Operating at the Sharp End: The Complexity of Human Error

Article · January 1994

2 authors:

Richard I. Cook
The Ohio State University

David D Woods
The Ohio State University

13 Operating at the Sharp End: The Complexity of Human Error

Richard I. Cook and David D. Woods
The Ohio State University

Studies of incidents in medicine and other fields attribute most bad outcomes to a category of human performance labeled *human error.* For example, surveys of anesthetic incidents in the operating room have attributed between 70% and 82% of the incidents surveyed to the human element (Chopra, Bovill, Spierdijk, & Koornneef, 1992; Cooper, Newbower, Long, & McPeek, 1978). Similar surveys in aviation have attributed more than 70% of incidents to crew error (Boeing Product Safety Organization, 1993). In general, incident surveys in a variety of industries attribute similar percentages of critical events to human error (for example, see Hollnagel, 1993, Table 1). The result is the perception, in both professional and lay communities, of a "human error problem" in medicine, aviation, nuclear power generation, and similar domains. To cope with this perceived unreliability of people, it is conventional to try to reduce or regiment the human's role in a risky system by enforcing standard practices and work rules and by using automation to shift activity away from people.

Generally, the "human" referred to when an incident is ascribed to human error is some individual or team of practitioners who work at what Reason calls the "sharp end" of the system (Reason, 1990; Fig. 13.1). Practitioners at the sharp end actually interact with the hazardous process in their roles as pilots, physicians, spacecraft 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. In order 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 behavior of sharp-end practitioners (Reason, 1990). This chapter examines issues surrounding human performance at the sharp end, including those described as errors and those considered expert.

Most people use the term *human error* to delineate 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 in incident-reporting systems in a variety of industries. For these investigators, human error is a specific variety of human performance that is, in retrospect, so clearly and significantly substandard and flawed that there is no doubt that the practitioner should have viewed it as substandard *at the time the act was committed.* The judgment that an outcome was due to human error is an attribution that (a) the human performance immediately preceding the incident was unambiguously flawed, and (b) the human performance led directly to the outcome.

But the term "human error" is 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 could have or nearly did end with a bad outcome. Thus, these judgments are made ex post facto, with the benefit of *hindsight* about the outcome or near miss. These factors make 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 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 human errors.

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 of people, organizations, and technologies functions and malfunctions (Hollnagel,1993; Rasmussen, Duncan, & Leplat, 1987; Reason, 1990; Woods, Johannesen, Cook, & Sarter, 1994).

Recent research into the evolution of system failures finds that the story of "human error" is markedly complex (Hollnagel,1993; Rasmussen et al., 1987; Reason, 1990; Woods et al., 1994). For example:

- The context in which incidents evolve plays a major role in human performance at the sharp end.
- 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.
- People at the blunt end create dilemmas and shape trade-offs among competing goals for those at the sharp end.
- The attribution of error after the fact is a process of social judgment rather than a scientific conclusion.

The goal of this chapter is to provide an introduction to the complexity of system failures and the term *human error.* It may seem simpler merely to attribute poor outcomes to human error and stop there. If one looks beyond the label, the swirl of factors and issues seems very complex. But it is in the examination of these deeper issues that one can learn how to improve the performance of large, complex systems.

We begin with an introduction to the complexity of error through several exemplar incidents taken from anesthesiology. Each of these incidents may be considered by some to contain one or more human errors. Careful examination of the incidents, however, reveals a more complicated story about human performance. The incidents provide a way to introduce some of the research results about the factors that affect human performance in complex settings such as medicine. Because the incidents are drawn from anesthesiology, most of the discussion is about human performance in the conduct of anesthesia, but the conclusions apply to other medical specialties and even to other domains.

The second part of the chapter deals more generally with the failures of large, complex systems and the sorts of problems those who would analyze human performance in such systems must encounter. It is significant that the results from studies in medicine and other domains such as aviation and nuclear power plant operation are parallel and strongly reinforcing. The processes of cognition are not fundamentally different between practitioners in these domains, and the problems that practitioners are forced to deal with are quite similar. We should not be surprised that the underlying features of breakdowns in these large, complex systems are quite similar.

Grappling with the complexity of human error and system failure has strong implications for the many proposals to improve safety by restructuring the training of people, introducing new rules and regulations, and adding technology. The third part of the chapter explores the consequences of these ideas for attempts to eliminate "human error" as a cause of large, complex system failures.

HUMAN PERFORMANCE AT THE SHARP END

What factors affect the performance of practitioