and get all of the pending cases completed so as to free up the people for other cases that are likely to follow?

It is not possible to measure accurately the likelihood of the various possible events that she considers. As in many such situations in medicine and elsewhere, she is attempting to strike a balance between common but lower consequence problems and rare but higher consequence ones. *Ex post facto* observers may view her actions as either positive or negative. On the one hand, her actions are decisive and result in rapid completion of the urgent cases. On the other hand, she has produced a situation where emergent cases may be delayed. The outcome influences how the situation is viewed in retrospect.

A critique often advanced in such situations is that "safety" should outweigh all other factors and be used to differentiate between options. Such a critique is usually made by naive individuals or administrative personnel not involved in the scene. *Safety is not a concrete entity, and the argument that one should always choose the safest path misrepresents the dilemmas that confront the practitioner.* The safest anesthetic is the one that is not given; the safest airplane is the one that never leaves the ground. All large, complex systems have intrinsic risks and hazards that must be incurred in order to perform their functions, and all such systems have had failures. The investigation of such failures and the attribution and effect by retrospective reviewers are discussed in Chapter 6.

What System Fails? Organizational and Cognitive Systems Perspectives

Figure 1 (p. 21) is deliberately designed to represent the entire ensemble of operational system and organizational context. The ensemble is represented in Figure 1 through a single shape (hence, the small icon used as a legend in the upper right corner). Organizational factors only operate *through* the constraints they impose on how the cognitive system at the sharp end adapts to meet the demands of the field of activity. However, one cannot understand or model a distributed cognitive system without reference to the larger organizational context in which it is embedded.

A purely ergonomic approach errs if it examines human-machine interaction independent of the organizational context.[18] Personally, we do not know how to study or model the "sharp end" without also developing an understanding of the organizational context in which these activities take place and which shapes them[19] (cf. for examples, Cook, Woods, and McDonald, 1991 and Moll van Charante et al., 1993 for how an understanding of the larger context was a part of studying groups of practitioners at the sharp end). But a purely organizational approach will miss or undervalue the adaptive response of the operational system to organizational constraints and to the demands of the field of activity. Organizational factors influence safety and risk *through* their impact on the distrihuted cognitive system at the sharp end. In the final analysis, it is sets of practitioners at the sharp end who confront directly and tangibly the possibility of negative outcomes. Analyses that disemhody the reality of personally confronting the consequences of decisions and actions miss a critical component of the incuhation and development of incidents.

Some would then see that the solution is to layer several different successive analyses centered around the individual, around the group or team, and around organizational processes. If this "onion skin" view is a simple accretion of independent perspectives, it misses the dynamic inter-relationships. But Hutchins (in press), Hollnagel (1993), and others propose that a synthesis is possihle if one sees (a) that individuals are always embedded in larger distributed systems and organizational contexts *and* (h) that an expanded cognitive language provides a tool for studies and models of the interactions within and among

[18] A purely ergonomic approach also errs if it only sees the interaction of individuals with particular devices in isolated tasks.

[19] Similarly, to study people doing cognitive work requires studying and modeling the field of activity in which they work in terms of the demands imposed on cognitive systems in general (e.g., Roth et al., 1992).

these "systems." The cognitive system synthesis provides a way to examine the linkages across problem demands and practitioner strategies, across practitioner activities and external artifacts or tools, across individuals and the distributed system in which they function, across operational systems and organizational contexts (e.g., Woods and Sarter, 1993). Again, operations is as much a distributed multi-agent system as management or design; management and design are just as much human activities as operations.

One can see the relationship of "sharp" and "blunt" ends as a kind of figure-ground. Strategies and activities at the "sharp" end stand out in relief against the ground of organizational context. When we shift our focal point to the "blunt end" itself, do we then study organizations directly? Operational systems are not restricted to personnel who manage a dynamic process—anesthesiologists, power plant operators, pilots, space mission controllers, etc. Designers[20] or managers are practitioners as well, who also function in an evolving world under various constraints. In other words, when we shift our focus from the operational system at the literal sharp end to designers or managers who work at the blunt end, we are now studying a different operational system that functions in a different organizational context. Recursively, what was the blunt end of one system becomes the sharp end of a new system. Figure 8 (p. 107) depicts how what is seen as field of activity, operational system, or organizational context, changes relative to one's focal point.

Design Failures

The system of people and artifacts involved in design activities also can be modeled as a cognitive system embedded in a larger organizational context (Figure 8, p. 107). This design system provides technological artifacts to the workplace where other people operate and manage some kind of process. However, design activities are about more than the artifacts themselves; they also, explicitly or implicitly, reshape the underlying field of practice. The design system, in a locally

[20] By designers we mean all of the people involved in the development and fielding of new technology. The kinds of people vary across many different engineering and non-engineering specialties. At various stages in the development and fielding, people will play different roles, e.g., people from the same specialty function differently early in concept identification as compared to late in integrating a "final" system into the field.

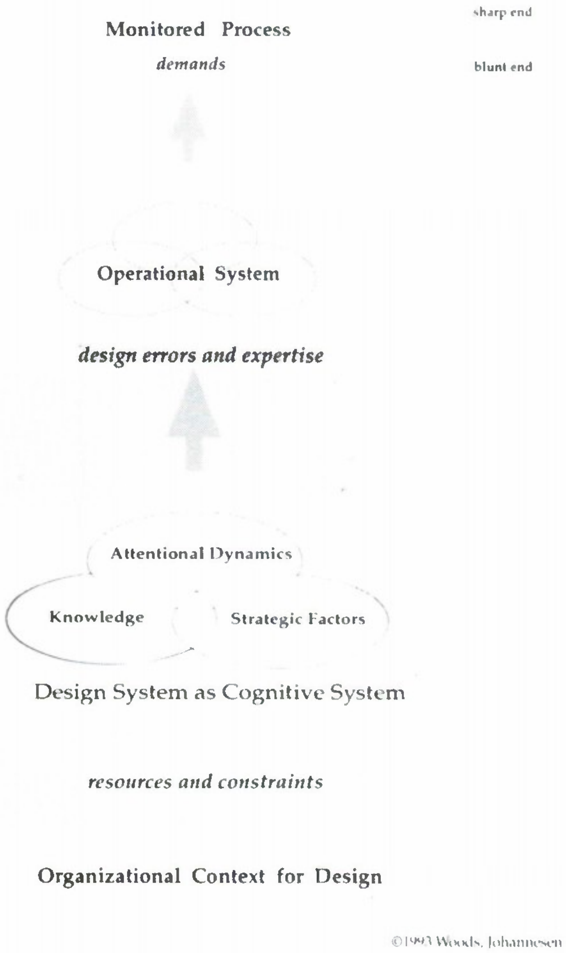


Figure 8. The system of people and artifacts involved in design activities also can be modeled as a cognitive system embedded in a larger organizational context.

rational way, attempts to develop technology to influence this underlying operational world hased on models about the relationship of technological artifacts and human performance, based on feedback about the effects of new technology on human performance in different settings, and based on multiple goals. Design systems exist in a larger organizational context that relates to economic, regulatory, and other factors that constrain the resources and frame the dilemmas faced by designers.

Mismatches between demands and resources can lead to "failures" in design where characteristics of new technological artifacts create or contribute to new forms of error and failure in the operational setting (such as mode errors; see Chapter 5). Design failures include introducing artifacts that, in actual practice, create new burdens and complexities (knowledge, attentional, or strategic). This kind of clumsy use of technology represents "design error" because it creates conditions that predictably lead to "operator error" under some circumstances. In other words, characteristics of technological artifacts function as a kind of latent failure – a condition present in the system that can lead to failure if other triggering and potentiating factors are present.

How do we establish that a design is erroneous from a cognitive system or human performance point of view? The criteria for calling a design erroneous are based on empirical results from investigations and experience with people who use such systems in various fields of practice. When these studies show how characeristics of computer-hased technology contribute to the potential for error and failure, we develop new knowledge about the relationship between technology and human performance. For example, studies have found that devices which provide poor feedhack to users about device state and behavior and which have multiple modes predictably produce mode errors (i.e., an action appropriate to one mode when the device is actually in another mode). Mode errors sometimes contribute to incidents and accidents if other potentiating factors are present. Thus, designing devices with the characteristics that encourage mode error is a kind of "design error." This perspective is detailed in Chapter 5 (cf., Figure 9, p. 125).

It is important to remember that the label "error" – even the labels "design error" or "management error" – should be a starting point, not the stopping rule, for investigations. Remember that errors, even design errors, are symptoms, not causes. What are the factors that govern the expression of design expertise and error? Design failures, when recognized as such, are governed by knowledge factors, attentional dynamics, or strategic factors. Designers may create devices with embedded error traps because of their buggy knowledge about the consequences for human cognition and behavior or a lack of knowledge about how to use technological powers to truly support cognitive activities. Researchers of human-computer cooperation may not have developed the necessary knowledge base.

Designers also may proliferate modes because of workload or attentional factors. Today most new systems are justified in part because of claimed or putative benefits on human performance. However, do the designers gather feedback about the actual effects of technology change on the operational systems in question? Designers may be fixated on their model of how technological artifacts should influence human performance rather than attuned to feedback about the actual reverberations of technology change. Attentional and workload factors may push designers away from gathering such feedback in their particular case or transferring knowledge developed from other cases.

Another factor that may lead designers to use technological possibilities clumsily is the method in which they make tradeoffs on strategic dilemmas. The cost of clumsy use of technology may not be appreciated (or may be rationalized away) and may not play a role in tradeoffs about how to prioritize resource investments during development. Given limited resources, time horizons, and the many different constraints to be satisfied for a successful final system, it is easy to see how designers can provide general purpose flexibility and defer responsibility for operability to trainers and practitioners.

Characteristics of a computer-based device will shape the cognition and behavior of practitioners regardless of whether designers explicitly attend to these factors (Woods, in press-b). For example, various design-shaping properties of the computer medium encourage certain patterns (e.g., they make it easy to proliferate modes or to provide ineffective feedback) unless designers explicitly invest their energy and imagination in developing alternative ways to use the power of new graphic and data processing technologies. Properties of the development environments also make it easy for designers to use technology clumsily. Woods et al. (1991) found that many design choices were made implicitly, based on what was easy to accomplish with a particular prototyping tool or softwarc development environment.

All these factors may operate togethcr in various degrces to produce the avalanche of clumsy computer-based devices that inundate beleaguered practitioners (Norman,1988). The relationship between computer technology and error is explored in the next chapter.

Interim Summary

Human performance in large systems and the failures of these systems are closely linked. The demands that operating large, complex systems place on human performance are mostly cognitive. The difference between expert and inexpert human performance depends on timely and appropriate action that in turn is shaped by knowledge factors, attentional dynamics, and stratcgic factors in relation to the demands of the problems and given the constraints imposed by organizational context. Our brief examination of the cognitive factors involved behind the label of human error has demonstrated that human performance is complex in proportion to the complexity of the domain itself. Analyses of the human role, especially those that take place after an incident or accident, must provide a satisfactory account of that complexity and its impact on the distributed cognitive system at the sharp end. The schema represented in Figure 1 (p. 21) can provide a framework for laying out the issues confronting practitioners at the sharp end.

5

THE IMPACT OF THE CLUMSY USE OF COMPUTER TECHNOLOGY ON COGNITION, BEHAVIOR, AND ERROR

Introduction

This chapter describes several classic deficiencies in computerized devices and how these negatively influence practitioner cognition and behavior. Characteristics of computerized devices that shape cognition and behavior in ways that increase the potential for error are one type of latent failure that can contribute to incident evolution. The presence of these characteristics, in effect, represents a failure of design in terms of operability (i.e., a kind of design "error"). We will show why these device characteristics are in error, and we will show how the failure to design for effective human-computer cooperation increases the risk of bad outcomes.

Technology Change Transforms Operational and Cognitive Systems

There are several possible motivations for studying an operational system in relation to the potential for error and failure. The occurrence of an accident or a near miss is a typical trigger for an investigation. Cumulated evidence from incident data bases may also provide a trigger to investigate "human error."

Another important trigger for examining the potential for system breakdown is at points of major technology change. Technology change is an intervention into an ongoing field of activity (Winograd and Flores, 1987; Flores, Graves, Hartfield, and Winograd, 1988). When developing and introducing new technology, one should realize that the technology change represents new ways of doing things; it does not preserve the old ways with the simple substitution of one medium for another (e.g., paper for computer-based).

Technological change is, in general, transforming the workplace through the introduction and spread of new computer-based systems. First, ubiquitous computerization has tremendously advanced our ability to collect, transmit, and transform data. In all areas of human endeavor, we are bombarded with computer-processed data, especially when anomalies occur (Woods, in press-b). But our ability to digest and interpret data has failed to keep pace with our abilities to generate and manipulate greater and greater amounts of data. Thus, we are plagued by data overload.

Second, user interface technology has allowed us to concentrate this expanding field of data into one physical platform, typically a single visual display unit (VDU). Users are provided with increased degrees of flexibility for data handling and presentation in the computer interface through window management and different ways to display data. The technology provides the capability to generate tremendous networks of computer displays as a kind of virtual perceptual field viewable through the narrow aperture of the VDU. These changes affect the cognitive demands and processes associated with extracting meaning from large fields of data (Woods, 1991; in press-b).

Third, heuristic and algorithmic technologies expand the range of subtasks and cognitive activities that can be automated. Automated resources can, in principle, offload practitioner tasks. Computerized systems can be developed that assess or diagnose the situation at hand, alerting practitioners to various concerns and advising practitioners on possible responses. These "intelligent" machines create joint cognitive systems that distribute cognitive work across multiple agents (Woods, 1986; Roth et al., 1987; Hutchins, 1990). Automated and intelligent agents change the composition of the team and shift the human's role within that cooperative ensemble.

Fourth, computerization and automation integrate or couple more closely together different parts of the system. Increasing the coupling within a system has many effects on the kinds of cognitive demands to be met by the operational system. With higher coupling, failures produce more side effects. A failure is more likely to produce a cascade of disturbances that spreads throughout the monitored process. Symptoms of faults may appear in what seem to be unrelated parts of the process (effects at a distance). These effects can make fault management and diagnosis much more complicated. Increased coupling may often create more opportunities for situations to arise with conflicts between different goals (cf., Woods, 1988). And increasing the coupling within a system changes the kinds of system failures one expects to see (Perrow, 1984; Reason, 1990). The latent failure model for disaster is derived from data on failures in highly coupled systems.

Technology change creates the potential for new kinds of error and system breakdown as well as changing the potential for previous kinds of trouble. Take the classic simple example of the transition from an analog alarm clock to a digital one. With the former, errors are of imprecision—a few minutes off one way or another; with the advent of the latter, precision increases, but it is now possible for order-of-magnitude errors where the alarm is set to go off exactly 12 hours off (i.e., by confusing PM and AM modes). Design needs to occur with the possibility of error in mind (Lewis and Norman, 1986). Analysis of the potential for system breakdown should be a part of the development process for all technology changes (Norman, 1983). This point should not be interpreted as part of a go/no go decision about new technology. It is not the technology itself that creates the problem; rather it is how the technological possibilities are utilized vis á vis the constraints and needs of the operational system (Norman, 1990a). *Design to reduce errors and to enhance error recovery is part of the process of using technology skillfully rather than clumsily.*

The Clumsy Use of Computer Technology

We usually focus on the perceived benefits of new automated or computerized devices and technological aids. Our fascination with the possibilities afforded by technology in general often obscures the fact that new computerized and automated devices also create new burdens and complexities for the individuals and teams of practitioners responsible for operating, troubleshooting, and managing high-consequence systems. The demands may involve new or changed tasks such as device setup and initialization, configuration control, or operating sequences. Cognitive demands change as well, creating new interface management tasks, new attentional demands, the need to track automated device state and performance, new communication or coordination tasks, and new knowledge requirements. These demands represent new levels and types of operator workload.

The dynamics of these new demands are an important factor because in complex systems, human activity ebbs and flows, with periods of lower activity and more self-paced tasks interspersed with busy, high-tempo, externally paced operations where task performance is more critical (Rochlin et al., 1987). Technology is often designed to shift workload or tasks from the human to the machine. But the critical design feature for well integrated cooperative cognitive work between the automation and the human is not the overall or time-averaged task workload. Rather, it is how the technology impacts low-workload and high-workload periods, and especially how it impacts the practitioner's ability to manage workload that makes the critical difference between clumsy and skillful use of the technological possibilities.

A syndrome, which Wiener (1989) has termed "clumsy automation," is one example of technology change that in practice imposes new burdens as well as some of the expected benefits. Clumsy automation is a form of poor coordination between the human and machine in the control of dynamic processes where the benefits of the new technology accrue during workload troughs, and the costs or burdens imposed by the technology occur during periods of peak workload, high-criticality, or high-tempo operations. Despite the fact that these systems are often justified on the grounds that they would help offload work from harried practitioners, we find that they in fact create new additional tasks, force the user to adopt new cognitive strategies, require more knowledge and more communication at the very times when the practitioners are most in need of true assistance (Cook, Woods, and Howie, 1990; Sarter and Woods, in press). This creates opportunities for new kinds of human error and new paths to system breakdown that did not exist in simpler systems (Woods et al., 1992).

Patterns in the Clumsy Use of Computer Technology

To illustrate these new types of workload and their impact on practitioner cognition and behavior let us examine two series of studies, one looking at pilot interaction with cockpit automation, and the other looking at physician interaction with new information technology in the operating room. Both series of studies found that the benefits associated with the new technology accrue during workload troughs, and the costs associated with the technology occur during high-criticality, or high-tempo operations (Wiener, 1989; Sarter and Woods, 1992; 1994; Cook et al., 1990; Moll van Charante et al., 1993).

Clumsy automation on the flightdeck

Results indicate that one example of clumsy automation can be seen in the interaction between pilots and flight management computers (FMCs) in commercial aviation (e.g., Sarter and Woods, 1992; 1994). Under low-tempo operations pilots communicate instructions to the FMCs which then "fly" the aircraft. Communication between pilot and FMC occurs through a multi-function display and keyboard. Instructing the computers consists of a relatively effortful process involving a variety of keystrokes on potentially several different display pages and a variety of cognitive activities such as recalling the proper syntax or where data is located in the virtual display page architecture. Pilots speak of this activity as "programming the FMC."

Cockpit automation is flexible also in the sense that it provides many functions and options for carrying out a given flight task under different circumstances. For example, the FMC provides at least five different mechanisms at different levels of automation for changing altitude. This customizability is construed normally as a benefit that allows the pilot to select the mode or option best suited to a particular flight situation (e.g., time and speed constraints). However, it also creates a variety of new demands. For example, pilots must know about the functions of the different modes, how to coordinate which mode to use when, and how to "bumplessly" switch from one mode or level of automation to another. In other words, the supervisor of automated resources must not only know something about how the system works, but also know how to work the system. Monitoring and attentional demands are also created as the pilots must keep track of which mode is active and how each active or armed mode is set up to fly the aircraft (Sarter and Woods, 1992).

In a series of studies on pilot interaction with this suite of automation and computer systems (Sarter and Woods, 1992; 1994), the data indicated that it was relatively easy for pilots to lose track of the automated systems' behavior during high-tempo and highly dynamic situations. For example, pilots would miss mode changes that occurred without direct pilot intervention during the transitions between phases of flight or during the high-workload descent and approach phases in busy airspace. These difficulties with system and mode awareness reduced pilots' ability to stay ahead of the aircraft. As a result, when the pace of operations increased (e.g., in crowded terminal areas where the frequency of changes in instructions from air traffic control increases), pilots tended to abandon the flexible but complex modes of automation and switch to less automated, more direct means of flight control. Note that pilots speak of this tactic as "escaping" from the FMC (Sarter and Woods, 1992).

The loss of system awareness may not affect the individual pilot only; it also can impact the shared cognition across the crew. Interacting with the automation through multi-function controls and displays tends to suppress cues about the activities and intent of the other human crew member (for a counter-example of a low technology cockpit subsystem see Hutchins, 1991). As a result, the crew's ability to maintain a shared frame of reference or common situation assessment can break down and degrade communication and coordination across the crew. The threat of breakdowns in shared cognition is particularly important in more dynamic and complex flight contexts where effective coordination across pilots is needed to cope with non-routine or novel events (e.g., Segal, 1993).

To utilize highly flexible systems, the practitioner must learn about all the available options, learn and remember how to deploy them across the variety of real operational contexts that can occur, and learn and remember the interface manipulations required to invoke different modes or features. All of this represents new burdens on the practitioner to set up and manage these capabilities and features. Data on pilot interaction with these types of systems indicates that pilots tend to become proficient or maintain their proficiency on a subset of modes or options. As a result, they try to manage the system within these stereotypical responses or paths, underutilizing system functionality. In addition, the results showed that some of the knowledge acquired in training was available only theoretically, but that this knowledge was inert, i.e., the practitioners were not capable of applying the knowledge effectively in differing flight contexts.

Clumsy automation in the operating room: I. Centralizing data display

Another study, this time in the context of operating room information systems, reveals some other ways that new technology creates unintended complexities and provokes practitioner coping strategies (Cook et al., 1990; Cook, Woods, McColligan, and Howie, 1991). In this case a new operating room patient-monitoring system was studied in the context of cardiac anesthesia. This and other similar systems integrate what was previously a set of individual devices, each of which displayed and controlled a single sensor system, into a single CRT display with multiple windows and a large space of menu-based options for maneuvering in the space of possible displays, options, and special features. The study consisted of observing how the physicians learned to use the new technology as it entered the workplace.

By integrating a diverse set of data and patient monitoring functions into one computer-based information system, designers could offer users a great deal of customizability and options for the display of data. Several different windows could be called depending on how the users preferred to see the data. However, these flexibilities all created the need for the physician to interact with the information system—the physicians had to direct attention to the display and menu system and recall knowledge about the system. Furthermore, the computer keyhole created new interface management tasks by forcing serial access to highly inter-related data and by creating the need to periodically declutter displays to avoid obscuring data channels that should be monitored for possible new events.

The problem occurs because of a fundamental relationship: the greater the trouble in the underlying system or the higher the tempo of operations, the greater the information processing activities required to cope with the trouble or pace of activities (Woods et al., 1992). For example, demands for monitoring, attentional control, information, and communication among team members (including human-machine communication) all tend to go up with the tempo and criticality of operations. This means that the burden of interacting with the display system tends to be concentrated at the very times when the practitioner can least afford new tasks, new memory demands, or diversions of his or her attention away from patient state to the interface per se.

The physicians tailored both the system and their own cognitive strategies to cope with this bottleneck. In particular, they were observed to constrain the display of data into a fixed spatially dedicated default organization rather than exploit device flexibility. They forced scheduling of device interaction to low-criticality, self-paced periods to try to minimize any need for interaction at high-workload periods. They developed stereotypical routines to avoid getting lost in the network of display possibilities and complex menu structures.

Clumsy automation in the operating room: II. Reducing the ability for recovery from error or failure

This investigation started with a series of critical incidents involving physician interaction with an automatic infusion device during cardiac surgery (Cook et al., 1992). The infusion controller was a newly introduced computer-based device used to control the flow of blood pressure and heart rate medications to patients during heart surgery. Each incident involved delivery of a drug to the patient when the device was supposed to be off or halted. Detailed debriefing of participants suggested that, under certain circumstances, the device would deliver drug (sometimes at a very high rate) with little or no evidence to the user that the infusion was occurring. A series of investigations were done including observation of device use in context to identify:

- characteristics of the device which make its operation difficult to observe and error prone and,
- characteristics of the context of cardiac anesthesiology which

interact with the device characteristics to provide opportunities for unplanned delivery of drug.[21]

In cardiac surgery, the anesthesiologist monitors the patient's physiological status (e.g., blood pressure, heart rate) and administers potent vasoactive drugs to control these parameters to desired levels based on patient baselines, disease type, and stage of cardiac surgery. The vasoactive drugs are administered as continuous infusion drips mixed with intravenous fluids. The devicc in question is one type of automatic infusion controller that regulates the rate of flow. The user enters a target in terms of drops per minute; the device counts drops that form in a drip chamber, compares this to the target, and adjusts flow. If the device is unable to regulate flow or detccts one of several different device conditions, it is programmed to cease operation and emit an audible alarm and warning message. Thc interface controls consist of three multi-function buttons and a small LCD panel which displays target rate and messages. In clinical use in cardiac surgery up to six devices may be set up with different drugs that may be needed during the case.

The external indicators of the device's state provide poor feedback and make it difficult for physicians to assess or track device behavior and activities. For example, the physician users were unaware of various controller behavioral characteristics such as overshoot at slow target rates, "seek" behavior, and erratic control during patient transport. Alarms were remarkably common during device operation. The variety of different messages were ambiguous—several different alarm messages can be displayed for the same underlying problem; the different messages depend on operating modes of the deviee which are not indicated to the user. Given the lack of visible feedback, when alarms recurred or a sequence occurred, it was very difficult for the physician to determine whether the device had delivered any drug in the intervening period.

The most intense periods of device use also were those time pcriods of highest cognitive load and task criticality for the physicians, i.e., the time period of coming off cardio-pulmonary bypass. It is preeisely during these periods of high workload that the automated devices are sup-

[21] This same device was referred to in the example used at the end of Chapter 2 and was involved in Incident #2 described in Chapter 4.

posed to provide assistance (less user workload through more precise flows, smoother switching between drip rates, etc.). However, this was also the period where the largest number of alarms occurred and where device troubleshooting was most onerous.

Interestingly, users seemed quite aware of the potential for error and difficulties associated with device setup which could result in the device not working as intended when needed. They sought to protect themselves from these troubles in various ways, although the strategies were largely ineffective.

In the incidents, misassemblies or device problems led to inadvertent drug deliveries. The lack of visible feedback led physicians to think that the device was not delivering drug and was not the source of the observed changes in patient physiology. Large amounts of vasoactive drugs were delivered to brittle cardiovascular systems, and the physicians were unable to detect that the infusion devices were the source of the changes. Luckily in all of the cases, the physicians responded appropriately to the physiological changes with other therapies and avoided any adverse patient outcomes. The investigations revealed that various device characteristics led to an increased potential for erroneous assessments of device state and behavior. This played a role in the incidents because it impaired the physician's ability to detect and recover from erroneous actions and failures. Because of these effects, the relevant characteristics of the device can be seen as deficiencies from a usability point of view; the device design is "in error."

The results of this series of studies directly linked, for the same device and context, characteristics of computerized devices to increased potential for erroneous actions and impaired ability to detect and recover from errors. Furthermore, the studies directly linked the increased potential for erroneous setup and the decreased ability to detect errors as important contributors to critical incidents. In other words, design errors functioned as latent failures.

The Impact of Clumsy Automation on Cognitive System Activities and Practitioner Behavior

There are some important patterns in the results from the above studies and others like them. One is that characteristics of computer-based devices and systems affect the potential for different kinds of erroneous actions and assessments. Characteristics of computer-based devices that influence cognition and behavior in ways that increase the potential for erroneous actions and assessments can be considered flaws in the human-computer cognitive system. These flaws represent one kind of source of latent failures that can reside within a complex human-machine system (Reason, 1990). Activating this type of latent failure in the presence of other potentiating factors leads incidents nearer to disaster.

A second pattern is that the computer medium shapes the constraints for design. In pursuit of the putative benefits of automation, user customizability, and interface configurability and given some fundamental properties of the computer as a medium for representation, it is easy for designers to unintentionally create a thicket of modes and options, to create a mask of apparent simplicity overtop of underlying device or interface complexity, to create a large virtual perceptual field hidden behind a narrow keyhole (Woods, in press-b).

One factor that contributes to clumsy use of technological possibilities is that new technology is often designed around "textbook" or routine scenarios (Roth et al., 1987; Woods, 1991). However, design basis scenarios may be insufficient to test the ability of the distributed human-machine cognitive system to handle difficult problems. Note that the distinction between a "textbook" or anticipated situation and one with unanticipated elements depends on the nature of the pre-planned routines available to guide problem solving. The demand-resource view suggests that the difficulty of a problem is in part a function of unanticipated situations or *complicating factors.* A complicating factor is some variation or difficulty that goes beyond the standard method for handling or responding to the situation (see the discussion of local rationality in Chapter 4). Examples range from the relatively simple (underspecified instructions or human execution errors) to the complex (multiple failures or novel situations). The need to test a human-machine system by sampling complicating factors is guided by the need to measure the "brittleness" of the distributed cognitive system.

However, there seems to be a basic correlation such that the more the trouble in the underlying system or the higher the tempo of operations, the greater the information-processing activities required to cope with the trouble or pace of activities. For example, demands for monitoring, attentional control, information, and communication among team members (including human-machine communication) all tend to go up with the tempo and criticality of operations. Thus, the costs associated with clumsy uses of the technology will be minimal during textbook operations but will increase during higher workload situations.

A result that occurred in all the above studies and has recurred in other studies of the impact of new technology on practitioner cognitive activities is that practitioners actively adapted or tailored the information technology provided for them to the immediate tasks at hand in a locally pragmatic way, usually in ways not anticipated by the designers of the information technology (Roth et al., 1987; Flores et al., 1988; Cook, Woods, McColligan, and Howie, 1991; Hutchins, 1990). Tools are shaped by their users (Woods et al., 1992).

New technology introduced for putative benefits in terms of human performance in fact introduced new demands and complexities into already highly demanding fields of practice. Practitioners developed and used a variety of strategies to cope with these new complexities. Because practitioners are responsible agents in the domain, they work to insulate the larger system from device deficiencies and peculiarities of the technology. This occurs, in part, because practitioners inevitably are held accountable for failure to correctly operate equipment, diagnose faults, or respond to anomalies even if the device setup, operation, and performance are ill-suited to the demands of the environment.

In all of these studies practitioners tailored their strategies and behavior to avoid problems and to defend against device idiosyncrasies. However, the results also show how these adaptations may be only partly successful. The adaptations could be effective, or only locally adaptive, in other words, brittle to various degrees (i.e., useful in narrow contexts, but problematic in others). Practitioner tailoring may be inadequate because it is incomplete or ineffective, for example increasing the exposure of the system to other hazards.

Finally, it would be easy to label the problems noted above as simply "human-computer interaction deficiencies." In some sense they are exactly that. But the label "human-computer interaction" (HCI) carries with it many different assumptions about the nature of the relationship between people and technology. The examples above illustrate deficiencies that go beyond the concepts typically associated with the label "computer interface" in several ways.

First, all of these devices more or less meet guidelines and common practices for human-computer interaction defined as simply making the needed data nominally available, legible, and accessible (see Woods, 1991, and Woods, in press-b, for general treatments of the limits of design for data availability). The characteristics of the above systems are problems because of the way they shape practitioner cognition and behavior in their field of activity. These are not deficiencies in an absolute sense; whether or not they are flaws depends on the context of use.

Thus, the problems noted above cannot be seen without understanding device use in context. Context-free evaluations are unlikely to uncover the important problems, determine why they are important, and identify criteria that more successful systems should meet (see Woods and Sarter, 1993 for the general case and Cook, Potter, Woods, and McDonald, 1991 for one specific one).

Third, the label HCI easily conjures up the assumption of a single individual alone, rapt in thought, but seeing and acting through the medium of a computerized device. The cases above and the examples throughout this volume reveal that failures and successes involve a system of people, machine cognitive agents, and machine artifacts embedded in context. Thus, it is important to see that the deficiencies, in some sense, are not in the computer-based device itself. Yes, one can point to specific aspects of devices that contribute to problems (e.g., multiple modes, specific opaque displays, or virtual workspaces that complicate knowing where to look next), but the proper unit of analysis is not the device *or* the human. "Cause" should not be attributed either to the design or to the people Rather, *the proper unit of analysis is the distributed cognitive system*—characteristics of artifacts are deficient because of how they shape cognition and behavior across a distributed set of agents. Re-design of a clumsy device really should be about re-design of the distributed cognitive system rather than about the artifact per se (although ultimately such a re-design eventually does require and depend on specific characteristics of artifacts). *Clumsiness is not really in the technology; clumsiness arises in how the technology is used relative to the context of demands and resources and agents and other tools* (e.g., Norman, 1990b).

It is important to highlight this last point because of a potential mis-interpretation. We are *not* advocating abandonment of advanced computer-based technology. Technology is just a kind of power. We are trying to illustrate the difference between using the power of technology clumsily and skillfully from the point of view of the operational system.

Behavior- and Cognition-Shaping Properties of Computer-Based Technology

A Map: The Impact Flow Diagram

Figure 9 (p. 125) provides an overall map of the process by which the clumsy use of new computer technology affects the cognition and behavior of people embedded in an operational system, creating the potential for latent failures which could contribute to incidents or accidents. The figure is a schematic of the results of research on the relationship of computer technology, cognition, practitioner behavior, and system failure. We will refer to it as the Impact Flow Diagram because it maps how technology impacts cognition in context, how cognition impacts behavior in operational contexts, and how behavior can contribute to incident evolution.

As illustrated in the Impact Flow Diagram (Figure 9, p. 125), there are a variety of characteristics of computer-based systems and devices that shape the cognitive activities of people. In particular, we can think of a computer-based information system in terms of how it represents the underlying process for someone in some goal and task context (cf., Woods, in press-b for a more complete description of this concept which is at the heart of representation aiding and design). Some properties of the information system as a representation are problematic because of their impact on cognitive activities. For example, one can examine a prototype computerized device and notice that there are a large number of windows that could be opened and manipulated on a single VDU. Research indicates that if the computerized device has this characteristic, then it is likely that users may experience problems getting lost in the large space of display options, and it is likely that users will face new interface management burdens to manipulate the interface itself,

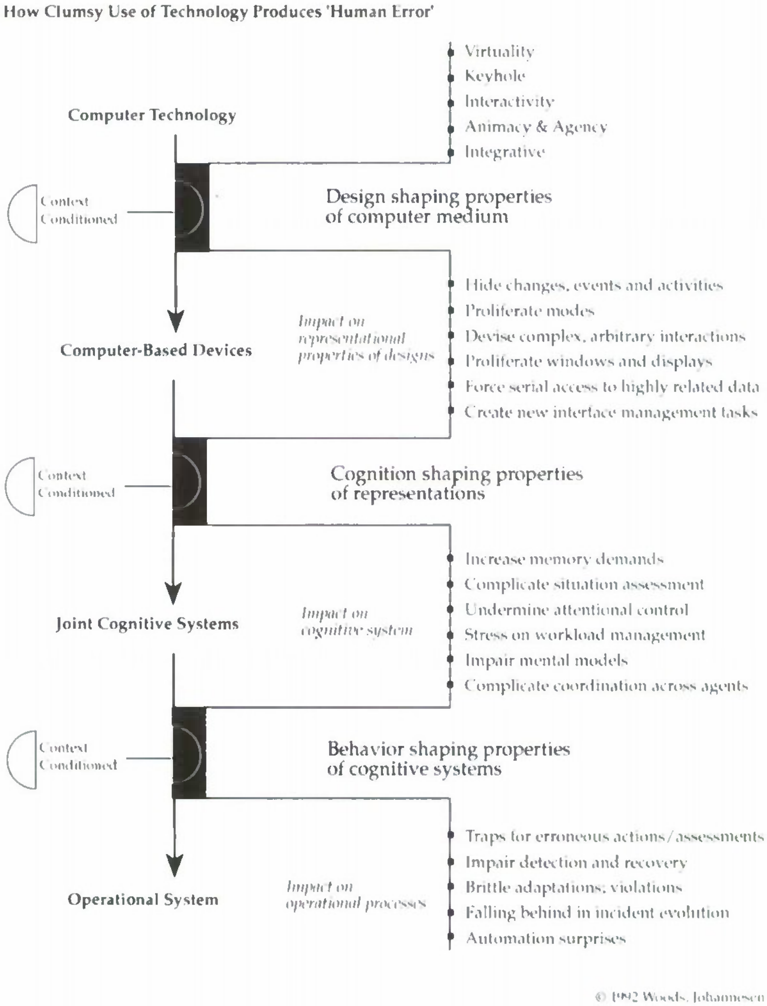

How Clumsy Use of Technology Produces 'Human Error'
Computer Technology
Virtuality
Keyhole
Interactivity
Animacy & Agency
Integrative
Context
Conditioned
Design shaping properties
of computer medium
Computer-Based Devices
Impact on
representational
properties of designs
Hide changes, events and activities
Proliferate modes
Devise complex, arbitrary interactions
Proliferate windows and displays
Force serial access to highly related data
Create new interface management tasks
Context
Conditioned
Cognition shaping properties
of representations
Joint Cognitive Systems
Impact on
cognitive system
Increase memory demands
Complicate situation assessment
Undermine attentional control
Stress on workload management
Impair mental models
Complicate coordination across agents
Context
Conditioned
Behavior shaping properties
of cognitive systems
Operational System
Impact on
operational processes
Traps for erroneous actions/assessments
Impair detection and recovery
Brittle adaptations; violations
Falling behind in incident evolution
Automation surprises
© 1992 Woods, Johannesen

Figure 9. This "Impact Flow Diagram" illustrates the relationship between the design-shaping properties of the computer as a medium, the cognition-shaping properties of representations in the computer medium, and the behavior-shaping properties of cognitive systems. The impact flow relationships define a latent failure path for the clumsy use of technology.

for example, de-eluttering the VDU surfaee (Cook et al., 1990; 1991h; Woods, in press-h). Negative consequenees will he larger if these data-management hurdens tend to congregate at high-eriticality, high-tempo periods of task performance. Another typieal prohlem is low ohservahility or opaque views where the computer graphics, through, for example, an over-relianee on displays of digital forms of raw values, give the illusion of informing the ohserver ahout the state of the underlying process when they aetually ohseure the ehanges, events, and aetivities in the underlying process (e.g., Potter et al., 1992).

These and other representational properties of computerized deviees are indicators of flaws heeause they contrihute to human-computer systems that tend to:

- make things invisihle, espeeially hiding "interesting" events, changes, and anomalies;
- proliferate modes;
- foree serial aecess to highly related data;
- proliferate windows and displays in virtual data spaee hehind a narrow aperture viewport;
- eontain complex and arhitrary sequences of operations, modes, and mappings;
- add new interfaee management tasks that tend to congregate at high-eriticality and high-tempo periods of the task;
- suppress cues ahout the aetivities of other team memhers, hoth machine and human (e.g., Norman, 1990a; 1990h).

Note that it is only hy examining how the computerized system represents the hehavior of the underlying process in question that one ean see these representational flaws. In other words, representational properties are hound to the context of the underlying proeess and the goals and tasks of the operational system that manages that proeess.

Characteristies of devices shape eognitive activities aeross the distrihuted system depending on the eontext of aetivities, demands, and goals in the partieular field of activity. The representational properties impaet eognitive systems (Figure 9, p. 125):

- through inereased demands on user memory,
- hy complicating situation assessment,
- hy undermining attentional control skills (where to focus when),
- add workload at high-criticality high-tempo periods,

- constrain the users' ability to develop effective workload management strategies,
- impair the development of accurate mental models of the function of the device and the underlying processes,
- decrease knowledge calibration (i.e., mislead users into thinking that their models are more accurate than they actually are),
- undermine the cognitive aspects of coordination across multiple agents.

These cognitive system changes are important because they influence how practitioners behave in various situations that can arise in a specific field of activity. In studies that look at the behavior-shaping consequences of these cognitive characteristics (e.g., Moll van Charante et al., 1993), one looks for:

- increased potential for different kinds of erroneous actions and erroneous assessments of process state (e.g., mode errors);
- impaired ability to detect and recover from failures, erroneous actions, or assessments;
- how the users tailor their behavior and the device to make it into a more usable tool, especially brittle tailoring that creates vulnerabilities to human-machine system breakdowns in special circumstances;
- increased risk of falling behind in incident evolution (loss of situation awareness and other breakdowns in attentional dynamics);
- automation surprises (Sarter and Woods, 1994) or other breakdowns in coordination across multiple agents;
- decreased learning opportunities.

When investigators work backwards from an accident, they typically find that one or more of these types of problems in operational processes were among the contributors to the incident evolution (e.g., Reason, 1990; Woods et al., 1987; Woods, 1991).

In other words, the Impact Flow Diagram traces a kind of latent failure chain. The clumsy use of technological possibilities shapes the cognition and behavior of the people embedded in the operational system in predictable patterns. There are design-shaping properties of the computer medium that make it easy for designers to create devices with typical flaws in human-computer cooperation. These characteristics are flaws because they create new cognitive demands and increase the stress on other cognitive activities. Behavior-shaping properties of cognitive systems link these effects to different kinds of operational consequences. As a result, these problems are latent failures that can contribute to incidents and accidents, if other potentiating factors are present.

There is a very ironic state of affairs associated with the clumsy use of technology. The very characteristics of computer-based devices that have been shown to complicate practitioners' cognitive activities and contribute to errors and failures, through studies of the device in context, are generally justified and marketed on the grounds that they reduce human workload and improve human performance. Examples of such putative claims include reduced skill requirements, greater attention to the joh, hetter efficiency, and reduced errors. Beware of superficial and context-free claims about the impact of new technology on human-machine systems. Understanding or predicting the effects of technology change requires one to study and to model distributed cognitive systems in context in order to see the cognition-shaping properties of the computer-hased representations and the behavior-shaping properties of the cognitive system. When our purpose is to help create new cognitive tools, we should start, not with context-free evaluations of the current or proposed prototype computer-based devices, hut by studying and modeling the distributed cognitive system, including the role of artifacts, in the context of the demands of the field of activity and the constraints imposed by the organizational context (see Figure 1, p. 21). Our goal is to understand the processes within a particular system that govern the expression of error and expertise. The resulting model of the distributed cognitive system guides the identification of ways to use technological powers skillfully to help the operational system function more effectively.

Classic Flaws in Human-Computer Cooperation: Designer "Error"

This section provides a brief discussion of how human computer cooperation flaws can he thought of as designer errors that arise from properties of the computer as a medium for representation and from factors that influence how new technology is deployed. These characteristics of computer-based devices are "flaws" because of how they shape the cognitive activities and behavior of practitioners.

A complete treatment of "flaws" in human-computer cooperation would require a volume in its own right. Plus, it evokes another need—aiding the designer of human-computer systems in the development of systems that improve operational performance in a particular setting. This problem of aiding design reaches out even further into the nature of the relationship of human factors as a profession to other technical and professional areas and communities. Thus, this section is not intended as a designer's guide, but simply to help the reader see how technology change influences people and can result in new types of errors or failure paths (the relationships in the Impact Flow Diagram). First, we will provide an overview of some of the typical flaws and how they arise. Second, we will provide a comprehensive treatment of one of these flaws—mode error, including potential countermeasures. Finally, we will use the mode error case to illustrate the relationships captured in the Impact Flow Diagram.

Penalties of Virtuality

"Every parameter you can control, you must control."[22]

A fundamental property of the computer as a medium for representation is *freedom from the physical constraints acting on the referent real world objects/systems* (Hochberg, 1986, p. 22-2 to 22-3). In many media (e.g., cinema), the structure and constraints operating in the physical world will ensure that much of the appropriate "information" about relationships in the referent domain is preserved in the representation. On the other hand in the computer medium, the designer of computer displays of data must do *all* of the work to constrain or link attributes and behaviors of the representation to the attributes and behaviors of the referent domain.

[22] W. Carlos' First Law of Digital Synthesized Music, 1992.

This property means that sets of displays of data in the computer medium can be thought of as a virtual perceptual field.[23] It is a perceivable set of stimuli, but it differs from a natural perceptual field and other media for representation because there is nothing inherent in the computer medium that constrains the relationship between things represented and their representation. This freedom from the physical constraints acting on the referent real world objects is a double-edged sword in human-computer cooperation, providing at the same time the potential for very poor representations and the potential for radically new and more effective representations.

The computer medium allows designers to combine multiple features, options, and functions onto a single physical platform. The same physical device can be designed to operate in many different contexts, niches, and markets simply by taking the union of all the features, options, and functions that are needed in any of these settings. In a sense, the computer medium allows one to create multiple virtual devices concatenated onto a single physical device. After all, the computer medium is multi-function—software can make the same keys do different things in different combinations or modes, or provide soft keys, or add new options to a menu structure; the CRT or other visual display unit (VDU) allows one to add new displays which can be selected if needed to appear on the same physical viewport. It is the *ne plus ultra* in modular media.

But this means that a practitioner cannot have the device in one context without also importing part of the complexity from all of the other contexts. Concatenating multiple virtual devices on a single platform forces practitioners concerned with only a single niche to deal with the complexity of all the other niches as well. This is in contradiction to what people are observed to do to cope with complexity—people divide up a domain to segregate the complexity in ways that are meaningful, i.e., into a series of local contexts. Furthermore, it is a funda-

[23] It would be perhaps more accurate to say that the computer medium can be thought of as an artificial perceptual field. Woods (in press-b) uses the word "virtual" to play off the current fashion in software. The use of virtual, as in virtual reality, is creating a new connotation for the word: virtual—giving the appearance or suggestion of a naturally occurring phenomenon while only approximating or even missing its essence; refers especially to computerized devices and systems.

mental research result that human cognition and behavior are conditioned to the context in which they occur.

The virtuality of computer-based information technology allows designers to develop new subsystems or devices with the appearance of simplicity by integrating diverse data, capabilities, and devices into a single multi-function CRT display and interface. But to do this pushes the designer to proliferate modes, to proliferate displays hidden behind the narrow viewport, to assign multiple functions to controls, to devise complex and arbitrary sequences of operation—in other words, to follow Don Norman's (1988) tongue-in-cheek advice on how to do things wrong in designing computer-based devices. Such systems appear on the surface to be simple because they lack large numbers of physical display devices and controls; however, underneath the placid surface of the CRT workstation there may be a variety of characteristics which produce cognitive burdens and operational complexities.

For example, it is easy to design systems where a few keys do many things in combination (e.g., Cook, Woods, McColligan, and Howie, 1991). But from the practitioners' perspective, this is very likely to create complex and arbitrary control sequences. The result can be memory burdens and fertile ground for a variety of phenotypical action errors such as omissions, repetitions, and for genotypical patterns in action errors, e.g., various slips of action, lapses (Reason and Mycielska, 1982). But practitioners will develop coping strategies to deal with the operational and cognitive clumsiness of these complexities—they create their own external memory aids (e.g., Norman, 1988).[24] An alternative technology-centered approach provides users with a generic keypad. The ultimate in flexibility in one sense, but, from a practitioner point of view, this makes all interactions the equivalent of "programming." As a result, the costs of interacting with the device's capabilities go up, which creates bottlenecks in high-tempo periods. How do practitioners cope? They *escape*—they abandon cooperative strategies with that device in high-tempo periods (Sarter and Woods, 1992; 1994).

[24] This leads to an approach to looking for excessive memory burdens imposed by computer-based devices analogous to the advice from "Deep Throat" in the Watergate investigations. As "Deep Throat" told Woodward to follow the money, so it is useful here to follow the paper—the notes attached near or on computer-based devices are strong clues about new memory burdens imposed by clumsy use of computer technology.

Keyhole property

Another important property of the virtual perceptual field of computer-based display systems is that the viewport size (the windows/ VDUs available) is very small relative to the large size of the artificial data space or number of data displays that potentially could be examined. In other words, the proportion of the virtual perceptual field that can be seen at the same time (physically in parallel) is extremely small. This property is often referred to as the keyhole effect (e.g., Woods, 1984). Given this property, shifting one's "gaze" within the virtual perceptual field is carried out by selecting another part of the artificial data space and moving it into the limited viewport.

To designers these properties appear quite simple—the computer medium seems to support multiple functions. A single physical device, the VDU, can be used to provide access in principle to any kind of view the designer, marketeer, or customer thinks relevant. But consider the cognition-shaping results of the characteristics of the typical computer-based systems produced with only this in mind.

The norm is that the observer can see only one small portion of the total data field at a time or a very small number of the potentially available displays (cf., Cook et al., 1990 and Woods et al., 1991 for examples). In addition, the default tendency is to use individual pieces of data as the base unit of organization with each piece of data placed in only one location within the virtual perceptual field (one "home"). The result appears to provide users with a great deal of user configurability; they can call up, into the physical viewports available, whichever view they desire to inspect at that time. Furthermore, since the data field is virtual, it is easy to proliferate displays and types of windows (via window management capabilities) each specialized for just one type of data.

An example

What are the cognitive consequences of the above properties? Computer-based devices with the characteristics discussed above impose cognitive burdens. The following example illustrates the extra workload that can be imposed when the structure of the interface forces serial search for highly inter-related data.

Consider, for example, an artificial intelligence (AI) system that is designed aeeording to these norms (from Woods et al., 1991). Raw data is the basic unit of display—shown as digital values, several tiled windows provide viewports on a single CRT, users can call up a variety of displays including many different menus, and many different displays that contain the sensor data on the state of the monitored process. This system also provides intelligent diagnosis and control eapabilities (in this sense we might speak of it as being "animate" and agent-like in that it ean act on its own).

When an event occurs, the affected parameter values (a number or numhers) change hue from white (meaning—"normal"), to red (meaning—"the component is being tested"), or purple (meaning—"a diagnosis has heen performed and the eomponent is in some sort of abnormal condition"). Assuming that the operator sees that an event has occurred at all (which is prohlematic on several grounds), all the practitioner knows is that this parameter is ahnormal; he or she does not know in what way it is abnormal or why the AI system considers the ehange important or interesting in the current context.

The practitioners have to decide, independent of the graphic and intelligent eapahilities of the "aiding" system, what other data to examine to pursue this event and apparent anomaly further. The users have to decide where to look next in the virtual perceptual field beyond the narrow keyhole. The users have to decide whether this change is even important in the partieular situation—should other events be investigated first? Is this ehange expeeted in the eurrent eontext? Does this signal warrant interrupting the ongoing lines of reasoning with regard to diagnosis or response selection?

If the practitioners decide to pursue the underlying event and its significance, they need to think of the other related data that will support their evaluations. They have then to think of where these data reside in the virtual field and how to eall up these displays. Relevant to this issue is whether the display ealled into the viewport contains reminders or prompts to the pertinent data, other displays, or navigation commands. Is the sequence direct or complicated, perhaps involving several layers of menu selections? For each menu or display ealled up, the practitioners must re-orient to the new view and search for the relevant item.

By the stage that the target data is found in the virtual space, practitioners may have opened several windows. Note how this creates a new operator interface management task—decluttering the workspace, where practitioners must remember to remove stale views and viewports. If decluttering is delayed, significant events in the monitored process may be missed. Or practitioners may only realize the need to declutter their virtual workspace when a new event has occurred that demands their attention. However, the decluttering task directs their attention to the interface itself when it should be focused on assessing the change in process state or evaluating how to respond to the change.

The structure of the computer information system forces practitioners into serial access to highly inter-related data. Users must search and assemble step by step and datum by datum the state of the process. Despite the graphic display capabilities of the system, the user must remember more not less (one example of what Norman calls the conspiracy against human memory in the design of computerized devices). The representation of the monitored process in the computer medium is underutilized as an external memory. The practitioners must build and maintain their own mental model of the state of the monitored process, assessments, and activities of the intelligent system. Practitioner attention is diverted to the computer system itself (where is a datum located in the virtual space? which menu provides access? how to navigate to that location?). New interface-management tasks are created such as decluttering. What makes this example particularly ironic is that advanced graphic and intelligent processing technologies are available. However, these technological powers not only do not support the relevant cognitive activities in dynamic fault management (the task domain), they actually create new cognitive demands. For example, they expand what Norman (1988) calls the "gulf of execution" (the difference between the practitioner's intentions and the actions allowed by the system) and the "gulf of evaluation" (the effort involved in interpreting the system's state relative to practitioner expectations).

Forcing serial access to highly related data

The structure of this computer-based system fragments data across different windows and displays, which forces the operator into a slow serial search to collect and then integrate related data. The proliferation of windows degrades rather than supports the cognitive component of interface navigation—knowing where to look next, and finding the right data at the right time (Woods, 1984; Elm and Woods, 1985).

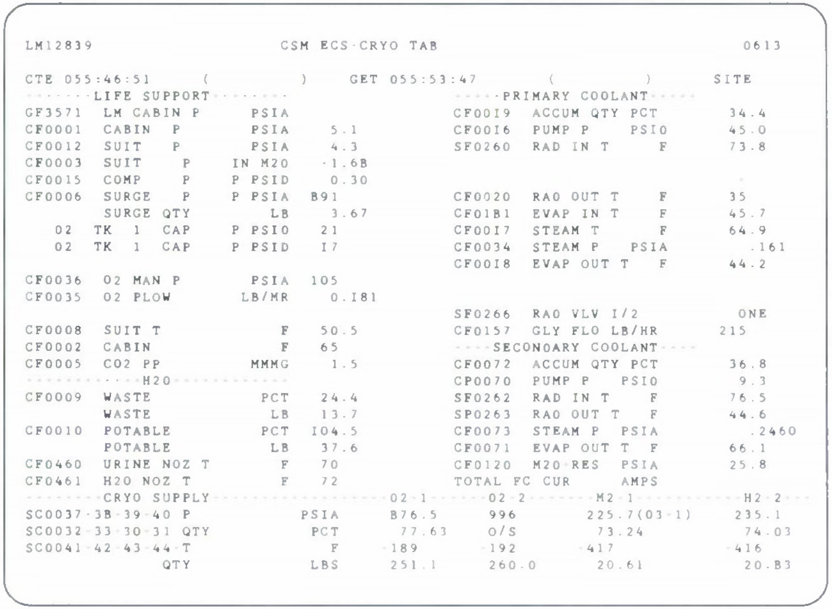
How do we know where to look next in a virtual perceptual field like this (cf., Woods, 1984)? Substantive tasks and sub-task coordination involve more data than can be displayed on even a windowed workstation. Thus, knowing where to look next in the data space available behind the limited viewports, and extracting information across multiple views is a fundamental cognitive activity. Yet, the default tendency in interface design is to leave out any orienting cues that indicate in mentally economical ways whether something interesting may be going on in another part of the virtual perceptual field (Woods, 1992). Instead, the processes involved in directing where-to-look-next are forced into a mentally effortful, high memory load, deliberative mode (in addition, the interface structure may create other cognitive problems in translating intentions into specific actions). The observer must remember where the desired data is located, to remember and execute the actions necessary to bring that portion of the field into the viewport, given he knows what data are potentially interesting to examine next (Woods et al., 1991). One can see the potential problems that derive from this keyhole property by imagining what it would be like to function with no peripheral vision or without other orienting perceptual systems to help determine where to look next, i.e., where to direct focal attention next.

To recap, the proliferation of windows and displays tends to fragment data across an increasingly complex structure of the virtual perceptual field. This forces serial access to highly inter-related data and increases the cognitive load in deciding where to look next (see Cook et al., 1990; Cook, Woods, McColligan, and Howie, 1991 for one example). While the device may possess great flexibility for users to tailor their workspace by manipulating the number, size, location and other window parameters, this flexibility creates new physical and cognitive tasks that can increase practitioner workload during high-tempo operations. If the extra interface management burdens tend to congregate at high-tempo, high-criticality periods, then there are constraints on practitioners' ability to find the right data at the right time. Practitioner attention shifts to the interface (where is the desired data located in the display space?) and to interface control (how do I navigate to that location in the display space?) at the very times where his or her attention needs to be devoted most to assessing and managing the monitored process.

Hiding interesting changes, events, and behaviors

Typically in computer-based representations, the basic unit of display remains an individual datum usually represented as a digital value, e.g., oxygen tank pressure is 297 p.s.i. (cf., Woods, 1991, or Woods et al., 1991 which contains examples of typical displays). Few attempts are made in the design of the representation of the monitored process to capture or highlight operationally interesting events—behaviors of the monitored process over time, (for one exception see Woods and Elias, 1988). This failure to develop representations that reveal change and highlight events in the monitored process has contributed to incidents where practitioners using such opaque representations miss operationally significant events (e.g., Freund and Sharar, 1990; Cook et al., 1992; Moll van Charante et al., 1993).

One well known accident where this representational deficiency contributed to the incident evolution (cf., Murray and Cox, 1989) was the Apollo 13 mission. In this accident, an explosion occurred in the oxygen portion of the cryogenics system (oxygen tank 2). The mission controller (the electrical, environmental, and communication controller or EECOM) monitoring this system was examining a screen filled with digital values (see Figure 10, p. 137, for a recreation of this display, the CSM ECS CRYO TAB display). After other indications of trouble in the spacecraft, he noticed, among a host of abnormalities in the systems he monitored, that oxygen tank 2 was depressurized (about 19 psi). In addition, most of the other mission controllers were seeing indications of trouble in the systems that they were responsible for. It took a precious 54 minutes as a variety of hypotheses were pursued before the team realized that the "command module was dying," and that an explosion in the oxygen portion of the cryogenics system was responsible for the extensive and evolving pattern of disturbances. The digital display had hidden the critical event: two digital values, out of 54 changing digital numbers, had changed anomalously (see Figures 10, p. 137; 11, p. 139; and 12, p. 140).



```
LM12839                          CSM ECS-CRYO TAB                                  0613

CTE 055:46:51     (              )    GET 055:53:47        (            )       SITE
-------LIFE SUPPORT---------                        -----PRIMARY COOLANT-----
GF3571  LM CABIN P      PSIA                        CF00I9  ACCUM QTY PCT        34.4
CF0001  CABIN  P        PSIA    5.1                 CF00I6  PUMP P     PSIO      45.0
CF0012  SUIT   P        PSIA    4.3                 SF0260  RAD IN T      F      73.8
CF0003  SUIT     P    IN M20   -1.6B
CF0015  COMP     P    P PSID    0.30                                              -
CF0006  SURGE    P    P PSIA  B91                   CF0020  RAO OUT T     F      35
        SURGE QTY         LB    3.67                CF01B1  EVAP IN T     F      45.7
   02  TK  1  CAP     P PSIO   21                   CF00I7  STEAM T       F      64.9
   02  TK  1  CAP     P PSID   I7                   CF0034  STEAM P    PSIA        .161
                                                    CF00I8  EVAP OUT T    F      44.2
CF0036  02 MAN P        PSIA  105
CF0035  02 PLOW        LB/MR    0.I81
                                                    SF0266  RAO VLV I/2           ONE
CF0008  SUIT T             F   50.5                 CF0157  GLY FLO LB/HR       215
CF0002  CABIN              F   65                   ----SECONOARY COOLANT----
CF0005  CO2 PP          MMMG    1.5                 CF0072  ACCUM QTY PCT        36.8
-------------H20------------                        CP0070  PUMP P    PSIO        9.3
CF0009  WASTE            PCT   24.4                 SF0262  RAD IN T     F       76.5
        WASTE             LB   13.7                 SP0263  RAO OUT T    F       44.6
CF0010  POTABLE          PCT  I04.5                 CF0073  STEAM P   PSIA         .2460
        POTABLE           LB   37.6                 CF0071  EVAP OUT T  F        66.1
CF0460  URINE NOZ T        F   70                   CF0120  M20-RES   PSIA       25.8
CF0461  H20 NOZ T          F   72                   TOTAL FC CUR      AMPS
--------CRYO SUPPLY-------------------02-1------02-2-------M2-1----------H2-2---
SC0037-3B-39-40 P             PSIA    B76.5     996        225.7(03-1)    235.1
SC0032-33-30-31 QTY            PCT     77.63    O/S         73.24          74.03
SC0041-42-43-44-T                F    -189      -192       -417           -416
            QTY                LBS    251.1     260.0       20.61          20.B3
```

© 1992 Woods and Holloway

Figure 10. Partial reconstruction of the computer display (display CSM ECS CRYO TAB) monitored by the electrical, environmental, and communication controller (EECOM) shortly before the explosion (55:54:44 mission time) that occurred during the Apollo 13 mission. Note oxygen tank 2 pressure is at 996 p.s.i.—high but still within accepted limits (From Woods, in press-b).

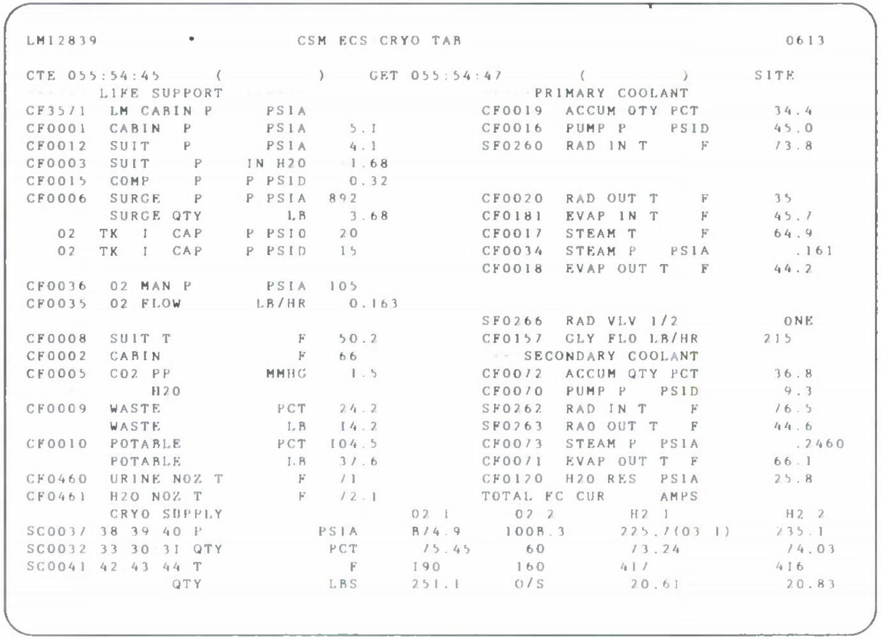
> So none of the three noticed the numbers for oxygen tank 2 during four particularly crucial seconds. At 55 hours, 54 minutes, and 44 seconds into the mission, the pressure stood at 996 p.s.i.—high but still within accepted limits. One second later, it peaked at 1,008 p.s.i. By 55:54:48, it had fallen to 19 p.s.i . . . . If one of them had seen the pressure continue on through the outer limits, then plunge, he would have been able to deduce that oxygen tank 2 had exploded (see Figure 13, p. 141). It would have been a comparatively small leap . . .to have put the whole puzzle of multiple disturbances across normally unconnected systems together (Murray and Cox, 1989, p. 406).

It was reported that the controller experienced a continuing nightmare for two weeks following the incident, in which "he looked at the screen only to see a mass of meaningless numbers. . . ." Finally, a new version of the dream came—he looked at the critical digitals "before the bang and saw the pressure rising. . . . Then the tank blew, and he saw the pressure drop and told Flight exactly what had happened" (Murray and Cox, 1989, p. 407).

The poor representation could be compensated for through human adaptability and knowledge; in other words, in Norman's terminology, knowledge-in-the-head can compensate for the absence of knowledge-in-the-world. But, what is the point of the computer as a medium for the display of data if it does not reduce practitioner memory loads? And yet, in computer system after computer system (e.g., Woods et al., 1991) we find that, despite the availability of new computational and graphic power, the end result is an increase in demands on practitioner memory. The contrast cannot be greater with studies of successful, but often technologically simple, cognitive artifacts that reveal how effective cognitive tools offload memory demands, support attentional control, and support the coordination of cognitive work across multiple agents (e.g., Hutchins, 1991).

To begin to move towards better representations that do not obscure the perception of events in the underlying system, there are three interrelated critical criteria in representation design (from Woods, in press-b):

1. Put data into context: (a) put a given datum into the context of related values; (b) collect and integrate data about important domain issues. Data are informative based on *relationships* to other data, rela-

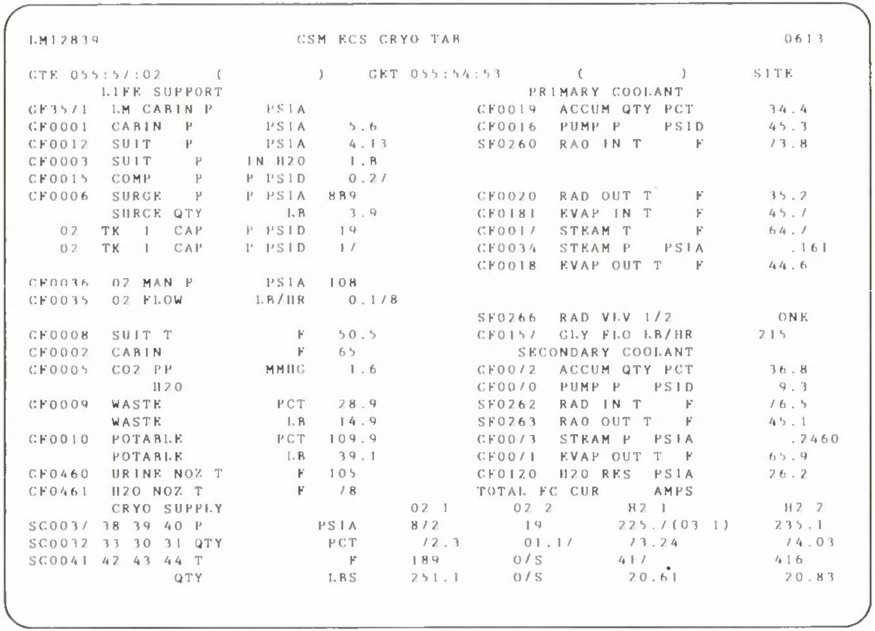


```
LM12839          •         CSM ECS CRYO TAB                                      0613

CTE 055:54:45      (              )    GET 055:54:47         (            )      SITE
         LIFE SUPPORT                             PRIMARY COOLANT
CF3571  LM CABIN P        PSIA                    CF0019  ACCUM OTY PCT          34.4
CF0001  CABIN   P         PSIA    5.1             CF0016  PUMP P    PSID         45.0
CF0012  SUIT    P         PSIA    4.1             SF0260  RAD IN T     F         73.8
CF0003  SUIT     P      IN H20    1.68
CF0015  COMP     P      P PSID    0.32
CF0006  SURGE    P      P PSIA  892               CF0020  RAD OUT T    F         35
        SURGE QTY           LB    3.68            CF0181  EVAP IN T    F         45.7
   02  TK  I  CAP       P PSIO   20               CF0017  STEAM T      F         64.9
   02  TK  I  CAP       P PSID   15               CF0034  STEAM P   PSIA           .161
                                                  CF0018  EVAP OUT T   F         44.2
CF0036  02 MAN P          PSIA  105
CF0035  02 FLOW          LB/HR    0.163
                                                  SF0266  RAD VLV 1/2             ONE
CF0008  SUIT T               F   50.2             CF0157  GLY FLO LB/HR        215
CF0002  CABIN                F   66                  SECONDARY COOLANT
CF0005  CO2 PP            MMHG    1.5             CF0072  ACCUM QTY PCT          36.8
            H20                                   CF0070  PUMP P   PSID           9.3
CF0009  WASTE              PCT   24.2             SF0262  RAD IN T     F         76.5
        WASTE               LB   14.2             SF0263  RAO OUT T    F         44.6
CF0010  POTABLE            PCT  104.5             CF0073  STEAM P  PSIA            .2460
        POTABLE             LB   37.6             CF0071  EVAP OUT T  F          66.1
CF0460  URINE NOZ T          F   71               CF0120  H20 RES  PSIA          25.8
CF0461  H2O NOZ T            F   72.1             TOTAL FC CUR      AMPS
        CRYO SUPPLY                     02 1      02 2          H2 1             H2 2
SC0037 38 39 40 P               PSIA    874.9     1008.3        225.7(03 1)      235.1
SC0032 33 30 31 QTY              PCT     75.45      60           73.24            74.03
SC0041 42 43 44 T                  F    190        160          417              416
               QTY               LBS    251.1      O/S           20.61            20.83
```

© 1992 Woods and Holloway

Figure 11. Partial reconstruction of the same computer display shown in Figure 10 (p. 137) during the pressure transient caused by the explosion (55:54:45 mission time). Note oxygen tank 2 pressure showed a peak at this point of 1,008 p.s.i. (From Woods, in press-b).

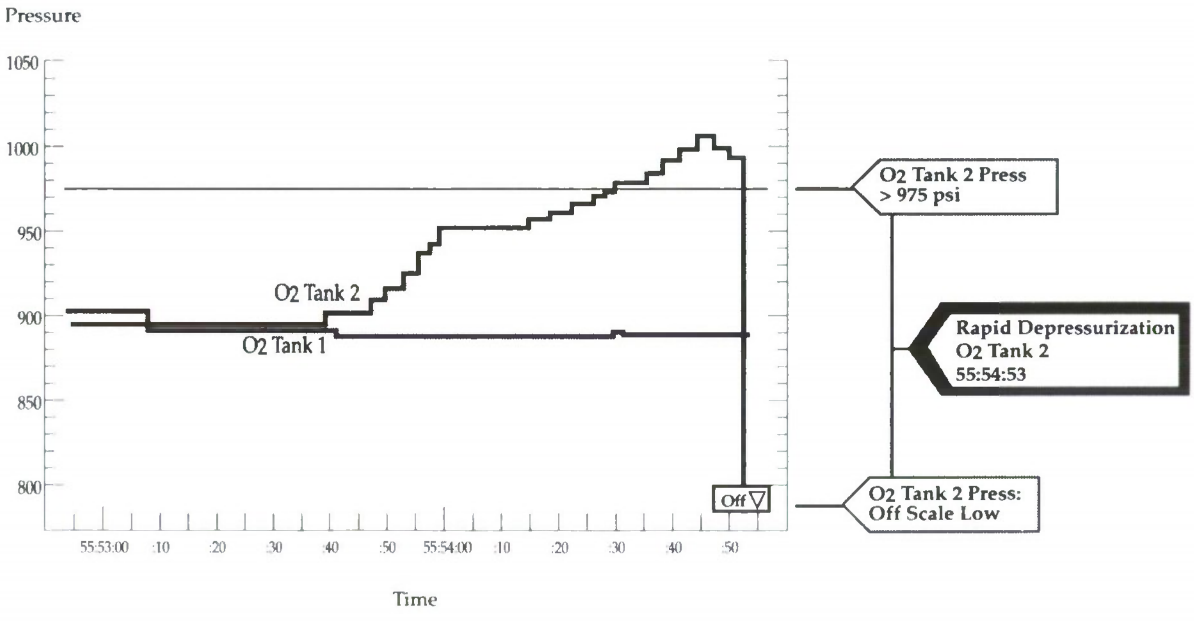


```
LM12839                         CSM ECS CRYO TAB                                          0613

GTE 055:57:02      (            )     GET 055:54:53        (           )        SITE
      LIFE SUPPORT                                       PRIMARY COOLANT
GF3571  LM CABIN P      PSIA                      CF0019  ACCUM QTY PCT        34.4
CF0001  CABIN  P        PSIA    5.6               CF0016  PUMP P    PSID       45.3
CF0012  SUIT   P        PSIA    4.13              SF0260  RAO IN T     F       73.8
CF0003  SUIT     P    IN H2O    1.8
CF0015  COMP     P    P PSID    0.27
CF0006  SURGE    P    P PSIA  889                 CF0020  RAD OUT T    F       35.2
        SURCE QTY         LB    3.9               CF0181  EVAP IN T    F       45.7
   02  TK  1  CAP     P PSID   19                 CF0017  STEAM T      F       64.7
   02  TK  1  CAP     P PSID   17                 CF0034  STEAM P   PSIA         .161
                                                  CF0018  EVAP OUT T   F       44.6
CF0036  07 MAN P        PSIA  108
CF0035  02 FLOW       LB/HR     0.178
                                                  SF0266  RAD VLV 1/2           ONE
CF0008  SUIT T             F   50.5               CF0157  GLY FLO LB/HR      215
CF0002  CABIN              F   65                   SECONDARY COOLANT
CF0005  CO2 PP          MMHG    1.6               CF0072  ACCUM QTY PCT        36.8
            H2O                                   CF0070  PUMP P    PSID        9.3
CF0009  WASTE            PCT   28.9               SF0262  RAD IN T     F       76.5
        WASTE             LB   14.9               SF0263  RAO OUT T    F       45.1
CF0010  POTABLE          PCT  109.9               CF0073  STEAM P   PSIA         .2460
        POTABLE           LB   39.1               CF0071  EVAP OUT T   F       65.9
CF0460  URINE NOZ T        F  105                 CF0120  H2O RES   PSIA       26.2
CF0461  H2O NOZ T          F   78                 TOTAL FC CUR      AMPS
        CRYO SUPPLY                   02 1        02 2         H2 1              H2 2
SC0037 38 39 40 P          PSIA       872          19          225.7(03 1)       235.1
SC0032 33 30 31 QTY         PCT        72.3        01.17        73.24             74.03
SC0041 42 43 44 T             F       189         O/S          417               416
             QTY            LBS       251.1       O/S           20.61             20.83
```

© 1992 Woods and Holloway

Figure 12. Partial reconstruction of the same computer display shown in Figures 10 and 11 (pp. 137 and 139) immediately after the explosion (55:54:48 mission time). Note oxygen tank 2 is depressurized (From Woods, in press-b).



© 1992 Woods and Holloway

Figure 13. Graphic overview of the behavior of oxygen tank pressure during the critical seconds of the Apollo 13 mission. (From Woods, in press-b).

tionships to larger frames of reference, and relationships to the interests and expectations of the observer. The challenge is the context sensitivity problem—what is interesting depends on the context in which it occurs.

2. Highlight change and events. Representations should highlight change/events and help reveal the dynamics of the monitored process. Events are temporally extended behaviors of the device or process involving some type of change in an object or set of objects. Recognizing an event involves recognition of both the object and the type of change. One key question is to determine what are "operationally interesting" changes or sequences of behavior, for example, highlight approach to a limit, highlight movement and rate of change, emphasize what event will happen next, and highlight significant domain events. Representing change and events is critical because the computer medium affords the possibility of dynamic reference—the behavior of the representation can refer to the structure and behavior of the referent objects and processes.

3. Highlight contrasts. Representations should highlight and support observer recognition of contrasts. Meaning lies in contrasts—*some departure from a reference or expected course.* Representing contrast means that one indicates the relation between the contrasting objects, states, or behaviors. One shows how the actual course of behavior follows or departs from some reference or expected sequence of behavior given the relevant context. Representing contrast indicates both the contrasting states or behavior and their relationship (how behavior departs or conforms to the contrasting case). Simply color coding a number or icon red (for danger), for example, shows that some anomaly is present, but it does not show the contrast of what is anomalous relative to what (Woods, 1992).

But given that the computer representation is free from the physical constraints acting on the referent objects, support for event perception in the computer medium requires the designer to actively identify operationally interesting changes or sequences of behavior and to actively develop representations that highlight these events to the observer given the actual task context. The default representations typically available do not make interesting events directly available for the practitioner to observe (Flach, Hancock, Caird, and Vicente, in press). Rather, practitioners are forced into a serial deliberative mode of cognition to abstract change and events from the displayed data (typically digital representations of sensed data).

This section has only briefly introduced a few of the typical ways that technology is used clumsily. The next section takes one human-computer problem—mode error, and attempts to explore it in more detail as an example and model for both practical concerns and research issues related to the clumsy use of technology.

Mode Error in Supervisory Control

In this section we try to provide a comprehensive overview of mode error and possible countermeasures. Mode error is one kind of breakdown in the interaction between humans and machines, especially computerized devices (Lewis and Norman, 1986). Norman (1988, p. 179) summarizes the source of mode error quite simply by suggesting that one way to create or increase the possibilities for erroneous actions is to ". . . change the rules. Let something be done one way in one mode and another way in another mode." When this is the case, a human user can commit an erroneous action by executing an intention in the way appropriate to one mode of the device when the device is actually in another mode. Put simply, multiple modes in devices create the potential for mode errors.[25]

Mode error is inherently a human-machine system breakdown. It requires a user who loses track of the system's active mode configuration and a machine that interprets user input differently depending on the current mode of operation. To understand the potential for mode error one needs to analyze the computer-based device in terms of what modes and mode transitions are possible, the context of how modes may come

[25] Another (more whimsical) way to understand mode error is to consider the following story based on the travels of a wordly cognitive systems researcher. "While traveling in Europe, I visited a castle built by a Danish King. The castle was large, with many rooms and corridors, and it was easy to get lost. It was remarkable, in part, for the multiple moats, designed to prevent attackers from gaining access to the King. At one point I discovered a collection of coats of arms to which I wanted to return. In looking for that room later, however, I found that I had already crossed over one of the moats and had been searching for the room in entirely the wrong complex of buildings. I was on the wrong side of the moat and had thus committed my first 'moat error'" (Richard Cook, personal communication, 1993).

into effect in dynamic scenarios, and how the mode of the device is represented in these contexts. Mode error is a failure of the distributed cognitive system (Hutchins, in press).

Several studies have shown how multiple modes can lead to erroneous actions and assessments (e.g., Lewis and Norman, 1986; Cook, Potter, Woods, and McDonald, 1991), and several design techniques have been proposed to reduce the chances for mode errors (Monk, 1986; Sellen, Kurtenbach, and Buxton, 1992). These studies also illustrate how evaluation methods can and should be able to identify computerized devices which have a high potential for mode errors.

Characteristics of the computer medium (e.g., its virtuality) and characteristics of design environments/processes make it easy for designers to proliferate modes and to create more complex interactions across modes. The result is new opportunities for mode errors to occur and new kinds of mode-related problems. Surprisingly, there is no single comprehensive treatment of mode errors or potential countermeasures available for designers (with the possible exception of Norman, 1988). Human-computer guidelines have been almost universally silent on the topic of mode errors, yet it is one of the common "design errors" in computer-based systems.

This section provides an overview of the current knowledge and understanding of mode error. We also suggest that it may be time to revise the traditional concept of mode error to account for changes in the nature of automated systems. Futhermore, we discuss possible ways to predict and prevent mode error especially in supervisory control of automated resources.

The Classic Concept of Mode Error

The concept of mode error was originally developed in the context of relatively simple computerized devices, such as word processors, used for self-paced tasks wherein the device only reacts to user inputs and commands. Mode errors in these contexts occur when an intention is executed in a way appropriate for one mode when, in fact, the system is in a different mode (see Norman, 1981). In this case, mode errors present themselves phenotypically as errors of commission. The mode error that precipitated the chain of events leading to the Strasbourg accident (Monnier, 1992; Lenorovitz, 1992a), in part, may have been of this form—the pilot appears to have entered the correct digits for the planned descent given the syntactical input requirements (33 was entered, intended to mean an angle of descent of 3.3 degrees); however, the automation was in a different descent mode which interpreted the entered digits as a different instruction (as meaning a rate of descent of 3300 feet per minute). Losing track of which mode the system is in is a critical component of a mode error. One part of this breakdown in situation assessment seems to be that device or system modes tend to change at a different rhythm relative to other user inputs or actions.

Norman (1981, 1988) classified mode errors as slips of action, but this seems problematic. In one sense a mode error involves a breakdown in going from intention to specific actions. But in another sense a breakdown in situation assessment has occurred—the practitioner has lost track of device mode. Mode errors emphasize that the consequences of an action depend on the context in which it is carried out. On the surface, the operator's intention and the executed action(s) appear to be in correspondence; the problem is that the meaning of action is determined by another variable—the system's mode status. This component makes it difficult to simply categorize mode errors as either a slip of action or an intention formation problem; elements of both seem to be present in a unique mix.

Designers should examine closely the mode characteristics of computerized devices and systems for the potential for creating this predictable form of human-computer breakdown. Multiple modes shape practitioner cognitive processing in two ways. First, the use of multiple modes increases memory and knowledge demands—one needs to know or remember the effects of inputs and the meanings of indications in the various modes. Second, it increases demands on situation assessment and awareness. The difficulty of these demands is conditional on how the interface signals device mode (observability) and on characteristics of the distributed set of agents who manage incidents. The difficulty of keeping track of which mode the device is in also varies depending on the task context (time-pressure, interleaved multiple tasks, workload).

Design countermeasures to the classic mode problems are straightforward in principle:

- Eliminate unnecessary modes (in effect, recognize that there is a cost in operability associated with adding modes for flexibility, marketing, and other reasons).
- Look for ways to increase the tolerance of the system to mode error. Look at specific places where mode errors could occur and (since these are errors of commission) be sure that (a) there is a recovery window before negative consequences accrue and (b) that the actions are reversible.
- Provide better indications of mode status and better feedback about mode changes.

Mode Error and Automated Systems

Human supervisory control of automated resources in event-driven task domains is a quite different type of task environment compared to the applications in the original research on mode error. Automation is often introduced as a resource for the human supervisor, providing him with a large number of modes of operation for carrying out tasks under different circumstances. The human's role is to select the mode best suited to a particular situation.

However, this flexibility tends to create and proliferate modes of operation which create new cognitive demands on practitioners (Woods, 1993b). Practitioners must know more—both about how the system works in each different mode and about how to manage the new set of options in different operational contexts. New attentional demands are created as the practitioner must keep track of which mode the device is in, both to select the correct inputs when communicating with the automation and to track what the automation is doing now, why it is doing it, and what it will do next. These new cognitive demands can easily congregate at high-tempo and high-criticality periods of device use, thereby adding new workload at precisely those time periods where practitioners are most in need of effective support systems.

These cognitive demands can be much more challenging in the context of highly automated resources. First, the flexibility of technology allows automation designers to develop much more complicated systems of device modes. Designers can provide multiple levels of automation and more than one option for many individual functions. As a result, there can be quite complex interactions across the various modes including "indirect" mode transitions. As the number and complexity of modes increase, it can easily lead to separate fragmented indications of mode status. As a result, practitioners have to examine multiple displays, each containing just a portion of the mode status data, to build a complete assessment of the current mode configuration.

Second, the role and capabilities of the machine agent in human-machine systems have changed considerably. In early devices, each system activity was dependent upon operator input; as a consequence, the operator had to act for an error to occur. With more advanced systems, each mode itself is an automated function which, once activated, is capable of carrying out long sequences of tasks autonomously *in the absence of additional commands from human supervisors*. For example, advanced cockpit automation can be programmed to automatically control the aircraft shortly after takeoff through landing. This is an increase in the apparent animacy and agency of the machine portion of a joint cognitive system. This increased capability of the automated resources themselves creates increased delays between user input and feedback about system behavior. As a result, the difficulty of error or failure detection and recovery goes up and inadvertent mode settings and transitions may go undetected for long periods. This allows for errors of omission (i.e., failure to intervene) in addition to errors of commission in the context of supervisory control.

Third, modes can change in new ways. Classically, mode changes only occurred as a reaction to direct operator input. In advanced technology systems, mode changes can occur indirectly based on situational and system factors as well as operator input. In the case of highly automated cockpits, for example, a mode transition can occur as an immediate consequence of pilot input. But it can also happen when a preprogrammed intermediate target (e.g., a target altitude) is reached or when the system changes its mode to prevent the pilot from putting the aircraft into an unsafe configuration.

This capability for "indirect" mode changes, independent of direct and immediate instructions from the human supervisor, drives the demand for mode awareness. Mode awareness is the ability of a supervisor to track and to anticipate the behavior of automated systems (Sarter and Woods, in press). Maintaining mode awareness is becoming increasingly important in the context of supervisory control of advanced technology which tends to involve an increasing number of interacting modes at various levels of automation to provide the user with a high degree of flexibility. Human supervisors are challenged to maintain awareness of which mode is active and how each active or armed mode is set up to control the system, the contingent interactions between environmental status and mode behavior, and the contingent interactions across modes. Mode awareness is crucial for any users operating a multi-mode system that interprets user input in different ways depending on its current status.

The complexity of modes, interactions across modes, and indirect mode changes create new paths for errors and failures. No longer are modes only selected and activated through deliberate explicit actions. Rather, pushing a button can result in the activation of different modes depending on the system status at the time of manipulation. The active mode that results may be inappropriate for the context, but detection and recovery can be very difficult in part due to long time-constant feedback loops.

An example of such an inadvertent mode activation contributed to a recent major accident in the aviation domain (the Bangalore accident, e.g., Lenorovitz, 1990). In that case, the pilot put the automation into a mode called OPEN DESCENT during an approach without realizing it. In this mode aircraft speed was being controlled by pitch rather than thrust (as would have been the case in the desirable mode for this phase of flight, i.e., in the SPEED mode). As a consequence, the aircraft could not sustain the glidepath and maintain the pilot-selected target speed at the same time. As a result, the flight director bars commanded the pilot to fly the aircraft well below the required profile to try to maintain airspeed. It was not until 10 seconds before impact that the crew discovered what had happened, too late for them to recover with engines at idle. How could this happen?

One contributing factor in this accident may have been several different ways of activating the OPEN DESCENT mode (i.e., at least five). The first two options involve the explicit manual selection of the OPEN DESCENT mode. In one of these two cases, the activation of this mode is dependent upon the automation being in a particular state.

The other three methods of activating the OPEN DESCENT mode are indirect in the sense of not requiring any explicit manual mode selection. They are related to the selection of a new target altitude in a specific context or to protections that prevent the aircraft from exceeding a safe airspeed. In this case, for example, the fact that the automation was in the ALTITUDE ACQUISITION mode resulted in the activation of OPEN DESCENT mode when the pilot selected a lower altitude. The pilot may not have been aware of the fact that the aircraft was within 200 feet of the previously entered target altitude (which is the definition of ALTITUDE ACQUISITION mode). Consequently, he may not have expected that the selection of a lower altitude at that point would result in a mode transition. Because he did not expect any mode change, he may not have closely monitored his mode annunciations, and hence missed the transition.

Display of data can play an important role when user-entered values are interpreted differently in different modes. In the following example, it is easy to see how this may result in unintended system behavior. In a current highly automated or "glass cockpit" aircraft, pilots enter a desired vertical speed or a desired flight path angle via the same display. The interpretation of the entered value depends on the active display mode. Although the verbal expressions for different targets differ considerably (for example: a vertical speed of two-thousand-five-hundred feet vs. a flight path angle of two-point-five degrees), these two targets on the display look almost the same (see Figure 14, p. 150). The pilot has to know to pay close attention to the labels that indicate mode status. He has to remember the indications associated with different modes, when to check for the currently active setting, and how to interpret the displayed indications. In this case, the problem is further aggravated by the fact that feedback about the consequences of an inappropriate mode transition is limited. The result is a cognitively demanding task; the displays do not support a mentally economical, immediate apprehension of the active mode (Woods, 1992). Cook , Potter, Woods, & McDonald (1991) also found a kind of mode problem in displays where the same alarm messages meant different things in different modes.

Coordination across agents in the distributed cognitive system is another important factor contributing to mode error in advanced systems.

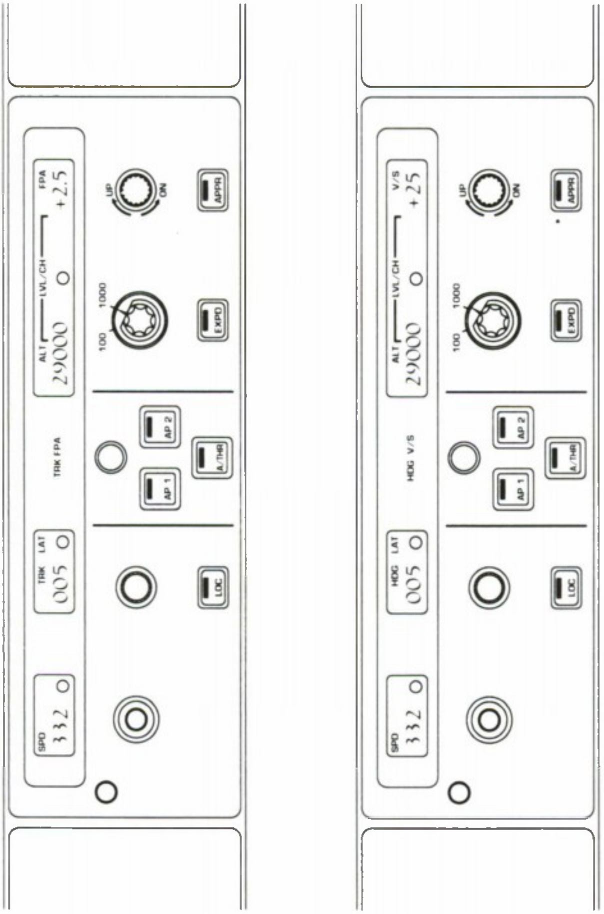


Figure 14. Example of multiple modes and the potential for mode error on the flight deck of an advanced technology aircraft. The same entry means different things in different modes (From Sarter and Woods, in press). Finding the different mode indications between the flight decks is left as an exercise to the reader.

Tracking system status and behavior becomes more difficult if it is possible for other users to interact with the system without the need for consent by all operators involved (the indirect mode changes are one human-machine example of this).

This problem is most obvious when two experienced operators have developed different strategies of system use. When they have to cooperate, it is particularly difficult for them to maintain awareness of the history of interaction with the system which may determine the effect of the next system input. In addition, the design of the interface to the automation may suppress cognitively economical cues about the activities of other agents within the distributed system (Hutchins, 1990; Woods, 1992).

The demands for mode awareness are critically dependent on the nature of the interface between the human and machine agents (and as pointed out above between human agents as well). If the computerized device also exhibits another of the HCI problems we noted earlier—not providing users with effective feedback about changes in the state of a device, automated system, or monitored process—then losing track of which mode the device is in may be surprisingly easy, at least in higher workload periods.

The above examples illustrate how a variety of factors can contribute to a lack of mode awareness on the part of practitioners. *Gaps or misconceptions in practitioners' mental models* may prevent them from predicting and tracking indirect mode transitions or from understanding the interactions between different modes. The *lack of salient feedback on mode status and transitions* (low observability) can also make it difficult to maintain awareness of the current and future system configuration. In addition to allocating attention to the different displays of system status and behavior, practitioners have to monitor *environmental states and events*, remember *past instructions to the system*, and consider possible *inputs to the system by other practitioners*. If they manage to monitor, integrate, and interpret all this information, system behavior will appear deterministic and transparent. However, depending on circumstances, missing just one of the above factors can be sufficient to result in an automation surprise and the impression of an animate system that acts independently of operator input and intention.

Recognizing the Potential for Mode Error

As illustrated in the above sections, mode error is a form of human-machine system breakdown. As systems of modes become more interconnected, more animate and agent-like (automated), and more event-driven, new types of mode-related problems are likely, unless the extent of communication between man and machine changes to keep pace with the new cognitive demands.

To uncover the potential for mode error in supervisory control of dynamic systems, it is essential that the dynamic behavior of devices be tested in the context of scenarios that go beyond textbook cases. Sarter and Woods(1994) have shown that one of the major mode-related problems for operators is to track mode transitions that do not immediately follow operator input. Therefore, it is not sufficient to look at mode annunciations statically without considering their behavior in times of transition. Mode annunciations have to be evaluated in a dynamic context to determine whether they succeed in capturing the operator's attention in times of change.

Dynamic testing involves both analytical and empirical approaches. The goal of the analytical approach is to lay out the functional structure of the system under all potential circumstances. One way to reach this goal is to create a state transition diagram which shows all possible system states and interactions between user and system. Such a diagram can help form hypotheses about potential problems in the system's functional design.

While this analytical approach focuses on problems related to the intended use of the system, the empirical approach emphasizes the need to work with practitioners to determine for what purposes and in what ways the system is actually being used. Hypotheses generated through the analytical approach can guide empirical explorations which, in turn, can reveal unanticipated difficulties and the users' response or adaptation to them. Empirical investigations also provide information about users' mental mod-els of the device—another valuable pointer to latent mode-related problems.

Countermeasures to Mode Error

Typically, recommendations for countermeasures against mode error fall into a few basic classes:

- reduce and simplify modes;
- provide better feedback on mode status, changes, and the implications of such changes;
- provide training that (a) supports acquisition and maintenance of better mental models of mode behavior and interactions in different contexts and (b) supports learning how to coordinate modes in different and sometimes infrequent contexts;
- use forcing functions;
- develop new patterns of coordination between human and machine agents;
- use machine intelligence to automate error detection.

Designers frequently fail to appreciate the cognitive and operational costs of more and more complex modes. Often, there are pressures and other constraints on designers that encourage mode proliferation. But in particular cases the benefits of increased functionality may be more than counterbalanced by the costs of learning about all the available functions, the costs of learning how to coordinate these capabilities in context, and the costs of mode errors.[26] Users frequently cope with the complexity of the modes by "re-designing" the system through patterns of use, e.g., few users may actually use more than a small subset of the resident options or capabilities (Rosson, 1983). However, a variety of pressures may still lead managers, designers, marketeers, and even practitioners to claim that there is a need for highly flexible systems with multiple capabilities, modes, and options.

Improved mental models

Mode errors tend to occur for two reasons: either the user misassesses the mode configuration of the system at a given point in time or he misses transitions (and the implications of such transitions) in mode status over time. The latter problem implies that the user does not pay attention to critical information at the right time. This occurs as a conjunction of several interacting factors. Knowledge factors can play a role in these breakdowns (cf., Sarter and Woods, 1992). One aspect of knowledge-related contributors seems to be gaps in practitioners' mental

[26] Unless these costs are rationalized away, i.e., the additional training burden ends up being accomplished through on-the-job training, and the costs of mode errors are attributed to the practitioner rather than to the design of the larger system.

model of the system. Another knowledge factor seems to be difficulties in learning how to coordinate and switch among the different modes and options in varying operational contexts—knowing how to work the system. The issue here is not that practitioners cannot use the system, but rather that they develop stereotypical methods and strategies based on the most frequently occurring situations. When events conspire to throw them off of these familiar methods and paths, a variety of troubles can arise (see Sarter and Woods, 1992; 1994; Cook et al., 1990; Cook, Woods, McColligan, and Howie, 1991 for more on this). Again, to achieve high reliability in human-machine systems, assessing operability in the context of situations that go beyond textbook cases is critical (see Chapter 4).

Improved feedback

Attentional dynamics play a critical role in mode problems. Given a busy environment (multiple tasks and monitoring for new events) and depending on the kind of feedback available, attention allocation strategies may not be sufficient, leading to missed mode transitions or to failures to appreciate the significance of a mode transition. Therefore, many have recommended that one way to reduce the risk of mode error is better display of mode status and behavior.

But the rub is determining what are better display techniques for mode status. Most displays of modes simply provide alphanumeric labels that designate current mode (many times with no positional, redundant, or analogical cues other than a propositional tag or label itself).[27] These are the types of displays that are typically available in the field studies that have documented mode error and awareness difficulties. To build effective indications of mode status, the first criterion is to develop displays that help practitioners detect and track mode *changes and transitions* (Norman, 1990a; Sarter and Woods, in press). In other words, highlight *events* not simply states; mode changes are important events which should stand out in any representation of the system (Woods, in press-b). Second, provide feedback that reveals the

[27] Unfortunately, we have, more than once, seen devices with hidden modes, e.g., Cook, Woods, McColligan, and Howie (1991).

implications of mode changes given the state of other inter-related factors and given possible future events or contingencies.

Third, accentuate the differences between modes. How do users know when to check mode status? Remember modes change more slowly than other task rhythms; practitioners are probably busy with other tasks and problems in situations where mode errors are particularly important relative to bad outcomes. Do not force monitoring behavior into an explicit decision to check whether mode status conforms to expectations. Do not force users to "read" the display closely (an act of focal attention) and invoke extensive knowledge of system function to interpret its significance in the current context every time they decide they should check on mode status. Use analogical representation techniques so that practitioners can simply apprehend mode status and changes as part of their scan of other system state variables (Woods, 1992; Woods, in press-b).

One approach to meeting the above criteria is to provide cues to signal mode status, automatic system activity, and mode changes that can be picked up through orienting perceptual systems (aural, kinesthetic or peripheral vision).[28] Monk (1986) suggests that the visual channel is not a good choice for conveying mode information in environments that already require considerable visual processing. Adding yet another visual source of information in such environments further challenges attentional dynamics related to knowing where to focus attention when (however, this discounts the possibilities for visual displays that support peripheral access; Woods, 1992). Such visual overload could result in problems caused by the need for making tradeoff decisions about what channels to attend to. It is also not advisable to use visual feedback that requires an act of focal attention to pick up the significance if it is not clear that the operator will be continuously attending to the display. Monk suggests that aural feedback (e.g., keying-contingent sound) might be a more useful feedback modality for such an environment. But auditory channels can be loaded as well. Aural feedback may also be too intrusive, forcing shifts of conscious attention which may be too distracting for mode indications relative to other activities in a time-shared multi-task environment.

[28] For an analogy, think of how we are implicitly aware of how we are physically situated in the world, e.g., indoors/outdoors, orientation, surfaces for support, without necessarily invoking conscious attention to these.

Another technique is to use kinesthetic feedback to increase mode awareness (Sellen et al., 1992). Their research focused on systems that involve only a limited number of modes which may not transfer to systems with more complex mode structures. However, the basic idea of using an otherwise free channel for mode information seems to be promising. In the aviation domain, kinesthetic feedback is successfully used for stall warnings, in which case the so-called "stick-shaker" (i.e., yoke vibration) warns the pilot of an imminent stall. This kind of feedback is difficult to ignore but may not disrupt other ongoing activities (e.g., communication with ATC).

Another dimension along which feedback can vary is in terms of who is generating it. Sellen et al. claim that user-maintained feedback is preferable to maintain mode awareness. It is questionable, however, whether this additional burden and responsibility are acceptable in most real-world settings. In many cases, system-provided feedback may be the only choice, and it will therefore be important to find ways to improve the communicative skills of machine agents.

Another direction for improved feedback would be to provide better indications of the consequences of mode changes for future system behavior. Displays could project how the automated mode will behave or control the underlying process relative to other armed or relevant modes or environmental conditions. A mode change in a highly automated system can be also a change in the mode of control. For example, one mode transition in the case of an automated aircraft is also a change from control of aircraft speed by pitch to control of speed by thrust. Indicating the change in controlling parameter, the new constraints that are relevant (e.g., the target or limit values may change as well), or the future behavior of the system based on the new mode, all could be useful ways of supporting error detection or failure detection.

Another concept is to provide displays that capture past instructions to the automation and the corresponding system behavior. Such "history of interaction" displays could provide a visual trace of past and even projected system behavior under the current mode configuration. Visualizations of the history of changes in mode configuration could support practitioners in the timely detection of future problems and of mode errors.

Forcing functions

Forcing functions constrain a sequence of user actions along particular paths. These constraints are designed to reduce the chances of specific actions leading to poor outcomes (Lewis and Norman, 1986). Forcing functions can take a variety of forms as pointed out by Lewis and Norman. The system can prevent the user from expressing impossible intentions (a "gag" response), it can react to illegal actions by doing nothing, or it can guess or explore with the user what the user's intention was and then help translate this intention into a legal action ("Self-correct," "Teach me," or "Let's Talk About It" styles). The problem with such forcing functions with respect to mode error is that they require (a) that there is only one legal action or strategy for each intention or (b) that a system is capable of inferring the user's intention so that it can judge the acceptability of the practitioner's actions. Such a system would also require access to information on the overall situation and context which may determine whether an action is appropriate. Without these capabilities, it would have to question almost any action, just in case, and it would become a nagging advisor, second-guessing all actions.

Coordinating human and machine agents

Aiding mode awareness is also connected to issues about how to coordinate the activities of human practitioners and machine agents (Billings, 1991). One approach that has been suggested is "management by consent" which requires that all members of the human team agree to any change in modes before it is activated. This approach could help the operators to build a memory trace of all prior system interactions. This should enable them to better predict future system behavior. The problem with this technique is that it involves the "dilemma of delegation." If automation and team work are supposed to reduce the burden on the operator by taking over and sharing tasks, then it seems counterproductive to require that all input be checked and agreed to by every member of the team.

Another interesting approach might also be to eliminate any defaults in mode settings. Past mode settings should always be deactivated once a new target is entered, thus forcing the practitioner to deliberately select a desired configuration for any system behavior to occur. This would contribute to a more consistent command structure. While this would be a cumbersome approach, it might prevent errors due to a lack of awareness of past mode settings.

Overall, what is important to note is how changes that seem to be just about technology (i.e., automation) raise questions about the human's role and about the coordination of people and machine agents in a distributed cognitive system. One does not just design a computerized artifact; one also is designing the operating conditions for a distributed cognitive system.

Mode Error as Designer Error

Mode error illustrates how the costs of clumsy use of technological possibilities are seen in "human error." Recognizing that such problems are symptoms of clumsy automation directs our attention to the people and organizations involved in design. What techniques will counter the proliferation of modes in the design process? How do we reduce designer "error" related to this aspect of the clumsy use of technology? Remember there are design-shaping aspects of the technology itself that encourage mode proliferation—the virtuality and keyhole properties—as well as the processes and organizational factors involved in developing computer-based devices (e.g., economic and marketing factors). Searching out each device with mode-related problems and developing custom countermeasures to combat mode error in that context are likely to be a very inefficient way to reduce mode error (we may not even be able to keep up the proliferation of ever more complex modes afforded by technological and other driving forces). How do we get designers to balance the costs of mode error against other costs and benefits in the design process? How do we make it easy for designers to use technological possibilities in ways that do not create the potential for mode errors? These questions pose interesting dilemmas for technologists, managers, and human factors researchers.

The Impact Diagram Revisited: The Case of Mode Error

At this stage we should be able to use the issue of mode error to walk through the Impact Diagram (Figure 9, p. 125). We have indicated how various properties of computer technology and the larger organizational context for design can encourage the proliferation of more complex device modes. One can suspect that a device, when fielded, will encourage the potential for mode error by examining its mode-related properties in relation to the demands of the field of practice—what modes are present, what mode transitions occur, are there indirect mode changes, how autonomous is the system in different modes, what situations include complicating factors that can challenge mode awareness, and how are the various modes and mode changes represented in these contexts?

The mode characteristics of the system shape the information processing involved in remembering and tracking device modes along with two other factors: the characteristics of the displays for indicating mode and the distributed set of agents who manage incidents. Mode proliferation has two kinds of impacts on the cognitive system. First, multiple modes increase memory and knowledge demands. One must know about the different modes, which actions do what in which mode, or which indications mean what in which mode. Second, mode proliferation increases demands on situation assessment and awareness as practitioners must keep track of what mode the device is in (a mode error is, in part, a breakdown in this situation assessment demand).

The difficulties in tracking device modes can vary depending on the task context (time pressure, interleaved multiple tasks, workload) and depending on how the interface depicts device mode. However, if the device also exhibits another of the flaws in computer-based representations that we noted earlier—not providing users with effective feedback about changes in the state of a device, automated system, or monitored process—then losing track of which mode the device is in may be surprisingly easy, at least in higher workload periods.

In terms of the behavior of people embedded within the operational system, the questions of interest include these: Do mode errors occur? What contextual factors contribute to their occurrence (workload, distractions)? What factors affect practitioners' ability to recover from mode errors when they do occur?

Design is shaped by properties of the computer medium and by the organizational context in which design occurs. These factors make it easy for designers to proliferate modes. The computer medium makes it easy to put several virtual devices on a single physical platform (this is attractive because one can build or market a single device for a variety of customers or a variety of niches, e.g., the operating room, critical care medicine, and nursing homes). But such generic devices are likely to possess multiple modes and potentially complex interactions between modes. Automation is another aspect of technology change that can lead to mode-related cognitive demands. When the mode of the system also can change in response to situation or system factors independent of practitioner input, the human practitioners face new demands for tracking system mode changes over time—mode awareness. As a result, we see new forms of breakdown: surprises created by indirect mode changes and errors of omission as well as commission in managing multiple modes.

The bottom line is that, as technological change proceeds, mode-related problems are becoming more and more commonplace (e.g., Lewis and Norman, 1986; Cook, Potter, Woods, and McDonald, 1991; Moll van Charante et al., 1993; Sarter and Woods, in press). Furthermore, we are already beginning to see incident and accident reports where mode-related problems are important contributing factors (e.g., Strasbourg: Monnier, 1992; or Bangalore: Lenorovitz, 1990).

Let us explore some of the reactions to one of these incidents (the Strasbourg aircraft crash) to see more about the dynamic that links technology change to error. As we described earlier., the accident investigation reports indicate a mode error in pilot interaction with cockpit automation seems to have been an important factor. The crew apparently entered a number thinking that the automation was in an *angle* of descent mode when it was actually in a *rate* of descent mode; their entry, 33, was interpreted as an instruction to fly at a rate of descent of 3300 feet/minute rather than the intended 3.3 degrees angle of descent. The crew's inability to detect the mode error within the time available, less than a minute, also played a role in the accident as well as a variety of other contributing factors (Monnier, 1992).

Following the accident several people in the aviation industry noted a variety of previous incidents in which similar mode errors had occurred:

> Firstly, British Airways had had an incident early in its A320 operation when the aircraft had inadvertently been flown on Rate of Descent when the pilots thought they were flying Flight Path Angle. This resulted in a ground proximity warning and subsequent go-around. . . . It then came to light that another operator had two similar incidents on record . . . (Seaman, 1992, p. 3).[29]

This captures the latent failure chain illustrated in Figure 9 (p. 125). The proliferation of modes and the opaque indications of the state and the behavior of the automation (Norman, 1990b; Sarter and Woods, in press) creates or exacerbates cognitive demands. Sometimes, given demanding and busy task contexts, a breakdown occurs—a mode error. This triggers a need for error detection and recovery. However, the automated system provides only weak feedback about the mode configuration, given the cognitive context, and provides little or no indication of the consequences of this mode configuration for the actual flight context. In other words, there is weak feedback and low observability in the interaction between the crew and their automated partners. Given that this erroneous action occurs, other contributors are necessary for the incident to evolve further along the path to disaster. The incidents on record reveal other cases where events proceeded far enough to be picked up by the formal incident-reporting mechanisms in this industry.

However, note that in the aftermath of the accident it is easy to focus on the particular manifestation of the mode-related problems. The external appearance of the error (phenotype) was a confusion of rate of descent and flight path angle modes. In one sense this is totally appropriate—one specific path for mode errors with safety consequences to occur in this setting is confusing these particular modes (given the current form of interaction and feedback between crew and automation). But in another sense this response is incomplete. Remember, incident-

[29] Remarks by Captain C. Seaman, Head of Safety, British Airways.

reporting systems had picked up several precursor incidents. If one can only interpret incidents in terms of phenotypes, then it is very difficult to see incidents as precursors of larger troubles, should other factors go wrong at the same time. Seeing these incidents as an indicator of the general question of mode-related problems in crew-automation interaction could spur a deeper examination of the potential for these kinds of problems throughout the human-machine system.

There is another reason why it is important to see the deeper error-related categories, or genotypes, indicated by specific data or incidents. The inherent variability of real systems and environments means that the particular incidents that have been observed may not be direct indicators of a particular vulnerability, but rather indicators of a type of problem that can contribute to incident evolution towards disaster. Wagenaar and Reason (1990) discuss this type/token problem at greater length.

The latent failure map depicted in the Impact Flow diagram (Figure 9, p. 125) points out that operational systems are adaptive. Practitioners attempt to adapt their behavior and to shape the artifacts they interact with to meet their responsibilitics and goals.

> I subsequently learnt that our own Training Captains had developed some ad hoc specific preventative training to avoid just this sort of event, even though there was a marked reluctance on the part of the BA A320 pilots, that I met, to acknowledge that there might be a shortcoming in ergonomic design (Seaman, 1992, p. 3).

People in operational systems have some perception of the hazards that affect their ability to meet their responsibilities and goals, such as clumsily designed technology. Based on these incomplete perceptions, people attempt to adapt through the means available to them. In this case, people in the training department developed some specific things that they thought might address the specific problem.

But there is danger if one only sees the specific external form of the error (the phenotype of the erroneous action).

> In other words, we are using the adaptability of the human being to make up for a shortfall in the system, a shortfall which sits there as a trap ready to catch a poor unsuspecting soul who may one day find it as part of an accident chain (Seaman, 1992, p. 4).

Thus, it is easy to fall back on individual people as causal units, rather than examine the larger system in which they are embedded. The incidents, accidents, and data from studies point to larger dynamics in pilot interaction with current cockpit automation, such as mode-related problems, that have ". . . much more to do with our failure, as an industry, to appreciate, recognize, and correct some of the traps that we were laying for flight crews to fall into" (Seaman, 1992, p. 4).

The next few sections explore further the issues about how practitioners adapt to accommodate new technology, and about why it seems so hard to appreciate the significance of flaws in human-computer cooperation.

Tailoring Tasks and Systems

Practitioners Adapt to Accommodate New Technology

In developing new information technology and automation, the conventional view seems to be that new technology makes for better ways of doing the *same* task activities. We often act as if domain practitioners were passive recipients of the "operator aids" that the technologist provides *for* them. However, this view overlooks the fact that the introduction of new technology represents a change from one way of doing things to another.

> The design of new technology is always an intervention into an ongoing world of activity. It alters what is already going on—the everyday practices and concerns of a community of people—and leads to a resettling into new practices. . . (Flores et al., 1988, p. 154).

Practitioners are not passive in this process of accommodation to change. Rather, they are an active *adaptive* element in the person-machine ensemble, usually the critical adaptive portion (e.g., Hutchins, 1990). Multiple studies have shown that practitioners adapt information technology provided for them to the immediate tasks at hand in a *locally* pragmatic way, usually in ways not anticipated by the designers of the information technology (Roth et al., 1987; Flores et al., 1988; Cook et al., 1990; Cook, Woods, McColligan, and Howie, 1991; Hutchins, 1990). Tools are shaped by their users. Or, to state the point more completely, artifacts are shaped into tools through skilled use in a field of activity. This process, in which an artifact is shaped by its use, is a fundamental characteristic of the relationship between design and use.

> There is always . . . a substantial gap between the design or concept of a machine, a building, an organisational plan or whatever, and their operation in practice, and people are usually well able to effect this translation. Without these routine informal capacities most organisations would cease to function (Hughes, Randall, and Shapiro, 1991, p. 319).

Studies have revealed several types of practitioner adaptation to the impact of new information technology, that Cook, Woods, McColligan, and Howie (1991) termed *system tailoring and task tailoring*. In system tailoring, practitioners adapt the device and context of activity to preserve existing strategies used to carry out tasks (e.g., adaptation focuses on the setup of the device, device configuration, how the device is situated in the larger context). In task tailoring, practitioners adapt their strategies, especially cognitive processing strategies, for carrying out tasks to accommodate constraints imposed by new technology.

System tailoring types of adaptations tend to focus on shaping the device itself to fit the strategies of practitioners and the demands of the field of activity. For example, in one study (Cook et al., 1990; Cook, Woods, McColligan, and Howie, 1991), practitioners set up the new device in a particular way to minimize their need to interact with the new technology during high criticality/tempo periods. This occurred despite the fact that the practitioners' configurations neutralized many of the putative advantages of the new system (the flexibility to perform greater numbers and kinds of data manipulation). Note that system tailoring frequently results in only a small portion of the "in principle" device functionality actually being used operationally.

Task tailoring types of adaptations tend to focus on how practitioners adjust their activities and strategies given constraints imposed by characteristics of the device. For example, serial display of data and the proliferation of windows create new data management tasks: (a) how to find related data through a narrow keyhole into a large virtual data space; (b) when and how to declutter the display as different views and windows accumulate. Practitioners may tailor the device itself, for example, trying to re-make it into a spatially dedicated, parallel-form display. But they may still need to tailor their activities. For example, they may need to learn when to schedule the new decluttering task (e.g., devising external reminders) to avoid being caught in a high criticality situation where their first need is to reconfigure the display so that they can "see" what is going on in the monitored process.

Task and system tailoring represent coping strategies for dealing with clumsy aspects of new technology. We have observed a variety of inter-related coping strategies employed by practitioners to tailor the system or their tasks (Roth et al., 1987; Cook et al., 1990; Cook, Woods, McColligan, and Howie, 1991; Sarter and Woods, 1991). One class of coping behaviors relates to workload management to prevent bottlenecks from occurring at high-tempo periods. For example, we have observed practitioners force device interaction to occur in low-workload periods to minimize the need for interaction at high-workload or high-criticality periods. We have observed practitioners abandon cooperative strategies and switch to single-agent strategies when the demands for communication with the machine agent are high, as often occurs during high-criticality and high-tempo operations.

Another class of coping strategies relates to spatial organization. We have consistently observed that users constrain "soft," serial forms of interaction and display into a spatially dedicated default organization.

Another consistent observation is that, rather than exploit device flexibility, we see practitioners externally constrain devices via ad hoc standards. Individuals and groups develop and stick with stereotypical routes or methods to avoid getting lost in large networks of displays, complex menu structures, or complex sets of alternative methods. For example, Figure 15 (p. 167) shows about 50% of the menu space for a computerized patient-monitoring information system used in operating rooms (Cook et al., 1990). We sampled physician interaction with the system for the first three months of its use during cardiac surgery. The highlighted sections of the menu space indicate the options that were actually used by physicians during this time period. This kind of data is typical—to cope with complexity, users throw away functionality to achieve simplicity of use tailored to their perceptions of their needs.

Studies of practitioner adaptation to clumsy technology consistently observe users invent "escapes"—ways to abandon high-complexity modes of operation and to retreat to simpler modes of operation when workload gets too high (Roth, et al., 1987; Cook et al., 1990; Cook, Woods, McColligan and Howie, 1991; Moll van Charante et al., 1993; Sarter and Woods, 1994).

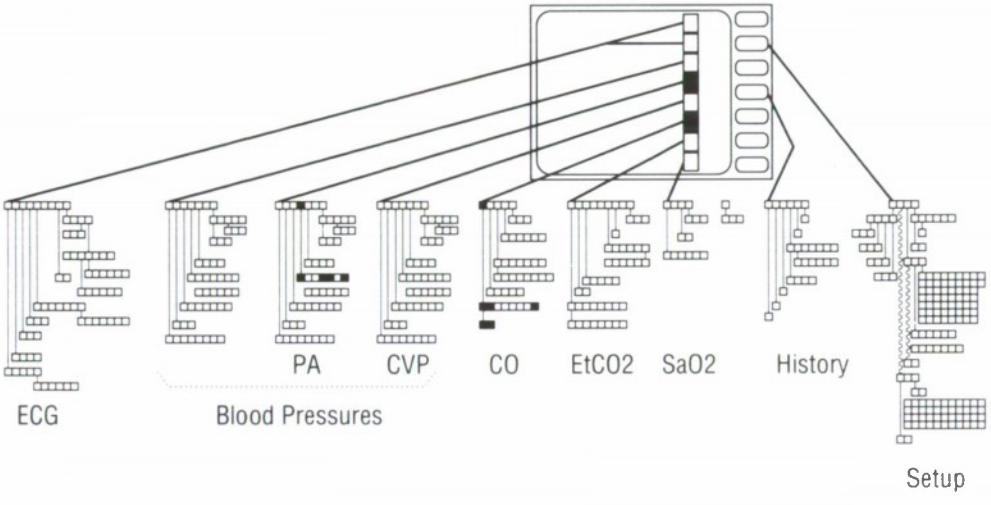
Finally, observations indicate that practitioners sometimes learn ways to "trick" automation, e.g., to silence nuisance alarms. Practitioners appear to do this in an attempt to exercise control over the technology (rather than let the technology control them) and to get the technology to function as a resource or tool for their ends (e.g., Roth et al., 1987).

Note these forms of tailoring are as much a group as an individual dynamic. Understanding how practitioners adaptively respond to the introduction of new technology and understanding the limits of their adaptations are critical for understanding how new automation creates the potential for new forms of error and system breakdown.

Brittle Tailoring as a Latent Failure

Practitioners (commercial pilots, anesthesiologists, nuclear power operators, operators in space control centers, etc.) are responsible, not just for device operation but also for the larger system and performance goals of the overall system. Practitioners tailor their activities to insulate the larger system from device deficiencies and peculiarities of the technology. This occurs, in part, because practitioners inevitably are held accountable for failure to correctly operate equipment, diagnose faults, or respond to anomalies even if the device setup, operation, and performance are ill-suited to the demands of the environment.

However, there are limits to a practitioner's range of adaptability, and there are costs associated with practitioners' coping strategies, es-



Copyright © 1990 R.I. Cook

Figure 15. How practitioners cope with complexity in computerized devices. This figure illustrates a portion of the menu space for a computerized patient monitoring information system. The highlighted areas are the items actually used by practitioners during observations of device use in cardiac surgery over three months. Note that the space of possibilities is very large compared with the portion practitioners actually use (From Cook et al., 1990).

pecially in non-routine situations when a variety of complicating factors occur (Woods, 1990a). These costs or limits represent a kind of latent failure in complex, high-consequence systems (Reason, 1990) whose effects are visible only when other events and circumstances combine with the latent failure to produce critical incidents. Thus, new burdens introduced by clumsy use of technology can create new pathways to disaster. Ironically, these types of incidents typically are labeled "human errors," while the human skills required to cope with the effects of these complexities are unappreciated except by the beleaguered practitioner. *Paradoxically, practitioners' adaptive, coping responses often help to hide the corrosive effects of clumsy technology from designers.*

Note the paradox: because practitioners are *responsible*, they work to smoothly accommodate new technology. As a result, practitioners' work to tailor the technology can make it appear smooth, hiding the clumsiness from designers.

We need to understand more about how practitioners adapt tools to their needs and to the constraints of their field of activity. These adaptations may be inadequate or successful, misguided or inventive, brittle or robust. When failures occur, understanding how those failures came about means understanding how the community of practitioners has tailored the artifacts and their strategies relative to the constraints of that field of activity. Research on computer-based artifacts should include investigations to understand the mutual shaping that goes on between practitioners and technological artifacts.

Why is it Hard to Find and to Appreciate the Significance of the Clumsy Use of Technology?

If flaws in the interaction of people and computer-based devices are so obvious in hindsight, why is it so hard to recognize them prior to the harsh glare of poor outcomes? If many of these flaws in computer-based devices are classic in that we see them repeated over and over again in different specific contexts and across different fields of activity, why then is it so easy to treat them as small local glitches rather than recognize them as a general type of design error?

Sources of Misattribution

The contribution of clumsy technology to incidents is easily missed because of a variety of factors. One is the belief that operahility is the responsibility of operational personnel and not the responsibility of designers. Because operators are responsible, they tailor the artifacts and their activities to try to make the system work. But their tailoring obscures the role of clumsy design. An outsider can easily focus on the contrast—operators usually make the system work; so failures can be attributed to the operators involved in a specific incident. However, a deeper understanding of the operational system uncovers the interacting constraints acting on it and uncovers how the operational system has adapted over time to try to meet the challenges and goals of the field of practice given the resources and constraints. Understanding the dynamics of the operational system (Figure 1, p. 21) and how it has adapted as a system to balance demands with resources and constraints usually reveals that a variety of systemic factors contribute to an incident, not simply "human error."

The complexity of individual incidents in particular fields of practice makes it easy to see them as idiosyncratic or unusual events and makes it easy to miss deeper contributors or larger trends. Even when design factors are raised, it is easy to rationalize why flaws in computer-based systems from an operahility point of view are small or unimportant contrihutors. For example, the HCI flaws can be rationalized as the "learning curve" for any new device. However, this misses the process by which operational systems adapt around clumsy computer-based systems. In addition, poor HCI is so common in many fields of practice that it becomes simply part of the job for practitioners to work around the clumsiness of these artifacts. For example, Cook, Woods, and McDonald (1991) examined all of the incidents that occurred over six months in one medical service. They found that when incidents involved clumsy computer-hased systems, factors related to human-computer interaction rarely were part of the institutional review. Instead, in cases involving clumsy technology, the focus was on the individuals involved and the need for them to adopt strategies to work around the clumsiness of these systems.

Myths About Cognitive Systems

Why is it so difficult to move beyond the person at the sharp end as a "cause" to an incident? In part, this difficulty is duc to some myths about how human cognition functions in context that can be grouped under the heading of equi-availability myths.

Equi-Availability Myth 1: If data is physically available, then its significance should be appreciated in all contexts.

This misses the fact that focusing in on the critical subset of relevant data, out of a very large field of available data, is a substantive cognitive activity in dynamic multi-task fields of activity (Woods, 1986; Woods, 1992; Woods, in press-a). In critical incidents it is usually the case that all of the data relevant in hindsight was physically available to the people at the sharp end, but the people did not find and interpret the right data at thc right time. In addition, this myth is a part of a general trend by some people to see "attention" as a motivational or effort factor (e.g., thc people involved in an incident didn't "try hard enough"). The term "vigilance" has been used in this sense in the anesthesiology community for example (Cook, Woods, and McDonald, 1991), in contrast to the technical sense of the term as a cognitive process and skill (Gopher, 1991).

Equi-Availability Myth 2: If people demonstrate knowledge in some context, then that knowledge should be available in all contexts.

Actually, the activation of knowledge-in-context is a fundamental cognitive process—human memory is a context-cued retrieval system (see the discussion of knowledge factors in Chapter 4). Education research has focused extensively on the problem of inert knowledge—knowledge that can be demonstrated in one context (e.g., test exercises) is not activated in other contexts where it is relevant (c.g., ill-structured problems). Inert knowledge consists of isolated facts that are disconnected from how the knowledge can be used to accomplish some purpose. This research emphasizes the necd to "conditionalize" knowledge to its use in different contexts as a fundamental tcaching strategy.

These two myths make it hard to see latent failures in human-machine system design. They are based on ignorance about the structure, function, and dynamics of cognitive systems. They fail to take into account how knowledge is activated in different contexts, how the focus of attention shifts when there are multiple channels and tasks to be juggled by practitioners, and how tradeoffs between different goals or possible outcomes are set when practitioners are faced with irreducible uncertainty and time pressure. These are the processes in distributed cognitive systems that govern the expression of expertise and of error.

Designer Error?

Finally, we should consider what the label of "designer error" means. It is easy to see the theme of this chapter as "designer error," rather than "operator error," is sometimes responsible for incidents. Or similarly, work on error and failure that implicates the importance of organizational factors (e.g., Reason, 1990) can easily be misinterpreted as simply substituting management error for operator error in many cases. Managers and designers and maintainers are human, as well. Operations is as much a distributed multi-agent system as management or design; management and design are just as much human activities as operations.

Labeling an incident as management or designer error risks the same traps as the indiscriminate use of the label operator error. They are all a form of assuming that the label "human error" is the end of an investigation rather than the beginning. Design failures, when recognized as such, can be governed by knowledge factors, attentional dynamics, strategic factors, demand-resource mismatches, and organizational constraints. The same factors govern the expression of error and expertise for designers and managers as well as for those embedded at the sharp end of systems.

6

THE COMPLEXITY OF ERROR

We have covered many different aspects of research on human error and the evolution of system failures up to this point. The results indicate that the story of human error is markedly complex (Rasmussen et al., 1987; Reason, 1990; Hollnagel, 1993). The story of error is complex because there are multiple contributors to an incident or disaster, each necessary but only jointly sufficient. Furthermore, the story of error is complex because:

- some of the contributors are latent, lying in wait for other triggering or potentiating factors,
- the human performance in question involves a distributed system of interacting people at the sharp end and organizational elements at the blunt end,
- the same factors govern the expression of both expertise and error,
- the context in which incidents evolve plays a major role in human performance at the sharp end,
- people at the blunt end create dilemmas and shape tradeoffs among competing goals for those at the sharp end, and
- the way technology is deployed shapes human performance, creating the potential for new forms of error and failure.

In this chapter, we will explore another factor that contributes to the complexity of error: the hindsight bias, which demonstrates that the attribution of error after-the-fact is a process of social and psychological judgment rather than an objective conclusion. We will explore the consequences of the hindsight bias for error analysis and conclude with some pointers about how to go behind the label "human error."

Evaluating Human Performance

Attributing System Failures to Practitioners

System failures, near failures, and critical incidents are the usual triggers for investigations of human performance. When critical incidents do occur, human error is often seen as a cause of the poor outcome. In fact, large complex systems can be readily identified by the percentage of critical incidents that are considered to have been caused by human error; the rate for these systems is typically about 75%. The repeated finding of about three-quarters of incidents arising from human error has built confidence in the notion that there is a human error problem in these domains. Indeed, the belief that fallible humans are responsible for large system failures has led many system designers to use more and more technology to try to eliminate the human operator from the system or to reduce the operator's possible actions so as to forestall these incidents.

Attributing system failure to the human operators nearest temporally and spatially to the outcome ultimately depends on the judgment by someone that the processes in which the operator engaged were faulty and that these faulty processes led to the bad outcome. Deciding which of the many factors surrounding an incident are important and what level or grain of analysis to apply to those factors is the product of *human* processes (social and psychological processes) of causal attribution. What we identify as *the cause of an incident* depends on with whom we communicate, on the assumed contrast cases or causal background for that exchange, and on the purposes of the inquiry.

For at least four reasons it is not surprising that human operators are blamed for bad outcomes. First, operators are available to blame. Large and intrinsically dangerous systems have a few, well identified humans at the sharp end. Those humans are closely identified with the system function so that it is unlikely that a bad outcome will occur without having them present. Moreover, these individuals are charged, often formally and institutionally, with ensuring the safe operation as well as the efficient functioning of the system. For any large system failure there will be a human in close temporal and physical relationship to the outcome (e.g., a ship captain, pilot, air traffic controller, physician, nurse).

The second reason that human error is often the verdict after accidents is that it is so difficult to trace backward through the causal chain of multiple contributors that are involved in system failure (Rasmussen,1986). It is particularly difficult to construct a sequence that goes past the people working at the sharp end of the system. To construct such a sequence requires the ability to reconstruct, in detail, the cognitive processing of practitioners during the events that preceded the bad outcome. The environment of the large system makes these sorts of reconstructions extremely difficult. Indeed, a major area of research is development of tools to help investigators trace the cognitive processing of operators as they deal with normal situations, with situations at the edges of normality, and with system faults and failures. The incidents described in Chapter 4 are unusual in that substantial detail about what happened, what the participants saw, and practitioner actions was available to researchers. In general, most traces of causality will begin with the outcome and work backwards in time until they encounter a human whose actions seem to be, in hindsight, inappropriate or sub-optimal. Because so little is known about how human operators process the multiple conflicting demands of large, complex systems, incident analyses rarely demonstrate the ways in which the actions of the operator made sense at the time.

The third reason that human error is often the verdict is paradoxical: human error is attributed to be the cause of large system accidents because human performance in these complex systems is so good. Failures of these systems are, by almost any measure, rare and unusual events. Most of the system operations go smoothly; incidents that occur do not usually lead to bad outcomes. These systems have come to be regarded as safe by *design* rather than by *control*. Those closely studying human operations in these complex systems are usually impressed by the fact that the opportunity for large-scale system failures is present all the time and that expert human performance is able to prevent these failures. As the performance of human operators improves and failure rates fall, there is a tendency to regard system performance as a marked improvemcnt in somc underlying quality of the system itself, rather than the honing of skills and expertise within the distributed operational system to fine edge. The studies of aircraft carrier flight operations by Rochlin ct al., (1987) point out that the qualities of human operators arc crucial to maintaining system performance goals and that, by most measures, failures should be occurring much more often than they do. As consumers of thc products from large complex systems such as health care, transportation, and defense, society is lulled by success into the bclief that these systems are intrinsically low-risk and that the expected failure rate should be zero. Only catastrophic failures receive public attention and scrutiny. The remainder of the system operation is generally regarded as unflawed because of the low overt failure rate, even though there arc many incidents that could become overt failures. Thorough accident analyses indicate that prior to an accident one can often find precursor incidents in which a similar set of circumstances or conditions arose, although the incident did not proceed as far along the accident chain.

This ability to trace backward with the advantage of hindsight is the fourth major reason that human error is so often the verdict after accidents. Studies have consistently shown that people have a tendency to judge the quality of a process by its outcome; information about outcome biases their evaluation of the process that was followed. Also, people have a tendency to consistently exaggerate what could have been anticipated in foresight (Fischhoff, 1975). Typically, hindsight bias in evaluations makes it seem that participants failed to account for information or conditions that should have been obvious[30] or behaved in ways that were inconsistent with thc (now known to bc) significant information. Thus, knowledge of a poor outcome biases the reviewer toward attributing failures to systcm opcrators. But to decide what would be obvious to practitioners in the unfolding problem requires investigating many factors about the evolving incident, the operational system and its organizational context such as the background of normal occurrences, routine practices, knowledge factors, attentional demands, strategic dilemmas, and other factors.

[30] When someone claims that something should have been obvious, hindsight bias is virtually always present.

The psychological and social processes involved in judging whether or not a human error occurred are critically dependent on knowledge of the outcome, something that is impossible before the fact. Indeed, *it is clear from the studies of large system failures that hindsight bias is the greatest obstacle to evaluating the performance of humans in complex systems.*

The Biasing Effect of Outcome Knowledge

Outcome knowledge influences our assessments and judgments of past events. These hindsight or outcome biases have strong implications for how we study and evaluate accidents, incidents, and human performance.

As mentioned in Chapter 2, whenever one discusses human error, one should distinguish between *outcome* failures and defects in the problem-solving *process*. Outcome failures are defined in terms of a categorical shift in consequences on some performance dimension. Generally, these consequences are directly observable. Outcome failures necessarily are defined in terms of the language of the domain, e.g., for anesthesiology sequelae such as neurological deficit, reintubation, myocardial infarction within 48 hours, or unplanned ICU admission. Military aviation examples of outcome failures include an unfulfilled mission goal, a failure to prevent or mitigate the consequences of some system failure on the aircraft, or a failure to survive the mission. An outcome failure provides the impetus for an accident investigation.

Process defects, on the other hand, are departures from some standard about *how* problems should be solved. Generally, the process defect, if uncorrected, would lead to, or increase the risk of, some type of outcome failure. Process defects can be defined in domain terms. For example in anesthesiology, some process defects may include insufficient intravenous access, insufficient monitoring, regional versus general anesthetic, and decisions about canceling a case. They may also be defined psychologically in terms of deficiencies in some cognitive or information processing function: for example, activation of knowledge in context, mode errors, situation awareness, diagnostic search, and goal tradeoffs.

People have a tendency to judge a process by its outcome. In the typical study, two groups are asked to evaluate human performance in cases with the same descriptive facts but with the outcomes randomly assigned to be either bad or neutral. Those with knowledge of a poor outcome judge the same decision or action more severely. This is referred to as the *outcome bias* (Baron and Hershey, 1988) and has been demonstrated with practitioners in different domains. For example, Caplan, Posner, and Cheney (1991) found an inverse relationship between the severity of outcome and anesthesiologists' judgments of the appropriateness of care. The judges consistently rated the care in cases with bad outcomes as substandard while viewing the same behaviors with neutral outcomes as being up to standard even though the care (i.e., the preceding human acts) was identical. Similarly, Lipshitz (1989) found the outcome bias when middle-rank officers evaluated the decisions made by a hypothetical officer. Lipshitz points out that judgment by outcomes is a fact of life for decision makers in politics and organizations. In other words, the label "error" tends to be associated with negative outcomes.

It may seem reasonable to assume that a bad outcome stemmed from a bad decision, but information about the outcome is actually irrelevant to the judgment of the quality of the process that led to that outcome (Baron and Hershey, 1988). The people in the problem do not intend to produce a bad outcome (Rasmussen et al., 1987). Practitioners at the sharp end are responsible for action when the outcome is in doubt and consequences associated with poor outcomes are highly negative. If they, like their evaluators, possessed the knowledge that their process would lead to a bad outcome, then they would use this information to modify how they handled the problem. *Ultimately, the distinction between the evaluation of a decision process and evaluation of an outcome is important to maintain because good decision processes can lead to bad outcomes and good outcomes may still occur despite poor decisions.*

Other research has shown that once people have knowledge of an outcome, they tend to view the outcome as having been more probable than other possible outcomes. Moreover, people tend to be largely unaware of the modifying effect of outcome information on what they believe they could have known in foresight. These two tendencies collectively have been termed thc *hindsight bias*. Fischhoff (1975) originally demonstrated the hindsight bias in a set of experiments that compared foresight and hindsight judgmcnts conccrning the likelihood of particular socio-historical events. Basically, the bias has been demonstrated in thc following way: participants are told about some event, and some are provided with outcome information. At least two different outcomes are used in order to control for one particular outcome bcing a priori morc likely. Participants are then asked to estimate the probabilities associated with the several possible outcomes. Participants given thc outcomc information are told to ignore it in coming up with their estimates, i.e., to respond as if they had not known the actual outcome, or in some cases are told to respond as they think others without outcome knowledge would respond. Those participants with the outcomc knowlcdge judge thc outcomes they had knowledgc about as more likely than the participants without the outcomc knowlcdge.

The hindsight bias has proven to bc robust; it has bcen demonstratcd for different types of knowledge: episodes, world facts (e.g., Wood, 1978; Fischhoff, 1977), and in some rcal-world settings. For example, several researchers have found that mcdical practitioners exhibitcd a hindsight bias when rating the likelihood of various diagnoses (cf., Fraser, Smith, and Smith, 1992).

Experiments on the hindsight bias havc shown that (a) pcople overestimate what they would have known in forcsight, (b) they also ovcrestimate what others knew in foresight (Fischhoff, 1975), and (c) thcy actually misremember what they themsclvcs knew in foresight (Fischhoff and Beyth, 1975).[31]

Fischhoff (1975) postulatcd that outcomc knowledge is immediatcly assimilated with what is alrcady known about the event. A process of retrospective sense-making may be at work in which the whole event, including outcome, is constructed into a coherent whole. This process could result in information that is consistent with the outcome being given more weight than information inconsistent with it.

[31] This misremembering may be linked to the work on reconstructive memory, in which a person's memories can be changed by subsequent information, e.g., leading questions may change eyewitnesses memories; see Loftus, E. (1979). *Eyewitness testimony*. Cambridge, MA: Harvard University Press.

> It appears that when we receive outcome knowledge, we immediately make sense out of it by integrating it into what we already know about the subject. Having made this reinterpretation, the reported outcome now seems a more or less inevitable outgrowth of the reinterpreted situation. Making sense out of what we are told about the past is, in turn, so natural that we may be unaware that outcome knowledge has had any effect on us. . . . In trying to reconstruct our foresightful state of mind, we will remain anchored in our hindsightful perspective, leaving the reported outcome too likely looking (Fischhoff, 1982, p. 343).

It may be that judges rewrite the story so that the information is causally connected to the outcome. A study by Wasserman, Lempert, and Hastie (1991) supports this idea. They found that people exhibit more of a hindsight bias when they are given a causal explanation for the outcome than when the outcome provided is due to a chance event (but see Hasher, Attig, and Alba, 1981, for an alternative explanation; see Hawkins and Hastie, 1990, for a summary).

Taken together, the outcome and hindsight biases have strong implications for error analyses.

- Decisions and actions having a negative outcome will be judged more harshly than if the same process had resulted in a neutral or positive outcome. We can expect this result even when judges are warned about the phenomenon and have been advised to guard against it (Fischoff, 1975, 1982).
- Judges[32] will tend to believe that people involved in some incident knew more about their situation than they actually did. Judges will tend to think that people should have seen how their actions would lead up to the outcome failure. Typical questions a person exhibiting the hindsight bias might ask are these: "Why didn't they see what was going to happen? It was so obvious!" Or, "How could they have done *X*? It was clear it would lead to *Y*!"

Hence it is easy for observers after-the-fact to miss or underemphasize the role of cognitive, design, and organizational factors in incident evolution. For example, a mode error was probably an important contribu-

[32] We use this term to mean any person who judges some action or decision.

tor to the Strasbourg crash of an Airbus A-320. As we have seen, this error form is a human-machine system breakdown that is tied to design problems. Yet people rationalize that mode error does not imply the need for design modifications:

> While you can incorporate all the human engineering you want in an aircraft, it's not going to work if the human does not want to read what is presented to him, and verify that he hasn't made an error.[33]

Similarly, in the aftermath of the AT&T's Thomas Street outage, it is easy to focus on individuals at the sharp end and ignore the larger organizational factors.

> Its terrible ah the incident in New York was (pause) all avoidable. The alarms were were ah ah disarmed; no one paid attention to the alarms that weren't disarmed; that doesn't have anything to do with technology, that doesn't have anything to do with competition, it has to do with common sense and attention to detail.[34]

In this case, as in others, hindsight biases the judgment of the commentator. A detailed examination of the events leading up to the Thomas Street outage shows how the alarm issue is, in part, a red herring and clearly implicates failures in the organization and management of the facility (see FCC, 1991).

In effect, judges will tend to *simplify* the problem-solving situation that was actually faced by the practitioner. The dilemmas facing the practitioner *in situ*, the uncertainties, the tradeoffs, the attentional demands, and the double binds, all may be under-emphasized when an incident is viewed in hindsight. A consideration of practitioners' resources and the contextual and task demands that impinge on them is crucial for understanding the process involved in the incident and for uncovering process defects.

[33] Remarks by Y. Benoist, Director of Flight Safety, Airbus Industrie, 1992; quoted in Lenorovitz (1992b).

[34] Remarks by Richard Liebhaber of MCI commenting on AT&T's Thomas Street outage that occurred on September 17, 1991; from the MacNeil-Lehrer Report, PBS.

In summary, these biases play a role in how practitioners' actions and decisions are judged after-the-fact. The biases illustrate that attributing human error or other causes (e.g., software error) for outcomes is a psychological and social process of judgment. These biases can lead us to summarize the complex interplay of multiple contributors with simple labels such as lack of attention or willful disregard. These biases can make us miss the underlying factors which could be changed to improve the system for the future, e.g., lack of knowledge or double binds induced by competing goals. Furthermore, the biases illustrate that the situation of an evaluator after-the-fact who does not face uncertainty and risk, and who possesses knowledge of outcome, is fundamentally different from that of a practitioner in an evolving problem.

So whenever you hear someone say (or feel yourself tempted to say) something like: Why didn't they see what was going to happen? It was so obvious! or, How could they have done *X*? It was clear it would lead to *Y*! Remember that error is the starting point of an investigation; remember that the error investigator builds a model of how the participants behaved in a locally rational way given the knowledge, attentional demands, and strategic factors at work in that particular field of activity. This is the case regardless of whether one is attributing error to operators, designers, or managers. In other words, it is the responsibility of the error investigator to explore how it could have been *hard* to see what was going to happen or *hard* to project the consequences of an action. This does not mean that some assessments or actions are not clearly erroneous. But adoption of the local rationality perspective is important to finding out how and why the erroneous action could have occurred. A local rationality analysis is essential to go beyond the usual window dressing of blame and train, a little more technology will be enough, or only follow the rules recommendations in order to develop effective countermeasures.

Some research has addressed ways to debias judges. Simply telling people to ignore outcome information is not effective (Fischhoff, 1975). In addition, telling people about the hindsight bias and to be on guard for it does not seem to be effective (Fischhoff, 1977; Wood, 1978). Strongly discrediting the outcome information can be effective (Hawkins and Hastie, 1990), although this may be impractical for conducting accident analyses.

The method that seems to have had the most success is for judges to consider alternatives to the actual outcome. For example, the hindsight bias may be reduced by asking subjects to explain how each of the possible outcomes might have occurred (Hoch and Lowenstein, 1989). Another relatively successful variant of this method is to ask people to list reasons both for and against each of the possible outcomes (von Winterfeldt and Edwards, 1986; Fraser et al, 1992).[35] This is an example of the general problem-solving strategy of considering alternatives to avoid premature closure.

This work has implications for debiasing judges in accident analysis. But first we need to ask the basic question: What standard of comparison should we use to judge processes (decisions and actions) rather than outcomes?

Standards for Assessing Processes Rather Than Outcomes

We have tried to make clear that one of the recurring problems in studying error is a confusion over whether the label is being used to indicate that an outcome failure occurred or that the process used is somehow deficient. The previous section showed that outcome knowledge biases judgments about the processes that led to that outcome. But it seems common sense that some processes are better than others for maximizing the chances of achieving good outcomes regardless of the presence of irreducible uncertainties and risks. And it seems self-evident that some processes are deficient with respect to achieving good outcomes, e.g., relevant evidence may not be considered, meaningful options may not be entertained, contingencies may not have been thought through. But how do we evaluate processes without employing outcome information? How do we know that a contingency should have been thought through except through experience? This is especially difficult given the infinite variety of the real world, and the fact that all systems are resource-constrained. Not all possible evidence, all possible hypotheses, or all possible contingencies can be entertained by limited resource systems. So the question is: What standards can be used to determine when a process is deficient?

[35] This technique is in the vein of a Devil's Advocate approach, which may be a promising approach to guard against a variety of breakdowns in cognitive systems (see Schwenk and Cosier, 1980).

There is a loose coupling between process and outcome; not all process defects are associated with bad outcomes, and good process cannot guarantee success given irreducible uncertainties, time pressure, and limited resources. But poor outcomes are relatively easy to spot and to aggregate in terms of the goals of that field of activity (e.g., lives lost, radiation exposure, hull losses, reduced throughput, costs, lost hours due to injuries). Reducing bad outcomes generally is seen as the ultimate criterion for assessing the effectiveness of changes to a complex system. However, the latent failure model of disasters suggests that measuring the reliability of a complex, highly coupled system in terms of outcomes has serious limitations. One has to wait for bad outcomes (thus one has to experience the consequences). Bad outcomes may be rare (which is fortunate, but it also means that epidemiological approaches will be inappropriate). It is easy to focus on the unique and local aspects of each bad outcome obscuring larger trends or risks. Bad outcomes involve very many features, factors, and facets.Which were critical? Which should be changed?

If we try to measure the processes that lead to outcomes, we need to define some standard about how to achieve or how to maximize the chances for successful outcomes given the risks, uncertainties, tradeoffs, and resource limitations present in that field of activity. The rate of process defects may be much more frequent than the incidence of overt system failures. This is so because the redundant nature of complex systems protects against many defects. It is also because the systems employ human operators whose function is, in part, to detect such process flaws and adjust for them before they produce bad outcomes (Incident #2 in Chapter 4 is an example of this).

Process defects can be specified locally in terms of the specific field of activity (e.g., these two switches are confusable). But they also can be abstracted relative to models of error and system breakdown (this erroneous action or system failure is an instance of a larger pattern or syndrom–mode error, latent failures, etc.). This allows one to use individual cases of erroneous actions or system breakdown, not as mere anecdotes or case studies, but rather as individual observations that can be compared, contrasted, and combined to look for, explore, or test larger concepts. It also allows for transfer from one specific setting to another to escape the overwhelming particularness of cases.

Standards for Evaluating Good Process

But specifying a process as defective in some way requires an act of judgment about the likelihood of particular processes leading to successful outcomes given different features of the field of activity. What dimensions of performance should guide the evaluation, e.g., efficiency or robustness; safety or throughput? This loose coupling between process and outcome leaves us with a continuing nagging problem. Defining human error as a form of process defect implies that there exists some criterion or standard against which the activities of the agents in the system have been measured and deemed inadequate. However, what standard should be used to mark a process as deficient?

We do not think that there can be a single and simple answer to this question. Given this, we must be very clear about what standards are being used to define error in particular studies or incidents; otherwise, we greatly retard our ability to engage in a constructive and empirically grounded debate about error. *All claims about when an action or assessment is erroneous in a process sense should be accompanied with an explicit statement of the standard used for defining departures from good process.*

One kind of standard about how problems should be handled is a *normative model of task performance*. This method requires detailed knowledge about precisely how problems should be solved, i.e., nearly complete and exhaustive knowledge of the way in which the system works. Such knowledge is, in practice, rare. At best, some few components of the larger system can be characterized in this exhaustive way. As a result, normative models rarely exist for complex fields of activity where bad outcomes have large consequences. There are great questions surrounding how to transfer normative models developed for much simpler situations to these more complex fields of activity (Klein et al., 1993). For example, laboratory-based normative models may ignore the role of time or may assume resource unlimited cognitive processing.

Another standard is the comparison of actual behavior to *standard operating procedures* (e.g., standards of care, policies, and procedures). These practices are mostly compilations of rules and procedures that are acceptable behaviors for a variety of situations. They include various protocols (e.g., the Advanced Cardiac Life Support protocol for cardiac arrest), policies (e.g., it is the policy of the hospital to have informed consent from all patients prior to beginning an anesthetic), and procedures (e.g., the chief resident calls the attending anesthesiologist to the room before beginning the anesthetic, but after all necessary preparations have been made).

Using standard procedures as a criterion may be of limited value either because they are codified in ways that ignore the real nature of the domain[36] or because the coding is underspecified and therefore too vague to use for evaluation. For example, one senior anesthesiologist replied, when asked about the policy of the institution regarding the care for emergent Caesarean-sections, "Our policy is to do the right thing." This seemingly curious phrase in fact sums up the problem confronting those at the sharp end of large, complex systems. It recognizes that it is impossible to comprehensively list all possible situations and appropriate responses because the world is too complex and fluid. Thus the person in the situation is required to account for the many factors that are unique to that situation. What sounds like a nonsense phrase is, in fact, an expression of the limitations that apply to all structures of rules, regulations, and policies (cf. e.g., Suchman, 1987; Roth et al., 1987).

One part of this is that standard procedures underspecify many of the activities and the concomitant knowledge and cognitive factors required to go from a formal statement of a plan to a series of temporally structured activities in the physical world (e.g., Roth et al., 1987; Suchman, 1987). As Suchman puts it, plans are resources for action, an abstraction or representation of physical activity. Procedures cannot, for both theoretical and practical reasons, completely specify all activity.

In general, procedural rules are underspecified and too vague to be used for evaluation if one cannot determine the adequacy of perfor-

[36] It is not unusual, for example, to have a large body of rules and procedures that are not followed because to do so would make the system intolerably inefficient. The work to rule method used by unions to produce an unacceptable slowdown of operations is an example of the way in which reference to standards is unrealistic. In this technique, the workers perform their tasks to an exact standard of the existing rules, and the system performance is so degraded by the extra steps required to conform to all the rules that it becomes non-functional (e.g., see Hirschhorn, 1993).

mance before the fact. Thus, procedural rules such as the anesthetic shall not begin until the patient has been *properly prepared* for surgery, or stop all *unnecessary* pumps are underspecified.[17] The practitioner on the scene must use contextual information to define when this patient is properly prepared or what pumps are unnecessary at this stage of a particular nuclear power plant incident. Ultimately, it is the role of the human at the sharp end to resolve incompleteness, apparent contradictions, and conflicts in order to satisfy the goals of the system.

A second reason for the gap between formal descriptions of work and the actual work practices is that the formal descriptions underestimate the dilemmas, interactions between constraints, goal conflicts, and tradeoffs present in the actual workplace (e.g., Cook, Woods, and McDonald, 1991; Hirschhorn, 1993). In these cases, following the rules may, in fact, require complex judgments as illustrated in the section on double binds (Chapter 4). Using standard procedures as a criterion for error may hide the larger dilemma created by organizational factors while providing the administrative hierarchy the opportunity to assign blame to operators after accidents (e.g., see Lauber, 1993 and the report on the aircraft accident at Dryden, Ontario; Moshansky, 1992).

Third, formal descriptions tend to focus on only one agent or one role within the distributed cognitive system. The operators' tasks in a nuclear power plant are described in terms of the assessments and actions prescribed in the written procedures for handling emergencies. But this focuses attention only on how the board operators (those who manipulate the controls) act during textbook incidents. Woods has shown through several converging studies of actual and simulated operator decision making in emergencies that the operational system for handling emergencies involves many decisions, dilemmas, and other cognitive tasks that are not explicitly represented in the procedures (see Woods et al., 1987, for a summary). Emergency operations involve many people in different roles in different facilities beyond the control room. For example, operators confront decisions about whether the formal plans are indeed relevant to the actual situation they are facing, and decisions about bringing additional knowledge sources to bear on a problem.

[17] These rules are taken from actual procedures used in anesthesiology and nuclear power emergencies respectively.

All these factors are wonderfully illustrated by almost any cognitive analysis of a real incident that goes beyond textbook cases. One of these is captured by a study of one type of incident in nuclear power plants (see Roth et al., 1992). In this case, in hindsight, there is a procedure that identifies the kind of problem and specifies the responses to this particular class of faults. However, handling the incident is actually quite difficult. First, as the situation unfolds in time, the symptoms are similar to another kind of problem with its associated procedures (i.e., the incident has a garden path quality; there is a plausible but erroneous initial assessment; see Chapter 4 for more on garden path problems). The relationship between what is seen, the practitioner's expectations, and other possible trajectories is critical to understanding the cognitive demands, tasks, and activities in that situation. Second, the timing of events and the dynamic inter-relationships among various processes contain key information for assessing the situation. This temporally contingent data is not well represented within a static plan, even if its significance is recognized by the procedure writers. Ultimately, to handle this incident, the operators must step outside of the closed world defined by the procedure system.

Standard practices and operating procedures may also miss the fact that for realistically complex problems there is often no one best method. Rather, there is an envelope containing multiple paths, each of which can lead to a satisfactory outcome (Rouse et al., 1984; Woods et al., 1987). Consider the example of an incident scenario used in a simulation study of cognition on the flightdeck in commercial aviation (Sarter and Woods, 1994; note that the simulated scenario was based, in part, on an actual incident). To pose a diagnostic problem with certain characteristics (e.g., the need to integrate diverse data, the need to recall and re-interpret past data in light of new developments, etc.), the investigators set up a series of events that would lead to the loss of one engine and two hydraulic systems (a combination that requires the crew to land the aircraft as soon as possible). A fuel tank is underfuelled at the departure airport, but the crew does not realize this, as the fuel gauge for that tank has been declared inoperative by maintenance. In any aircraft, there are standards for fuel management, i.e., how to feed fuel from the different fuel tanks to the engines. The investigators expected the crews to follow the standard procedures, which in this context would lead to the engine loss, the loss of one of the hydraulic systems, and the associated cognitive demands. And this is indeed what happened except for one crew. This one flight engineer, upon learning that one of his fuel tank gauges would be inoperative throughout the flight, decided to use a non-standard fuel management configuration to ensure that, just in case of any other troubles, he would not lose an engine or risk a hydraulic overheat. In other words, he *anticipated* some of the potential interactions between the lost indication and other kinds of problems that could arise and then shifted from the standard fuel management practices. Through this non-standard behavior, he prevented all of the later problems that the investigators had set up for the crews in the study.

Did this crew member commit an error? If one's criterion is departure from standard practices, then his behavior was erroneous. If one focuses on the loss of indication, the pilot's adaptation anticipated troubles that might occur and that might be more difficult to recognize given the missing indication. By this criterion, it is a successful adaptation. But what if the pilot had mishandled the non-standard fuel management approach (a possibility since it would be less practiced, less familiar)? What if he had not thought through all of the side effects of the non-standard approach, did the change make him more vulnerable to other kinds of troubles?

Consider another case, this one an actual aviation incident from 1991 (we condensed the following from an unpublished incident report to reduce aviation jargon and to shorten and simplify the sequence of events):

> Climbout was normal, following a night heavy weight departure under poor weather conditions, until approximately 24,000 ft when *numerous* caution/warning messages began to appear on the cockpit's electronic caution and warning system (CRT-based information displays and alarms about the aircrafts mechanical, electric, and engine systems). The first of these warning messages was OVHT ENG 1 NAC, closely followed by BLEED DUCT LEAK L, ENG 1 OIL PRESSURE, FLAPS PRIMARY, FMC L, STARTER CUTOUT 1, and others. Additionally, the #1 engine generator tripped off the line (generating various messages), and the #1 engine amber REV indi-

> cation appeared (indicating a #1 engine reverse). In general, the messages indicated a deteriorating mechanical condition of the aircraft. At approximately 26,000 ft, the Captain initiated an emergency descent and turnback to the departing airport. The crew, supported by two augmented crew pilots (i.e., a total of four pilots), began to perform numerous (over 20) emergency checklists (related to the various warnings messages, the need to dump fuel, the need to follow alternate descent procedures, and many others). In fact, the aircraft had experienced a serious pylon/wing fire. Significantly, there was no indication of fire in the cockpit information systems, and the crew did not realize that the aircraft was on fire until informed of this by ATC during the landing roll out. The crew received and had to sort out 54 warning messages on the electronic displays, repeated stick shaker activation, and abnormal speed reference data on the primary flight display. Many of these indications were conflicting, leading the crew to suspect number one engine problems when that engine was actually functioning normally. Superior airmanship and timely use of all available resources enabled this crew to land the aircraft and safely evacuate all passengers and crew from the burning aircraft.

The crew successfully handled the incident; the aircraft landed safely and passengers were evacuated successfully. Therefore, one might say that no errors occurred. On the other hand, the crew did not correctly assess the source of the problems, they did not realize that there was a fire until after touchdown, and they suspected number one engine problems when that engine was actually functioning normally. Should these be counted as erroneous assessments? Recall, though, that the display and warning systems presented an "electronic system nightmare" as the crew had to try to sort out an avalanche of low level and conflicting indications in a very high-workload and highly critical situation.[38]

[38] The incident occurred on a flight with two extra pilots aboard (the nominal crew is two). They had to manage many tasks in order to make an emergency descent in very poor weather and with an aircraft in deteriorating mechanical condition. Note the large number of procedures which had to be coordinated and executed correctly. How the extra crew contributed to the outcome or how well a standard sized crew would have handled the incident would be an interesting question to pursue using the neutral observer criteria (see the next section).

The above incidents help to exemplify several points. Assessing good or bad process is extremely complex; there are no simple answers or criteria. Standard practices and procedures provide very limited weak criteria for defining errors as bad process. What can one do then? It would be easy to point to other examples of cases where commentators would generally agree that the cognitive process involved was deficient on some score. One implication is to try to develop other methods for studying cognitive processes that provide better insights about why systems fail and how they may be changed to produce higher reliability human-machine systems (Rochlin et al., 1987; Reason, 1990).

Neutral Observer Criteria

The practitioners at the sharp end are embedded in an evolving context. They experience the consequences of their actions directly or indirectly. They must act under irreducible uncertainty and the ever-present possibility that in hindsight their responses may turn out wrong. As one critical care physician put it when explaining his field of medicine:

We're the ones who have to do something. It is their job to interpret situations that cannot be completely specified in detail ahead of time. Indeed, it is part of practitioners' tacit job description to negotiate the tradeoffs of the moment.

It is easy when blessed with the luxury of hindsight to lose the perspective of someone embedded within an evolving situation and experiencing the full set of interacting constraints that they must act under. But this is the perspective that we must capture if we are to understand how an incident evolved toward disaster. One technique for understanding the situated practitioner represents a third approach to develop a standard of comparison. One could use an empirical approach, one that asks what would other similar practitioners have thought or done in this situation? De Keyser and Woods (1990) called this kind of empirically based comparison the *neutral observer criterion*. To develop a neutral observer criterion, one collects *data* to compare practitioner behavior during the incident in question to the behavior of similar practitioners at various points in the evolving incident and in similar or contrasting cases. In practice, the comparison is usually accomplished by using the judgment of similar practitioners about how they would behave under similar circumstances. Neutral observers make judgments or interpretations about the state of the world, relevant possible future event sequences, and relevant courses of action. The question is whether the path taken by the actual problem-solver is one that is plausible to the neutral observers. One key is to avoid contamination by the hindsight bias; knowledge about the later outcome may alter the neutral observer's judgment about the propriety of earlier responses. One function of neutral observers is to help define the envelope of appropriate responses given the information available to the practitioner at each point in the incident. Another function is to capture the real dilemmas, goal conflicts, and tradeoffs present in the actual workplace. In other words, the purpose is to capture the ways that formal policies and procedures underspecify the demands of the field of practice.

An example occurred in regard to the Strasbourg aircraft crash (Monnier, 1992). Mode error in pilot interaction with cockpit automation seems to have been a contributor to this accident. Following the accident several people in the aviation industry noted a variety of precursor incidents for the crash where similar mode errors had occurred, although the incidents did not evolve as far toward negative consequences. This data provides us with information about what other similar practitioners have done, or would have done, when embedded in the context of commercial air transport. It indicates that a systemic vulnerability existed based on the design, rather than a simple case of human error.

Our research, and that of others, is based on the development of neutral observer criteria for actions in complex systems. This method involves comparing actions that wcre taken by individuals to those of other similar practitioners placed in the same or contrasting situation. Note that this is a strong criterion for comparison and it requires that the evaluators possess or gather the same sort of expertise and experience as was employed during the incident. It does not rely on comparing practitioner behaviors with theory, rules, or policies. It is particularly effective for situations where the real demands of the system are poorly understood and where the pacc of system activity level is fast (i.e., in large, complex systems).

Error Analysis as Causal Judgment

Error and accident analysis is one case where lay people, scientists, engineers, managers, or regulators make causal judgments or attributions. Causal attribution is a psychological and social judgment process that involves isolating one factor from among many contributing factors as a cause for the event to be explained. Strictly speaking, there are almost always several necessary and sufficient conditions for an event. But people distinguish among these necessary and sufficient conditions focusing on some as causes and relegating others to a background status as enabling conditions. In part, what is perceived as cause or enabling condition will depend on the context or causal background adopted (see Hart and Honore, 1959; also see Cheng and Novick, 1992). Consider a classic example used to illustrate this point. Oxygen is typically considered an enabling condition in an accident involving fire, as in the case of a dropped cigarette. However, people would generally consider oxygen as a cause if a fire broke out in a laboratory where oxygen was deliberately excluded as part of an experiment.

Current models of causal attribution processes hold that people attempt to explain the *difference* between the event in question and some contrasting case (or set of cases). Rather than explaining an event per se, one explains why the event occurs in the target case and not in some counterfactual *contrast case* (Hilton, 1990).[39] The critical point is that there are degrees of freedom in how an event, such as an accident, is explained, and the explanation chosen depends, in part, on the contrasting case or cases adopted. Thus, in a neutral observer approach, the investigator tries to obtain data on different kinds of contrast cases, each of which may throw into relief different aspects of the dynamics of the incident in question.

[39] Some relevant factors for establishing a causal background or contrast case are the dimensions originally proposed by Kelley (1973): consensus, distinctiveness, and consistency. Consensus refers to the agreement between the responses of other people and the response of a particular person regarding a particular stimulus on a particular occasion; distinctiveness refers to the disagreement between the particular person's responses to some particular stimulus and other stimuli on the particular occasion; and consistency refers to the agreement between the way a particular person responds to a particular stimulus on different occasions (see Cheng and Novick, 1992)

Interactional or contrast case models of causal attribution help us to understand the diversity of approaches and attitudes toward human error and disasters. If someone asks another person why a particular incident occurred and if the shared background between these people is that causes of accidents are generally major equipment failures, environmental stresses, or misoperation, then it becomes sensible to respond that the incident was due to human error. If one asks why did a partieular ineident oceur, when the shared background concerns identifying who is financially responsible (e.g., a legal perspeetive), then it becomes sensible to expect an answer that specifies the person or organization that erred. If questioner and respondent only appear to have a shared background (because both use the words human error) when they, in fact, have different frames of referenee for the question, then it is not surprising to find confusion.

In some sense, one could see the research of the 1980s on error as framing a different background for the question: Why did this ineident oeeur? The eausal baekground for the researehers involved in this intensive and eross-diseiplinary examination of error and disaster was: How do we develop higher reliability human-maehine systems? This eausal baekground helped to point these researehers toward system-level factors in the management and design of the complex processes. In addition, when this question is posed by social and behavioral seientists, they (not so surprisingly) find socio-technical contributors, as opposed to reliability engineers who pointed to a different set of factors (Hollnagel, 1993). The benefit of the soeio-teehnieal background as a frame of referenee for eausal attribution is that it heightens our ability to go beyond the attribution of human error in analysis of risk and in measures to enhance safety.[40]

The baekground for a neutral observer approaeh to analyzing eognitive process and error comes from the local rationality assumption, i.e., people do reasonable things, given their knowledge, objectives, point of view, and limited resources. However, an accident is by definition unintentional; people do not intend to act in ways that produce negative eonsequences (excepting sabotage). Error analysis traces the problem-solving process to identify points at which limited knowledge and

[40] It seems lo us thal psychological processes of causal attribution apply as well to researchers on human error as they do to non-behavioral scientisls. One could imagine a corollary to William Jamcs' Psychologists Fallacy in which psychologists supposc that they are immune from lhe psychological processes lhat they sludy.

processing lead to breakdowns. Process-tracing methods are used to map out how the incident unfolded over time, what the available cues were, which cues were actually noticed by participants, and how they were interpreted. Process tracing attempts to understand why the particular decisions/actions were taken, i.e., how did it make sense to the practitioners embedded in the situation (Woods, 1993a).

The relativistic notion of causal attribution suggests that we should seek out and rely on a broad set of contrast cases in explaining the sequence of events that led to an outcome. We explain why the practitioners did what they did by suggesting how that behavior could have been locally rational. To do this we need to understand behavior in the case in question relative to a variety of different contrast cases–what other practitioners would have done in the situation or in similar situations.

Error as Information

One of the themes reverberating throughout this book is that human error represents a symptom rather than a cause. In this view error is, in part, a form of information about the functioning of the system in which those people are embedded (Rasmussen, 1986). We can use the information role to go behind the label "human error" and learn about how to improve human-machine, socio-technical systems. Lanir (1986) has developed a framework that captures how organizations can react to disaster.

A Fundamental Surprise

On March 28, 1979, the U.S. nuclear industry and technologists were rocked by the Three Mile Island accident (TMI). The consternation that resulted was due to more than the fact that it was the worst nuclear accident up to that time or the radiological consequences per se. Rather, the accident is a case of what Lanir (1986) terms *fundamental surprise*.[41] A fundamental surprise, in contrast to a situational surprise, is a sudden revelation of the incompatibility between one's self-perception and his environmental reality. Examples include the launch of Sputnik for the U.S., and the Yom Kippur war for Israel.

[41] Perhaps the best way to grasp Lanir's concept of fundamental surprise is through an apocryphal story about Noah Webster, the well-known lexicographer (from Lanir, 1986). Lanir tells the story and then explains the concept this way: *(continued on p. 196)*

One day, he arrived home unexpectedly to find his wife in the arms of his servant. "You surprised me," said his wife. "And you have astonished me," responded Webster. Webster's precise choice of words captured an important difference between his situation and that of his wife.

One difference between surprise and astonishment is the different level of intensity associated with the two: astonishment is more powerful and extensive than surprise. Indeed, Mr. Webster's situation possesses an element of shock. His image of himself and his relations with his wife were suddenly and blatantly proven false. This was not the case for Mrs. Webster who, although surprised by the incident, still could maintain her image of herself, her environment, her husband, and the relations between them. Indeed, even if Mrs. Webster had taken all the steps she viewed as necessary to prevent the incident, she had to assume that there was some possibility of her unfaithfulness eventually being revealed. For Mrs. Webster, the failure was due to an external factor. Although she was uncertain about the external environment she was not uncertain about herself.

In contrast, Mr. Webster's astonishment revealed unrecognized uncertainty extending far beyond his wife, his servant, or other external factors. For him, comprehending the event's significance required a holistic re-examination of his self-perceptions in relation to his environment. Although this surprise offered Mr. Webster a unique opportunity for self awareness, it came at the price of refuting his deepest beliefs.

A second distinction between surprise and astonishment lies in one's ability to define in advance the issues for which one must be alert. Surprises relate to specific events, locations, and time frames. Their demarcations are clear. Therefore, it is possible, in principle, to design early warning systems to prevent them. In contrast, events providing astonishment affect broad scopes and poorly demonstrated issues. Mr. Webster's shocking incident revealed only the "tip of an iceberg."

Another distinction concerns the value of information. Mrs. Webster lacked one item of information which, had she had it in advance, would have allowed preventing her surprise: the information that her husband would return early that day. No single piece of information could have prevented Mr. Webster's astonishment. In most cases, the critical incident is preceded by precursors from which an outside observer could have deduced the state of the couple's relations. Such observers should be less prone to the tendency to interpret information in ways that suit one's own world view, belittling or even ignoring the diagnostic value of information that contradicts it.

A fourth distinction between fundamental surprise and astonishment is in the ability to learn from the event. For Mrs. Webster, the learning process is simple and direct. Her early warning mechanisms were ineffective. If given a second chance, she might install a mechanism to reduce the possibility of being caught in a similar situational surprise.

Mr. Webster might attempt an explanation that would enable him to comprehend it without having to undergo the painful process of acknowledging and alerting a flawed world view. For example, he might blame the servant for "attacking his innocent wife." If it were established that the servant was not primarily at fault, he might explain the incident as an insignificant, momentary lapse o , his wife's behalf. In more general terms, we may say that Mr. Webster's tendency to seek external, incidental reasons reflects the human tendency to behave as though astonishment is merely a surprise and, thus, avoid recognition of the need to experience painful "self" learning. Lanir refers to Mrs. Webster's type of sudden discovery as a "situational surprise" and Mr. Webster's sudden revelation of the incompatibility of his self-perception with this environmental reality as a "fundamental surprise."

The TMI accident was more than an unexpected progression of faults; it was more than a situation planned for but handled inadequately; it was more than a situation whose plan had proved inadequate. The TMI accident constituted a fundamental surprise in that it revealed a basic incompatibility between the nuclear power industry's view of itself and reality. Prior to TMI, the industry could and did think of nuclear power as a purely technical system where all problems were in the form of some engineering technical area or areas, and the solutions to these problems lay in those engineering disciplines. TMI graphically revealed the inadequacy of that world view because the failures were in the socio-technical system and not due to pure technical factors (a single equipment or mechanical flaw) or to a purely human failure (gross incompetence or deliberate failures).

Prior to TMI, the pre-planning for emergencies consisted of considering large equipment failures; however, it did not consider a compounding series of small failures interacting with inappropriate human assessments of the situation and therefore erroneous actions. Prior to TMI, risk analysis also focused on large machine failures, not on the concatenation of several small failures, both machine and human. The kind of interaction between human and technical factors that actually occurred was inconceivable to the nuclear industry as a whole prior to TMI.

The post-TMI nuclear industry struggled to cope with, and adjust to, the revelations of TMI. The process of adjustment involved the phases associated with fundamental surprise described by Lanir. First, the *surprise event* itself occurs. Second, reaction spills over the boundaries of the event itself to include issues that have little to do with the triggering event-*crises*. Third, these crises provide the opportunity for *fundamental learning* which, in turn, produces practical *changes* in the world in question. Finally, the changes are absorbed and a new equilibrium is reached.

The immediate investigations of the TMI accident focused heavily on the mutual interaction between technical systems and people. The proposed changes that resulted from these investigations addressed the basic character of the joint human-machine system. These included providing new kinds of representations of the state of the plant, restructuring the guidance for board operators on how to handle abnormal conditions, and restructuring the organization of people in various facilities and their roles in handling different problems created by accidents.

However, in the process of carrying through on these and other lessons learned, the U.S. nuclear industry shifted direction and treated the accident as if it was nothing more than a situational surprise. They began to focus on localized and purely technological solutions, what could be termed the fundamental surprise error, after Lanir's analysis (cf., Reason, 1990).[42] This occurred despite the fact that the revelations of TMI continued to re-occur in other major incidents in the U.S. nuclear industry (e.g., the Davis-Besse nuclear power plant incident, see US NRC, 1985) as well as in other risky technological worlds. While the post-TMI changes clearly have improved aspects of the socio-technical system through such things as new sensors, new analyses of possible accident conditions, new guidance on how to respond to certain accident conditions, and changes in emergency notification procedures, the basic socio-technical system for operating plants and responding to failures has not changed (Moray and Huey, 1988).

As this case illustrates, incidents and accidents are opportunities for learning and change. But learning from the fundamental surprise may be partial and ineffective. The fundamental surprise often is denied by those involved. They interpret or redefine the incident in terms of local and specific factors as if it were only a situational surprise. The narrower interpretation can lead to denial of any need to change or to attribution of the cause to local factors with well bounded responses–the fundamental surprise error.

The label human error is a good example of a narrow interpretation that avoids confronting the challenges raised by the fundamental surprise. The fundamental surprise associated with the failures of large complex systems is that one must look for reliability in the larger system of interacting people and machines (cf., recall the examples of human-machine system failures re-interpreted as simply human error cited earlier in this chapter). If the source of the incident is human error, then only local responses are needed which do not change the larger organization or system. Curing human error in this local sense only requires sanctions against the individuals involved, injunctions to

[42] In general, the fundamental surprise error is re-interpreting a fundamental surprise as merely a situational surprise which then requires no response or only a limited response. For the context of complex system failures, research results indicate that a specific version of this error is re-interpreting a human-machine *system* breakdown as being due to purely human factors.

try harder or follow the procedures more carefully, or some remedial adjustments in the training programs. Even more comfortable for the technologist is the thought that human error indicates that the people in the system are an unreliable component. This leads to the idea that just a little more technology will be enough (Woods, 1991), that purely technological responses without consideration of human-machine systems or larger organizational factors can produce high-reliahility organizations. As a result of these rationalizations, the opportunity to learn from the fundamental surprise is lost.

As in the case of TMI, disasters in a variety of industries have been and continue to be unforeseen. As in the case of TMI, these accidents point to the interaction of people, technology, and the larger organization in which practitioners at the sharp end are embedded (Reason, 1990). The Thomas St. network failure challenges the larger organization and management systems (FCC, 1991); the Strasbourg and Bangalore crashes (Monnier, 1992; Lenorovitz, 1990) point to the human-machine cognitive system and the problems that can arise in coordination between people and automatic systems with many interacting modes. Before the fact, the accidents are largely inconceivable to the engineering and technological communities. As a result, Wagenaar and Groeneweg (1987) and Lanir (1986) have termed these accidents as impossible, in the sense that the event is outside the closed world of a purely technical language of description. The challenge of fundamental surprise is to acknowledge these impossible events when they occur and to use them as sources of information for expanding the language of description. For us, the challenge is to expand the language of description to include systems of intertwined people and machines as in the cognitive system language used in Chapters 4 and 5.

What is Human Error, Anyway?

There are at least two different ways of interpreting human performance in complex systems. The conventional way views human performance as the source of errors that can he eliminated by restricting the range of human activity or eliminating the performer from the system. According to this view, human error is seen as a distinct category that can be counted and tabulated.

The second approach views human performance as the means for resolving the uncertainties, conflicts, and competing demands inherent in large, complex systems (Hollnagel, 1993). Regulatory bodies, administrative entities, economic policies, and technology development organizations can affect both the conflicts practitioners confront and the resources available to practitioners for resolving those conflicts. The analyses guided by this approach explicitly avoid the term "human error" because it obscures more than it reveals.

The label "human error" is a judgment made in hindsight. After the outcome is clear, any attribution of error is a social and psychological judgment process, not a narrow, purely technical, or objective analysis. Different judges with different background knowledge of the events and context, or with different goals, will judge the performance of human practitioners differently. In a real sense, then, for scientists and investigators *there is no such thing as human error* (cf. Hollnagel, 1993). Human error does not comprise a distinct category of human performance. Recognizing the limits of the label "human error" can lead us in new more fruitful directions for improving the performance of complex systems.

As the many incidents sprinkled throughout this book suggest, human performance is not simply either adequate or inadequate. Nor is it either faulty or fault-free. Rather, *human performance is as complex and varied as the domain in which it is exercised.* Credible evaluations of human performance must be able to account for all of the complexity that confronts the practitioner. This is precisely what most evaluations of human performance do not do; they simplify the situations and demands confronting practitioners until it is obvious that the practitioners have erred. By stripping away the complexities and contradictions inherent in operating these large systems, the evaluators (a) eliminate the richness of detail that might help to show how the activities of the practitioners were locally rational and (b) fail to see the bottlenecks and dilemmas that challenge practitioner expertise and skill.

So how should we view a failure in a large, complex system? If a bad outcome is seen as yet another incident involving one or more human errors by some practitioners (i.e., if we adopt the conventional view), what shall we do then? The options are few. We can try to train people to remediate the apparent deficiencies in their behavior. We can try to remove the culprits from the scene or, at least, prevent these kinds of people from becoming practitioners. We can try to police practitioner activities more closely.

However, many of the changes oecurring in large complex systems, including those made in the name of reducing the human error problem, may make these systems more brittle and inerease the apparent contribution of human error (Cook and Woods, 1994). In rcsponse to incidents, organizations generate more rules, regulations, policies, and procedures that make it more likely that practitioners will be found to have erred by post incident analyses (i.e., erred in the sense of being discrepant with some aspect of standard policies). Emphasis on increasing efficiency generates more pressure on praetitioners, exacerbating double binds. Increased use of technology can create new burdcns and complexities for already beleaguered practitioners, and create new modes of failure. Even the burgeoning volume of data and knowledge in every field of practice plays a role: for example, inereasing the likelihood of inert knowledge problems (Feltovieh et al., 1989). In the face of these pressures, a quality management system that steadfastly maintains that human error is the root cause of system failures can be relied on to generate a huge volume of error statistics.

This book suggests quite a different approach. System failures can be viewed as a form of information about the system in which people are embedded. They do not point to a single independent (and human) component (a culprit) as the source of failure. Instead, system failures indicate the need for an analysis of the decisions and actions of individuals and groups embedded in the larger system that provides resources and imposes constraints. To study human performance and system failure requires studying the funetion of the system in which practitioners are embedded. In general, failures tell us about situations where knowledge is not brought to bear effectively, where the attcntional demands are extreme, or where the *n-tuple* bind is created. Knowledge of these systemie features allows us to see how human behavior is shaped and to examine alternatives for shaping it differently.

In this view, the behavior that people, in hindsight, call human error is the end result of a large number of factors coming to bear at the sharp end of practice. Social and psychological processes of causal attribution lead us to label some practitioner actions as human error and to regard other actions as acceptable performance. Hindsight bias leads us to see only those forks in the road that practitioners decided to take—we see "the view from one side of a fork in the road, looking back" (Lubar, 1993, p. 1168). This view is fundamentally flawed because it does not reflect the situation confronting the practitioners at the scene. The challenge we face as evaluators of human performance is to re-construct what the view was like or would have been like had we stood on the same road.

The schema of knowledge factors, attentional dynamics, and strategic factors presented in Chapter 4 provides one means of categorizing the activities of teams of practitioners.[43] The model of large system failure arising from the concatenation of the consequences of multiple small latent conditions provides an explanation for the mysteriously unique appearance of failures. The latent failure model also explains the limited success achieved by the pursuit of first causes. It also suggests that the role of human practitioners in large systems may be, in part, to uncouple elements of the system to minimize the propagation of consequences from latent failures resident in the system (Perrow, 1984). The cognitive systems perspective integrates problem demands, cognitive factors at the sharp end, and the organizational factors which influence the tradeoffs and dilemmas faced by practitioners. The Impact Flow diagram (Figure 9, p. 125) shows how technology change, especially the clumsy use of technological possibilities, shapes the cognition and behavior of practitioners. The social and psychological processes of causal attribution provide a model for studying how people come to label some human assessments and actions human error. All of these concepts provide the means to go behind the label "human error."

If You Think You Have a Human Error Problem, What Should You Do?

How should one proceed if there is a perception of a human error problem to be investigated? Or more broadly, how should one proceed to develop high reliability organizations (Rochlin et al., 1987)?

[43] The team need not be entirely human; the same schema may be used for evaluating the performance of machine expert systems and the performance of teams of human and machine cognitive agents.

In many cases, the concerned parties are already investing effort to collect reports on the incidents that occur within the system in question (e.g., the Aviation Safety Reporting System at NASA; or see Boeing, 1993; Cooper et al., 1984). There are many issues associated with how to collect incident data well (e.g., anonymity for the reporters). But even when the collection mechanism is well tuned, tracking sets of accidents, by itself, will not be enough to help answer the important questions about high-reliability human-machine systems. One problem is that it may detect risks too late, after the costs of the consequences of accidents have been incurred. After the fact, with benefit of hindsight, we often look back and find precursor incidents and signals that could have indicated a pre-existing vulnerability. We need conceptual frameworks for seeing and appreciating the significance of such precursors. For example, the concept of mode error helps us understand a variety of specific erroneous actions and incidents on highly automated flightdecks such as the Strasbourg or Bangalore accidents (see Chapter 5). A second problem is that incident data bases typically are organized, indexed, and reported only in terms of the language of domain, which means they capture only the external expression or phenotype of erroneous actions. Thus, it is easy to see the risks too narrowly. One can get lost in the variety and particularness of the incidents captured in this way and miss larger, deeper patterns that are precursors to accidents.

If you perceive a human error problem, the first step is to recognize that the label in itself is no explanation and no guide to countermeasures. When you hear or are tempted to use the label human error, stop; whenever you are tempted to say, how could these practitioners (whether operators, designers, or managers) have been so blind or so ignorant or whatever, stop; remember erroneous actions are the starting point for an investigation.

But then, how should one proceed? Erroneous assessments and actions are symptoms about underlying mismatches in the operational system in question. Investigate the background of the erroneous actions to discover these mismatches. To do this, the investigation will need to look at more than just the error itself; the systems in which the practitioners are embedded need to be studied. Build a model of how the participants behaved in a locally rational way given the knowledge, attentional demands, and strategic factors at work in that particular field of activity (see Chapter 4). This means one must understand the knowledge, attentional, and strategic demands that operate in this field of activity normally, at the margins of normality, and in different kinds of abnormal conditions. In other words, investigate the larger system in which the incident occurred, and do not focus exclusively on the particular incident and participants.

Explore how it could have been hard to see what was going to happen or hard to project the consequences of an action. Find ways to avoid taking the position of the omniscient observer, to avoid the hindsight bias. Try to understand what it is like to act in the field of activity, to confront uncertainty under time pressure and with limited resources. The demands of the situation may include many constraints such as uncertainty, time pressure, goal conflicts, and limited resources which are not usually considered in normative models of behavior (e.g., Woods, 1988; Klein et al., 1993). To accomplish this, one must skirt the dangers of the outcome or hindsight biases. Outcome knowledge biases our view of the events and processes leading up to that outcome. We weigh evidence differently than the participants. We underemphasize the role of resource and attentional constraints. We overestimate the knowledge available to the people in the situation. In general, judges with outcome knowledge will tend to simplify the problem-solving situation that was actually faced by the practitioners. The uncertainties, the large data space, and the number of potential actions and diagnostic paths that were faced by the practitioners all may be underemphasized when a task or incident is viewed in hindsight with knowledge of outcome.

Go beyond phenotypical or domain language descriptions and look for genotypical patterns in your incident data. Incidents, analyzed at a deeper level, can be used as a kind of data to reveal more than the risk inherent in some particular incident. Seen in this way, incidents can point to instances of larger trends or function as evidence for the role of different kinds of error genotypes. Do not become fixated on the risk inherent in that particular incident alone; look at the risk of latent failure types and other genotypical factors that pushes incident evolution farther down the path toward disaster (e.g., Woods, 1990a; Hollnagel, 1993).

Expertise and error are context bound. A great deal of information is lost if erroneous actions are counted and aggregated as if they were homogenous. The contextual information is critical in the search for deeper patterns. First, understand the context in which the behavior occurs, the particularness and variety of contributing factors. Then escape from particularness through using and modifying concepts about deeper patterns. The patterns are not context free, but cross-contextual. Seeing the patterns requires looking and abstracting across particular contexts.

To find patterns in the stream of events and incidents, one needs conceptual looking glasses. Thus, study and apply knowledge of different kinds of genotypical patterns that lead to erroneous actions and assessments. These concepts provide a mechanism for abstracting larger patterns from the flow of behavior. Using concepts to abstract from individual cases is critical if one is to recognize the latent vulnerabilities signaled by these accident precursors (Reason, 1990).

Invest in expanding the research base about systematic patterns of breakdown in distributed cognitive systems. This means we need a kind of complementarity between local concerns (e.g., How do I improve this particular system?) and a broader, longer-term view. Each particular setting also functions as a kind of laboratory for learning about the deeper patterns so that we can make long-run progress as well (Woods, 1993a).

Do not rely just on incident data. Tracking sets of incidents, by itself, will not be enough to help answer the important questions about high-reliability human-machine systems. Go out and look at the relevant operational system in other ways, for example, through audits for latent failures (Reason, 1990), through directed surveys, and through field studies, among other approaches. Looking at incidents too narrowly or at a superficial level makes it very easy to miss the role of latent factors, especially the clumsy use of technology (for an example, see Moll von Charante et al., 1993).

The topic of automation surprises provides a good example of methods that can be orchestrated to seek out and understand latent factors. Based on a variety of concerns about the impact of new levels of automation in commercial air transport, Wiener (1989) surveyed pilot opinion about the new generation of glass cockpits. His survey included a tantalizing qucstion and responsc. A fairly largc number of respondents agreed with thc statement: In thc automation, there are still things that happen that surprisc mc. The fact that pilots reported being surprised by the bchavior of thc automated systems intrigued Sarter and Woods (1992): In what circumstances did these surprises occur? Were there patterns? What factors influenced this breakdown in awareness or coordination bctwccn thc human crcw and their automated partners? Sarter and Woods used several techniques (soliciting specific cases from pilots; observing pilots during their glass cockpit training) to build a corpus of automation surprises to help answer these questions. The results helped in the design of a field experimcnt to invcstigatc pattcrns in pilot coordination with cockpit automation (Sartcr and Woods, 1994). Undcrstanding thc pattcrns underlying automation surprises (c.g., indirect modc changes) lcd to new rcsearch directions, (c.g., mode awarcncss), ncw design directions (e.g., techniques to providc cnhanccd fccdback about modcs, mode transitions and displays of what may happcn ncxt; scc Chaptcr 5), and new training directions (e.g., exploratory training to cnhance the flexible use of knowledge, Feltovich ct al., 1993).

One theme of this book has bccn that the clumsy use of computer tcchnology is a kind of latcnt failurc which can contributc to incidents and accidents (sec Chaptcr 5 and Figurc 9, p. 125). The effccts of thc clumsy usc of tcchnological possibilitics can bc sccn without waiting for accidents to occur. First, wc can look for cognition-shaping characteristics of computer-based devices. For example, one can examine a prototypc computcrizcd dcvicc and noticc that there arc a largc number of windows that could bc opcncd and manipulated on a single VDU. But if this capability is orchcstratcd clumsily, thcn thc systcm will force scrial acccss to highly rclatcd data and crcatc ncw interface management burdcns, for examplc, dc-cluttering thc VDU surfacc (Cook ct al., 1990; Cook, Woods, McColligan, and Howic, 1991; Woods, in press-b). Negative consequences will be larger if these data management burdens tcnd to congrcgatc at high-criticality, high-tcmpo periods of task pcrformancc.

Sccond, wc can mcasurc thc impact of thc clumsy usc of technology in tcrms of the impact on the cognitivc activitics of agcnts in the distributcd systcm. Properties of the computcr-bascd tcchnology may increase demands on user memory, undermine attentional control skills (where to focus when), or impair the development of accurate mental models of the function of the device and the underlying processes.

A third place to see the impact of clumsy use of technology, short of waiting for incidents, is to examine the behavior of the people embedded in an operational system. If memory demands are high, practitioners are likely to develop their own aiding strategies (e.g., notes, external reminders) to compensate or to simplify how they use the technological devices to reduce the need to remember so much. If there is a proliferation of displays, windows, and options, practitioners have been observed to tailor the device and their strategies to reduce the knowledge and attentional demands. For example, they may set up the device in ways to avoid interacting with it during high-tempo periods.

However, the ability to adapt around the clumsiness and complexities is limited; user tailoring may be ineffective or brittle if certain events or circumstances arise (see Chapter 5). Various erroneous actions may slip through practitioners' defenses and be revealed as mode errors or automation surprises. Poor feedback about the state of the computer-based systems may hinder their ability to detect and recover from failures, erroneous actions or assessments. Examining the operational system for these kinds of effects of the clumsy use of technological possibilities can reveal latent problems before one or another becomes a contributing factor in a more serious incident.

See incidents and investigations of your operational system as opportunities to learn–to engage in fundamental learning about your organization, to learn how it constrains or supports the people at the sharp end. Recognize and explore goal conflicts. Hiding or suppressing dilemmas and tradeoffs will exacerbate their potential for havoc. This point can be stated another way: beware of the fundamental surprise error in your response to an incident or an accident. (The fundamental surprise error is to re-interpret an event that challenges basic assumptions as if it were merely due to narrow local factors.) Incidents are, by definition, unpleasant surprises (Lanir, 1986). But, however unpleasant, incidents are opportunities to learn. Because they are unpleasant, it is tempting to see surprises only in terms of specific, local, and well bounded channels. However, the result is only the usual recommendations of "blame and train," "a little more technology will be enough," "be more vigilant (try harder)," or "only follow the rules." The outcome and hindsight biases, in particular, undermine our ability to learn fundamentally from failure. Start with the assumption that incidents are evidence of a fundamental surprise. Then ask, How did events go beyond our model of where accidents come from? How are these events evidence of other failure modes/paths or other factors that we have not recognized or invested resources to address? Incidents are opportunities for the organization to learn about itself and its relationship to larger systems and stakeholders.

A corollary to the fundamental surprise error is the fallacy of thinking that "just a little more technology will be enough." Human error is not some deficiency or flaw or weakness that resides inside people. It cannot be treated by appealing to technology and trying to eliminate people as unreliable or unpredictable system elements. "Human error" is a symptom, a kind of information about the functioning of the system in which those people are embedded (Rasmussen, 1986). One must understand this system, which is fundamentally a distributed multi-agent system, a human-machine system, a cognitive system, to know how to use technological powers skillfully as opposed to clumsily (see Chapter 4).

Recognize that there are no ahsolute, single "causes" for accidents. There are many factors that contrihute to incidents and disasters. Therefore, always keep a set in mind; do not focus on only a single factor in isolation (Chapter 3). Which of these many factors we focus on and the level or grain of analysis that we apply to those factors are the products of *human* processes (social and psychological processes) of causal attribution. What we identify as causes depends on whom we are communicating to, on the assumed contrast cases or causal hackground for that exchange, and on the purposes of the inquiry. Thus, the suhset of contrihutors to an incident or disaster that are seen as causal will he different depending on the purposes of the investigation. If the investigation takes place for liability purposes, the concerns will be to decide who pays for damages or consequences, how to limit liability judgments, or how to deflect responsibility for damages to other parties. If it is done for funereal[44] purposes, the concerns will he how do we put the losses, often very personal losses, hehind us, reassert our faith and trust in using the implicated operational system, and resume normal activities (e.g., after an aviation disaster, people still need to get back on an airplane and make use of air transport systems). If it is done for political (power) purposes, incidents may be used as clubs or levers in struggles for control within or across organizations. But if the goal is improving the reliability of the distributed human-machine system, the concerns should be to learn about how the overall system is vulnerable to failure, to develop effective strategies for change, and to prioritize investment.

Perhaps the greatest clue to the reliability of an organization lies in its reaction to failure. Do not use investigations simply to justify the organization's motives (all of us, wherever placed in an organization, want to and try to do a good job). Do not investigate with the *a priori* goal of finding out how *others* failed. Rochlin et al. (1987) and Westrum (1993) report on specific organizations where timely and accurate information flow is rewarded and valued, even when the information is about one's own erroneous actions or about problems in the system (e.g., the Aviation Safety Reporting System (ASRS) for commercial air transport). These organizations did not react to such information with punishment for the involved parties or policies that would have the effect of suppressing information flow.

From these and other cases, one important measure of the reliability of an organization, in the sense of resilience or robustness, may be how it responds to evidence of failures. Lower reliability organizations tend to react with a search for culprits. Their reactions can take several forms: exhortations (or punishments) presumed to increase practitioners' vigilance or attention to detail,[45] removal or exile of the culprits,[46] or mechanisms to attempt to regiment operators in order to protect management from the apparent unreliability or unpredictability of operational personnel (i.e., either injunctions to closely follow the rules or the introduction

[44] Accident investigation, in part, can fulfill some of the roles of a funeral after a tragedy, a ritual marking the resolution of the tragedy that assists the healing process and promotes one's ability to go on with normal activities.

[45] This can take a variety of forms such as bulletins issued to be more careful in certain operational contexts, notices sent to read the manual (as occurred in the case investigated in Moll van Charante et al., 1993), or exhortations to follow standard policies more closely.

[46] The view seems to be our people do not err; if they do, we fire them. (See Norman, 1990b.)

of more automation). On the other hand, higher reliability organizations tend to see failures as opportunities to learn and change (Rochlin et al., 1987; Westrum, 1993).

Learning from error is difficult, both for individuals and for organizations. As systems become more complex and highly coupled (Perrowian complexity; Perrow, 1984; Woods, 1988), the ability to recognize failure is degraded. When failures involve multiple factors, it is easy to interpret or rationalize them in many different ways. And it also becomes more difficult to respond constructively to failure (e.g., seeing possible directions for change), in part, because the greater coupling increases the reverberations of change. As Perrowian complexity increases, the specialization of agents in the system goes up as well. As a result, no one person or group can see the whole situation.

But we have a responsibility, which is driven by the consequences that can accompany failure, to maximize the information value of such potentially expensive feedback. Achieving higher reliability in human-machine systems demands that we look directly, honestly, and intensely in every way at incidents, disasters, and their precursors. If we label events as human error and stop, what have we learned? As the many examples and concepts in this volume illustrate, the answer is very little. The label "human error" is a judgment made in hindsight. Failures occur in systems that people develop and operate for human purposes. Such systems are not and cannot be purely technological; they always involve people at various levels and in various ways. We cannot pretend that technology alone, divorced from the people who develop, shape, and use it, will be enough. Failure and success are both forms of information about the *system* in which people are embedded. The potential for constructive change lies behind the label "human error."

7

REFERENCES

Adler, P. (1986). New technologies, new skills. *California Management Review, 29,* 9-28.

Axelrod, R. (1984). *The evolution of cooperation.* New York: Basic Books.

Baars, B. (1992). *Experimental slips and human error: Exploring the architecture of volition.* New York: Plenum Press.

Baron, J., and Hershey, J. (1988). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology, 54*(4), 569-579.

Bereiter, S., and Miller, S. (1988). Sources of difficulty troubleshooting automated manufacturing systems. In W. Karwowski, H. Parsaei, and M. Wilhelm (Eds.), *Ergonomics of hybrid automated systems.* Amsterdam: Elsevier Science.

Billings, C. (1991). *Human-centered aircraft automation: A concept and guidelines* (NASA Tech. Memo. 103885). Moffett Field, CA: NASA Ames Research Center.

Boeing Product Safety Organization. (1993). *Statistical summary of commercial jet aircraft accidents: Worldwide operations, 1959-1992.* Seattle, WA: Boeing Commercial Airplanes.

Boring, E. G. (1950). *A history of experimental psychology* (2nd Ed.). New York: Appleton-Century-Crofts.

Bransford, J., Sherwood, R., Vye, N., and Rieser, J., (1986). Teaching and problem solving: Rcsearch foundations. *American Psychologist, 41,* 1078-1089.

Brown, J. S., Moran, T. P., and Williams, M. D. (1982). *The semantics of procedures.* Palo Alto, CA: Xerox Research Center.

Cacciabue, P. C., Decortis, F., Drozdowicz, B., Masson, M., and Nordvik, J.-P. (1992). COSIMO: A cognitive simulation model of human decision making and behavior in accident management of complex plants. *IEEE Transactions on Systems, Man and Cybernetics, 22* (5), 1058-1074.

Caplan, R., Posner, K., and Cheney, F. (1991). Effect of outcome on physician judgments of appropriateness of care. *Journal of the American Medical Association, 265,* 1957-1960.

Cheng, P., and Novick, L. (1992). Covariation in natural causal induction. *Psychological Review, 99,* 365-382.

Chi, M. T. H., Glaser, R., and Farr, M. (1988). *The nature of expertise.* Hillsdale, NJ: Erlbaum.

Chopra, V., Bovill, J. G., Spicrdijk, J., and Koornneef, F. (1992). Reported significant observations during anaesthesia: A prospective analysis over an 18-month period. *British Journal of Anaesthesia, 68,* 13-17.

Cook, R. I., McDonald, J. S., and Smalhout, R. (1989). *Human error in the operating room: Identifying cognitive lock up* (CSEL Tech. Rep. 89-TR-07). Columbus, OH: Thc Ohio Statc University, Department of Industrial and Systems Engineering, Cognitive Systems Engineering Laboratory.

Cook, R. I., Woods, D. D., and Howie, M. B. (1990). The natural history of introducing new information technology into a high risk environment. In *Proceedings of the 35th Annual Meeting of the Human Factors Society* (pp. 429-433). Santa Monica, CA: Human Factors Society.

Cook, R. I., Woods, D. D., and McDonald, J. S. (1991). *Human performance in anesthesia: A corpus of cases* (CSEL Tech. Rep. 91-TR-03). Columbus, OH: The Ohio State University, Department of Industrial and Systems Engineering, Cognitive Systems Engineering Laboratory.

Cook, R. I., Woods, D. D., McColligan, E., and Howie, M. B. (1991). Cognitive consequences of "clumsy" automation on high workload, high consequence human performance. In R. T. Savely (Ed.), *Fourth Annual Workshop on Space Operations, Applications and Research (SOAR '90)* (NASA Report CP-3103). Washington, DC: National Aeronautical and Space Administration.

Cook, R. I., Potter, S. S., Woods, D. D., and McDonald, J. S. (1991). Evaluating the human engineering of microprocessor-controlled operating room devices. *Journal of Clinical Monitoring, 7,* 217-226.

Cook, R. I., Woods, D. D., and Howie, M. B. (1992). Unintentional delivery of vasoactive drugs with an electromechanical infusion device. *Journal of Cardiothoracic and Vascular Anesthesia, 6*, 1-7.

Cook, R. I., and Woods, D. D. (1994). Operating at the sharp end: The complexity of human error. In M. S. Bogner (Ed.), *Human error in medicine.* Hillsdale, NJ: Erlbaum.

Cooper, J. B., Newbower, R. S., Long, C. D., and McPeek, B. (1978). Preventable anesthesia mishaps: A study of human factors. *Anesthesiology, 49*, 399-406.

Cooper, J. B., Newbower R. S., and Kitz R. J. (1984). An analysis of major errors and equipment failures in ancsthcsia management: Conditions for prevention and detection. *Anesthesiology, 60,* 43-42.

Corker, K., Davis, L., Papazian, B., and Pcw, R. (1986). *Development of an advanced task analysis methodology and demonstration for army aircrew/aircraft integration* (BBN Report 6124). Cambridge, MA: Bolt, Beranek, & Newman.

De Keyser, V., and Woods, D. D. (1990). Fixation errors: Failures to revise situation assessmcnt in dynamic and risky systems. In A. G. Colombo and A. Saiz dc Bustamantc (Eds.), *System reliability assessment* (pp. 231-251). Drodrccht, Thc Nctherlands: Kluwer Academic.

Dorner, D. (1983). Heuristics and cognition in complex systems. In R. Groner, M. Groner, and W. F. Bischof (Eds.), *Methods of heuristics.* Hillsdale, NJ: Erlbaum.

Dougherty, E. M. (1990). Guest editorial. *Reliability Engineering and System Safety, 29,* 283-299.

Doughcrty, E. M., and Fragola, J. R. (1990). *Human reliability analysis: A systems engineering approach with nuclear power plant applications.* New York: Wilcy.

Edwards, W. (1984). How to make good decisions. *Acta Psychologica, 56,* 5-27.

Elm, W. C., and Woods, D. D. (1985). Gctting lost: A case study in interface design. *Proceedings of the 29th Annual Meeting of the Human Factors Society* (pp. 927-931). Santa Monica, CA: Human Factors Society.

Endsley, M. R. (1988, May). *Situation awareness global assessment technique (SAGAT).* Paper presented at thc National Acrospace and Electronic conference (NAECON), Dayton, OH.

Federal Communications Commission. (1991). *Report on the September 17, 1991 AT&T NY power outage at Thomas Street switching station and network disruption.* Washington, DC: Common Carrier Bureau.

Feltovich, P. J., Spiro, R. J., and Coulson, R. (1989). The nature of conceptual understanding in biomedicine: The deep structure of complex ideas and the development of misconceptions. In D. Evans and V. Patel (Eds.), *Cognitive science in medicine: Biomedical modeling.* Cambridge, MA: MIT Press.

Feltovich, P. J., Spiro, R. J., and Coulson, R. (1993). Learning, teaching and testing for complex conceptual understanding. In N. Fredericksen, R. Mislevy, and I. Bejar (Eds.), *Test theory for a new generation of tests.* Hillsdale, NJ: Erlbaum.

Fischer, U., Orasanu, J.M., and Montvalo, M. (1993). Efficient decision strategies on the flight deck. In *Proceedings of the Seventh International Symposium on Aviation Psychology.* Columbus, OH: The Ohio State University, Aviation Psychology Laboratory.

Fischhoff, B. (1975). Hindsight–foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance, 1*(3), 288-299.

Fischhoff, B. (1977). Perceived informativeness of facts. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 349-358.

Fischhoff, B. (1982). For those condemned to study the past: Heuristics and biases in hindsight. In D. Kahneman, P. Slovic, and A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases.* Cambridge, England: Cambridge University Press.

Fischhoff, B. (1986). Decision making in complex systems. In E. Hollnagel, G. Mancini, and D. D. Woods (Eds.), *Intelligent decision making in process control environments.* Berlin: Springer-Verlag.

Fischhoff, B., and Beyth, R. (1975). "I knew it would happen" Remembered probabilities of once-future things. *Organizational Behavior and Human Performance, 13,* 1-16.

Fischhoff, B., and Beyth-Marom, R. (1983). Hypothesis evaluation from a Bayesian perspective. *Psychological Review, 90*(3), 239-260.

Fitts, P. M., and Jones, R. E. (1947). *Analysis of factors contributing to 460 "pilot-error" experiences in operating aircraft controls* (Memorandum Report TSEAA-694-12). Wright Field, OH: U.S. Air Force Air Materiel Command, Aero Medical Laboratory.

Flach, J., Hancock, P., Caird, J., and Vicente, K. (Eds.). (In press). *An ecological approach to human-machine systems I: A global perspective.* Hillsdale, NJ: Erlbaum.

Flores, F., Graves, M., Hartfield, B., and Winograd, T. (1988). Computer systems and the design of organizational interaction. *ACM Transactions on Office Information Systems, 6,* 153-172.

Fraser, J. M., Smith, P. J., and Smith, J. W., Jr. (1992). A catalog of errors. *International Journal of Man-Machine Studies, 37*(3), 265-393.

Freund, P. R., and Sharar, S. R. (1990). Hyperthermia alert caused by unrecognized temperature monitor malfunction. *Journal of Clinical Monitoring, 6,* 257.

Gaba, D. M., Maxwell, M., and DeAnda, A. (1987). Anesthetic mishaps: Breaking the chain of accident evolution. *Anesthesiology, 66,* 670-676.

Gaba, D. M., and DeAnda, A. (1989). The response of anesthesia trainees to simulated critical incidents. *Anesthesia and Analgesia, 68,* 444-451.

Gentner, D., and Stevens, A. L. (Eds.) (1983). *Mental models.* Hillsdale, NJ: Erlbaum.

Gopher, D. (1991). The skill of attention control: Acquisition and execution of attention strategies. In D. Meyer and S. Kornblum (Eds.), *Attention and Performance XIV*. Hillsdale, NJ: Erlbaum.

Haber, R. N. (1987). Why low-flying fighter planes crash: Perceptual and attentional factors in collisions with the ground. *Human Factors, 29*, 519-532.

Hart, H. L. A., and Honore, A. M. (1959). *Causation in the Law.* Oxford, England: Clarendon Press.

Hasher, L., Attig, M. S., and Alba, J. W. (1981). I knew it all along: Or did I? *Journal of Verbal Learning and Verbal Behavior, 20,* 86-96.

Hawkins, S., and Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin, 107*(3), 311-327.

Herry, N. (1987). Errors in the execution of prescribed instructions. In J. Rasmussen, K. Duncan, and J. Leplat. (Eds.), *New technology and human error.* Chichester, England: Wiley.

Hilton, D. (1990). Conversational processes and causal explanation. *Psychological Bulletin, 197*(1), 65-81.

Hirschhorn, L. (1993). Hierarchy vs. bureaucracy: The case of a nuclear reactor. In K. H. Roberts (Ed.), *New challenges to understanding organizations.* New York: McMillan.

Hoch, S. J., and Lowenstein, G. F. (1989). Outcome feedback: Hindsight and information. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15*(4), 605-619.

Hochberg, J. (1986). Representation of motion and space in video and cinematic displays. In K. R. Boff, L. Kaufman, and J. P. Thomas, (Eds.), *Handbook of perception and human performance: Volume I. Sensory processes and perception.* New York: Wiley.

Hollister, W. M. (Ed.). (1986). *Improved guidance and control automation at the man-machine interface* (AGARD Advisory Rep. No. 228). Neuilly-Sur-Seine, France: Advisory Group for Aerospace Research & Development.

Hollnagel, E. (1991a). The phenotype of erroneous actions. In G. R. Weir and J. L. Alty (Eds.), *Human-computer interaction and complex systems.* London: Academic Press.

Hollnagel, E. (1991b). Cognitive ergonomics and the reliability of cognition. *Le Travail Humain, 54*(4).

Hollnagel. E. (1993). *Human reliability analysis: Context and control.* London: Academic Press.

Hollnagel, E., and Woods, D. D. (1983). Cognitive systems engineering: New wine in new bottles. *International Journal of Man-Machine Studies, 18,* 583-600.

Hughes, J., Randall, D., and Shapiro, D. (1991). CSCW: Discipline or paradigm? A sociological perspective. In L. Bannon, M. Robinson, and K. Schmidt (Eds.), *E-CSCW '91: The Second European Conference on Computer Supported Cooperative Work* (pp. 309-323). Amsterdam: Kluwer.

Hughes, J., Randall, D., and Shapiro, D. (1992). Faltering from ethnography to design. *Computer Supported Cooperative Work (CSCW) Proceedings.*(pp. 1-8). New York: Association for Computing Machinery.

Hutchins, E. (1990). The technology of team navigation. In J. Galegher, R. Kraut, and C. Egido (Eds.), *Intellectual teamwork: Social and technical bases of cooperative work.* Hillsdale, NJ: Erlbaum.

Hutchins, E. (1991). *How a Cockpit Remembers its Speed* (Tech. Rep.). San Diego, CA: University of California, Distributed Cognition Laboratory.

Hutchins, E. (In press). *Cognition in the wild.* Cambridge, MA: MIT Press.

Johnson, P. E., Duran, A. S., Hassebrock, F., Moller, J., Prietula, M., Feltovich, P. J., and Swanson, D. B. (1981). Expertise and error in diagnostic reasoning. *Cognitive Science, 5,* 235-283.

Johnson, P. E., Moen, J. B., and Thompson, W. B. (1988). Garden path errors in diagnostic reasoning. In L. Bolec and M. J. Coombs (Eds.), *Expert system applications.* New York: Springer-Verlag.

Johnson, P. E., Jamal, K., and Berryman, R. G. (1991). Effects of framing on auditor decisions. *Organizational Behavior and Human Decision Processes, 50,* 75-105.

Johnson, P. E., Grazioli, S., Jamal, K., and Zualkernan, I. (1992). Success and failure in expert reasoning. *Organizational Behavior and Human Decision Processes, 53,* 173-203.

Kelley, H. H. (1973). The process of causal attribution. *American Psychologist, 28,* 107-128.

Kemeny, J. G., et al. (1979). *Report of the President's Commission on the accident at Three Mile Island.* New York: Pergamon Press.

Kieras, D. E., and Polson, P. G. (1985). An approach to the formal analysis of user complexity. *International Journal of Man-Machine Studies, 22,* 365-394.

Klein, G. A. (1989). Do decision biases explain too much? *Human Factors Society Bulletin, 32* (5), 1-3.

Klein, G. A., and Crandall, B. (in press). The role of mental simulation in problem solving and decision making. In J. Flach, P. Hancock, J. Caird, and K. Vicente (Eds.), *An ecological approach to human-machine systems I: A global perspective.* Hillsdale NJ: Erlbaum.

Klein, G. A., Orasanu, J., and Calderwood, R. (Eds.). (1993). *Decision making in action: Models and methods.* Norwood, NJ: Ablex.

Lanir, Z. (1986). *Fundamental surprise.* Eugene, OR: Decision Research.

Lanir, Z., Fischhoff, B., and Johnson, S. (1988). Military risk taking: C_3I and the cognitive function of boldness in war. *Journal of Strategic Studies, 11* (1), 96-114.

Lauber, J. K. (1993, April). *A safety culture perspective.* Paper presented at Flight Safety Foundation's 38th Corporate Aviation Safety Seminar, Irving, TX.

Layton, C., Smith, P. J., and McCoy, E. (1994). Design of a cooperative problem solving system for enroute flight planning: An empirical evaluation. *Human Factors, 36* (1), 94-119.

Lenorovitz, J. M. (1990). Indian A320 crash probe data show crew improperly configured aircraft. *Aviation Week and Space Technology, 132*(6/25/90), 84-85.

Lenorovitz, J. M. (1992a). Confusion over flight mode may have role in A320 crash. *Aviation Week and Space Technology, 137(2/3/92),* 29-30.

Lenorovitz, J. M. (1992b). French government seeks A320 changes following Air Inter crash report. *Aviation Week and Space Technology, 137*(3/2/92), 30-31.

Leveson, N. G., and Turner, C. S. (1992). *An investigation of Therac-25 accidents* (UCI Tech. Rep. No. 92-108). Irvine, CA: University of California.

Lewis, C., and Norman, D. A. (1986). Designing for error. In D. A. Norman and S. W. Draper (Eds.), *User-centered system design: New perspectives of human-computer interaction* (pp. 411-432). Hillsdale, NJ: Erlbaum.

Lipshitz, R. (1989). "Either a medal or a corporal": The effects of success and failure on the evaluation of decision making and decision makers. *Organizational Behavior and Human Decision Processes, 44,* 380-395.

Lubar, S. (1993). Review of "The evolution of useful things" by Henry Petroski. *Science, 260* (May 21), 1166-1168.

Mach, E. (1905). *Knowledge and error.* Dordrecht, The Netherlands: Reidel Publishing Company. (English Translation, 1976)

March, J. G. (1978). Bounded rationality, ambiguity and the engineering of choice. *The Bell Journal of Economics, 9,* 587-608.

McRuer, D., et al. (Eds.). (1992). *Aeronautical technologies for the twenty-first century* (pp. 243-267). Washington, DC: National Academy Press.

Moll van Charante, E., Cook, R. I., Woods, D. D., Yue, Y., and Howie, M. B. (1993). Human-computer interaction in context: Physician interaction with automated intravenous controllers in the heart room. In H. G. Stassen (Ed.), *Analysis, design and evaluation of man-machine systems 1992.* New York: Pcrgamon Press.

Monk, A. (1986). Mode errors: A user centred analysis and some preventative measures using keying-contingent sound. *International Journal of Man Machine Studies, 24,* 313-327.

Monnier, A. (1992). *Rapport preliminaire de la commission d'enquete administrative sur l'accident du Mont Sainte Odile du 20 Janvier 1992* (Conference de Pressc du 24 fevrier 1992). Paris, France: Ministere de L'equipement, du logement, des transports ct de l'espace.

Moray, N. (1984). Attention to dynamic visual displays in man-machine systems. In R. Parasuraman and D. R. Davies (Eds.), *Varieties of attention.* New York: Academic Press.

Moray, N., Dessouky, M., Kijowski, B., and Adapathya, R. (1991). Strategic behavior, workload and performance in task scheduling. *Human Factors, 33,* 607-629.

Moray, N., and Huey, B. (Eds.). (1988). *Human factors research and nuclear safety.* Washington, DC: National Academy Press.

Moshansky, V. P. (1992). *Final report of the commission of inquiry into the Air Ontario crash at Dryden, Ontario.* Ottawa, Canada: Minister of Supply and Services.

Muir, B. (1987). Trust between humans and machines. *International Journal of Man-Machine Studies, 27,* 527-539.

Murray, C., and Cox, C. B. (1989). *Apollo: The race to the moon.* New York: Simon & Schuster.

National Transportation Safety Board. (1991). *Avianca, the airline of Colombia, Boeing 707-321B, HK 2016, Fuel Exhaustion, Cove Neck, NY, January 25, 1990* (Report No. AAR-91/04). Washington, DC: Author.

National Transportation Safety Board. (1984). *Eastern Air Lines Lockheed L-1011, N334EA Miami International Airport, FL, May 5, 1983* (Report No. AAR 84/04). Washington, DC: Author.

National Transportation Safety Board. (1986a). *China Airlines B-747-SP, 300 NM northwest of San Francisco, CA, February 19, 1985* (Report No. AAR-86/03).Washington, DC: Author.

National Transportation Safety Board. (1986b). *Delta Air Lines Lockheed L-1011-385-1, Dallas-Fort Worth Airport, TX, August 2, 1985* (Report No. AAR-86/05). Washington, DC: Author.

National Transportation Safety Board. (1990). *Marine accident report: Grounding of the U.S. Tankship Exxon Valdez on Bligh Reef, Prince William Sound, near Valdez, Alaska, March 24, 1989* (Report No. NTSB/MAT-90/04). Washington, DC: Author.

National Transportation Safety Board. (1993). *US Air Flight 405 LaGuardia Airport, March 22, 1992* (Report No. AAR-93/02). Washington, DC: Author.

Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology.* San Francisco: Freeman.

Newell, A. (1982). The knowledge level. *Artificial Intelligence, 18,* 87-127.

Newell, A., and Simon, H. A. (1963). GPS, a program that simulates human thought. In E. A. Feigenbaum and J. Feldman (Eds.), *Computers and thought.* New York: McGraw Hill.

Newell, A., and Simon, H.A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice-Hall.

Norman, D. A. (1981). Categorization of action slips. *Psychological Review, 88,* 1-15.

Norman, D. A. (1983). Design rules based on analysis of human error. *Communications of the ACM, 26,* 254-258.

Norman, D. A. (1988). *The psychology of everyday things.* New York: Basic Books.

Norman, D. A. (1990a). The "problem" of automation: Inappropriate feedback and interaction, not "overautomation." *Philosophical Transactions of the Royal Society of London, B 327,* 585-593.

Norman, D. A. (1990b). Commentary: Human error and the design of computer systems. *Communications of the ACM, 33(1),* 4-7.

Norman D. A. (1992). *Turn signals are the facial expressions of automobiles.* New York: Addison-Wesley.

Oaksford, M., and Chater, N. (1992). Bounded rationality in taking risks and drawing inferences (Notes and Comments I). *Theory and Psychology,* 2(2), 225-230.

Orasanu, J. M. (1990). *Shared mental models and crew decision making* (CSL Rep. 46). Princeton, NJ: Princeton University, Cognitive Science Laboratory.

Payne, J. W., Bettman, J. R., and Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory and Cognition, 14*(3), 534-552.

Payne J. W., Johnson E. J., Bettman J. R., and Coupey, E. (1990). Understanding contingent choice: A computer simulation approach. *IEEE Transactions on Systems, Man, and Cybernetics, 20*, 296-309.

Perkins, D., and Martin, F. (1986). Fragile knowledge and neglected strategies in novice programmers. In E. Soloway and S. Iyengar (Eds.), *Empirical studies of programmers.* Norwood, NJ: Ablex.

Perkins, D. N., and Salomon, G. (1989). Are cognitive skills context-bound? *Educational Researcher,* January-February, 16-25.

Perrow, C. (1984). *Normal accidents. Living with high-risk technologies.* New York: Basic Books.

Pew, R. W., Miller, D. C., and Feehrer, C. E. (1981). *Evaluation of proposed control room improvements through analysis of critical operator decisions* (NP-1982). Palo Alto, CA: Electric Power Research Institute.

Potter, S. S., Woods, D. D., Hill, H., Boyer, R., and Morris, W. (1992). Visualization of dynamic processes: Function-based displays for human-intelligent system interaction. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, Chicago, IL* (pp. 1504-1509). New York: Institute of Electrical and Electronic Engineers.

Rasmussen, J. (1985). Trends in human reliability analysis. *Ergonomics, 28* (8), 1185-1196.

Rasmussen, J. (1986). *Information processing and human-machine interaction: An approach to cognitive engineering.* New York: North-Holland.

Rasmussen, J., and Batstone, R. (1989). Why do complex organizational systems fail? *Environment Working Paper No. 20.* Washington, DC: The World Bank.

Rasmussen, J., Duncan, K., and Leplat, J. (Eds.) (1987). *New technology and human error.* Chichester, England: Wiley.

Reason, J. (1990). *Human error.* Cambridge, England: Cambridge University Press.

Reason, J. (1993). The identification of latent organizational failures in complex systems. In J. A. Wise, V. D. Hopkin and P. Stager (Eds.), *Verification and Validation of Complex Systems: Human Factors Issues.* Berlin: Springer-Verlag.

Reason, J., and Mycielska, K. (1982). *Absent minded? The psychology of mental lapses and everyday errors.* Englewood Cliffs, NJ: Prentice Hall.

Resnick, L., Levine, J., and Teasley, S. D. (1991). *Perspectives on socially shared cognition.* Washington, DC: American Psychological Association.

Roberts, K. H., and Rousseau, D. M. (1989). Research in nearly failure-free, high-reliability organizations: Having the bubble. *IEEE Transactions in Engineering Management, 36,* 132-139.

Rochlin, G., LaPorte, T. R., and Roberts, K. H. (1987). The self-designing high reliability organization: Aircraft carrier flight operations at sea. *Naval War College Review, (Autumn),* 76-90.

Rochlin, G. (1991). Iran Air Flight 655 and the USS Vincennes. In T. LaPorte (Ed.), *Social responses to large technical systems.* Doedrecht, The Netherlands: Kluwer Academic.

Rogers, W. P., et al. (1986). *Report of the Presidential Commission on the Space Shuttle Challenger Accident.* Washington, DC: Government Printing Office.

Rosson, M. (1983). Patterns of experience in text editing. *CHI'83: Human Factors in Computing Systems* (pp. 171-175). New York: ACM.

Roth, E. M., Bennett, K. B., and Woods, D. D. (1987). Human interaction with an "intelligent" machine. *International Journal of Man-Machine Studies, 27,* 479-525.

Roth, E. M., and Woods, D. D. (1988). Aiding human performance: I. Cognitive analysis. *Le Travail Humain, 51*(1), 39-64.

Roth E. M., Woods, D. D., and Pople, H. E., Jr. (1992). Cognitive simulation as a tool for cognitive task analysis. *Ergonomics, 35,* 1163-1198.

Rouse, W. B., et al. (1984). *A method for analytical evaluation of computer-based decision aids* (NUREG-CR-3655). Washington, DC: U.S. Nuclear Regulatory Commission.

Rouse, W. B., and Morris, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin, 100,* 359-363.

Sarter, N. B., and Woods, D. D. (1991). Situation awareness: A critical but ill-defined phenomenon. *International Journal of Aviation Psychology, 1,* 43-55.

Sarter, N. B.,·and Woods, D. D. (1992). Pilot interaction with cockpit automation I: Operational experiences with the Flight Management System. *International Journal of Aviation Psychology, 2,* 303-321.

Sarter N. B., and Woods D. D. (1994). Pilot interaction with cockpit automation II: An experimental study of pilots' mental model and awareness of the Flight Management System (FMS). *International Journal of Aviation Psychology, 4* (1), *1-28.*

Sarter, N. B., and Woods, D. D. (in press). "How in the world did we get into that mode?" Mode error and awareness in supervisory control. *Human Factors (Special Issue on Situation Awareness).*

Schwenk, C., and Cosier, R. (1980). Effects of the expert, devil's advocate and dialectical inquiry methods on prediction performance. *Organizational Behavior and Human Decision Processes, 26, 409-424.*

Schwid, H. A., and O'Donnell, D. (1992). Anesthesiologist's management of simulated critical incidents. *Anesthesiology, 76*, 495-501.

Seaman, C. (1992). Safety services human factors research. British Airlines, *Flywise, 17(December).*

Segal, L. D. (1993). Automation design and crew coordination. In *Proceedings of the Seventh International Symposium on Aviation Psychology.* Columhus, Ohio: The Ohio State University, Aviation Psychology Laboratory.

Seifert, C. M., and Hutchins, E. (1992). Error as opportunity: Learning in a cooperative task. *Human-Computer Interaction, 7*(4), 409-435.

Sellen, A. J., Kurtenbach, G. P., and Buxton, W. A. S. (1992). The prevention of mode errors through sensory feedback. *Human-Computer Interaction, 7,* 141-164.

Senders, J., and Moray, N. (1991). *Human error: Cause, prediction, and reduction.* Hillsdale, NJ: Erlbaum.

Simon, H. (1957). *Models of man (Social and rational).* New York: Wiley.

Simon, H. (1969). *The sciences of the artificial.* Cambridge, MA: MIT Press.

Singleton, W. T. (1973). Theoretical approaches to human error. *Ergonomics, 16,* 727-737.

Smith, E. E. and Goodman, L. (1984). Understanding written instructions: The role of an explanatory schema. *Cognition and Instruction, 1,* 359-396.

Spiro, R. J., Coulson, R. L., Feltovich, P. J. and Anderson, D. K. (1988). Cognitive flexibility theory: Advanced knowledge acquisition in ill-structured domains. *Proceedings of the Tenth Annual Conference of the Cognitive Science Society* (pp. 375-383). Hillsdale, NJ: Erlbaum.

Stevens, S. S. (1946). Machines cannot fight alone. *American Scientist, 34,* 389-400.

Suchman, L. (1987). *Plans and situated actions. The problem of human machine communication.* Cambridge, England: Cambridge University Press.

Tenney, Y. J., Jager Adams, M., Pew, R. W., Huggins, A. W. F., and Rogers, W. H. (1992). *A principled approach to the measurement of situation awareness in commercial aviation* (NASA Contractor Report No. NAS1-18788). Hampton, VA: NASA Langley Research Center.

Thordsen, M. L., and Klein, G. (1989). Cognitive processes of the team mind. In *Proceedings of the IEEE Conference on Systems, Man and Cybernetics* (pp. 46-49). New York: Institute of Electrical and Electronic Engineers.

Turner, B. A. (1978). *Man-made disasters.* London: Wykeham.

Tversky, A. and Kahnemann, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124-1131.

U.S. Department of Defense. (1988). *Report of the formal investigation into the circumstances surrounding the downing of Iran Air Flight 655 on 3 July 1988.* Washington, DC: Author.

U.S. House of Representatives Committee on Armed Services. (1987). *Report on the staff investigation into the Iraqi attack on the USS Stark.* Washington, DC: Government Printing Office.

U.S. Nuclear Regulatory Commission. (1985). *Loss of main and auxiliary feedwater at the Davis-Besse Plant on June 9, 1985.* (NUREG-CR-1154). Washington, DC: Author.

von Winterfeldt, D., and Edwards, E. (1986). *Decision analysis and behavioral research.* Cambridge, England: Cambridge University Press.

Wagenaar, W., and Keren, G. (1986). Does the expert know? The reliability of predictions and confidence ratings of experts. In E. Hollnagel, G. Mancini, and D. D. Woods (Eds.), *Intelligent decision making in process control environments.* Berlin: Springer-Verlag.

Wagenaar, W., and Groeneweg, J. (1987). Accidents at sea: Multiple causes and impossible consequences. *International Journal of Man-Machine Studies, 27,* 587-598.

Wagenaar, W., and Reason, J. (1990). Types and tokens in road accident causation. *Ergonomics, 33,* 1365-1375.

Wasserman, D., Lempert, R. O., and Hastie, R. (1991). Hindsight and causality. *Personality and Social Psychology Bulletin, 17*(1), 30-35.

Westrum, R. (1993). Cultures with requisite imagination. In J. A. Wise, V. D. Hopkin and P. Stager (Eds.), *Verification and validation of complex systems: Human factors issues.* Berlin: Springer-Verlag.

Wickens, C. D. (1991). *Engineering psychology and human performance.* Columbus, OH: Charles Merrill.

Wiener, E. L. (1989). *Human factors of advanced technology ("glass cockpit") transport aircraft* (Tech. Rep. 117528). Moffett Field, CA: NASA Ames Research Center.

Wiener, N. (1964). *God and Golem.* Cambridge, MA: MIT Press.

Winograd, T., and Flores, F. (1987). *Understanding computers and cognition.* Reading, MA: Addison-Wesley.

Wood, G. (1978). The "knew-it-all-along" effect. *Journal of Experimental Psychology: Human Perception and Performance, 4*(2), 345-353.

Woods, D. D. (1982). Operating decision making behavior during the steam generator tube rupture at the Ginna nuclear power station. In W. Brown and R. Wyrick (Eds.), *Analysis of steam generator tube rupture events at Oconee and Ginna.* Institute of Nuclear Power Operations. 82-030. (Also, Westinghouse Research and Development Center Report 82-1C57-CONRM-R2.)

Woods, D. D. (1984). Visual momentum: A concept to improve the cognitive coupling of person and computer. *International Journal of Man-Machine Studies, 21,* 229-244.

Woods, D. D. (1986). Paradigms for intelligent decision support. In E. Hollnagel, G. Mancini, and D. D. Woods (Eds.), *Intelligent decision support in process environments.* New York: Springer-Verlag.

Woods, D. D. (1988). Coping with complexity: The psychology of human behavior in complex systems. In L. P. Goodstein, H. B. Andersen, and S. E. Olsen (Eds.), *Tasks, errors, and mental models.* New York: Taylor & Francis.

Woods, D. D. (1990a). Modeling and predicting human error. In J. L. Elkind, S. Card, J. Hochberg, and B. Huey (Eds.), *Human performance models for computer-aided engineering.* New York: Academic Press.

Woods, D.D. (1990b). Risk and human performance: Measuring the potential for disaster. *Reliability Engineering and System Safety, 29*, 387-405.

Woods, D. D. (1991). The cognitive engineering of problem representations. In G. R. Weir and J. L. Alty (Eds.), *Human-computer interaction and complex systems*. London: Academic Press.

Woods, D. D. (1992). *The alarm problem and directed attention* (CSEL Tech. Rep. 92-TR-01). Columbus, OH: The Ohio State University, Department of Industrial and Systems Engineering, Cognitive Systems Engineeering Laboratory.

Woods, D. D. (1993a). Research methods for the study of cognition outside of the experimental psychology laboratory. In G. A. Klein, J. Orasanu, and R. Calderwood (Eds.), *Decision making in action: Models and methods.* Norwood, NJ: Ablex.

Woods, D. D. (1993b). The price of flexibility in intelligent interfaces, *Knowledge-Based Systems, 6*, 1-8.

Woods, D. D. (1993c). *Cognitive systems in context* (CSEL Tech. Rep. 93-TR-01). Columbus, OH: The Ohio State University, Department of Industrial and Systems Engineering, Cognitive Systems Engineering Laboratory.

Woods, D. D. (1994). Some observations from studying cognitive systems in context [Keynote Address]. In *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society.* Hillisdale, NJ: Erlbaum.

Woods, D. D. (In press-a). Cognitive demands and activities in dynamic fault management: Abductive reasoning and disturbance management. In Stanton, N. (Ed.), *The human factors of alarm design.* London: Taylor & Francis.

Woods, D. D. (In press-b). Towards a theoretical base for representation design in the computer medium: Ecological perception and aiding human cognition. In J. Flach, P. Hancock, J. Caird, and K. Vicente (Eds.), *An ecological approach to human-machine systems I: A global perspective.* Hillsdale, NJ: Erlbaum.

Woods, D. D., Cook, R. I., and Sarter, N. (1992). *Clumsy automation, practitioner's adaptive response and human error* (CSEL Tech. Rep. 92-TR-04). Columbus, OH: The Ohio State University, Department of Industrial and Systems Engineering, Cognitive Systems Engineering Laboratory .

Woods, D. D., and Elias, G. (1988). Significance messages: An integral display concept. In *Proceedings of the 32nd Annual Meeting of the Human Factors Society* (pp. 1350-1354). Santa Monica, CA: Human Factors Society.

Woods, D. D., O'Brien, J., and Hanes, L. F. (1987). Human factors challenges in process control: The case of nuclear power plants. In G. Salvendy (Ed.), *Handbook of human factors/ergonomics.* New York: Wiley.

Woods, D. D., Pople, H. E., Jr., and Roth, E. M. (1990). *The Cognitive Environment Simulation as a tool for modeling human performance and reliability* (NUREG-CR-5213). Washington, DC: U.S. Nuclear Regulatory Commission.

Woods, D. D., Potter, S. S., Johannesen, L., and Holloway, M. (1991). *Human interaction with intelligent systems. Volumes I and II* (CSEL Tech. Reps. 91-TR-01 and 02). Columbus, OH: The Ohio State University, Department of Industrial and Systems Engineering, Cognitive Systems Engineering Laboratory .

Woods, D. D., and Roth, E. M. (1986). *Models of cognitive behavior in nuclear power plant personnel* (NUREG-CR-4532). Washington, DC: U.S. Nuclear Regulatory Commission.

Woods, D. D., and Roth, E. M. (1988). Cognitive Systems Engineering. In M. Helander (Ed.), *Handbook of human-computer interaction.* New York: Elsevier.

Woods, D. D., and Roth, E. M. (In press). Symbolic AI-based computer simulations as a tool for investigating the dynamics of joint cognitive systems. In J.-M. Hoc, P. C. Cacciabue, and E. Hollnagel (Eds.), *Expertise and technology: Cognition and human-computer cooperation.* Hillsdale, NJ: Erlbaum.

Woods, D. D. and Sarter, N. B. (1993). Evaluating the impact of new technology on human-machine cooperation. In J. Wise, V. D. Hopkin,and P. Stager (Eds.), *Verification and validation of complex and integrated human-machine systems.* Berlin: Springer-Verlag.

Woods, D. D., Wise, J. A., and Hanes, L. F. (1982). *Evaluation of safety parameter display concepts* (Tech. Rep. NP-2239). Palo Alto, CA.: Electric Power Research Institute.

Wright, D., Mackenzie, S. J., Buchan, I., Cairns, C. S., and Price, L. E. (1991). Critical incidents in the intensive therapy unit. *Lancet 338,* 676-678 .

Yue, L., Woods, D. D., and Cook, R. I. (1992). *Cognitive engineering of the human computer interface: Redesign of an infusion controller in cardiac anesthesiology* (CSEL Tech. Rep. 92-TR-03). Columbus, OH: The Ohio State University, Department of Industrial and Systems Engineering, Cognitive Systems Engineering Laboratory .

8

ACCIDENTS INDEX

9

AUTHOR INDEX

10

SUBJECT INDEX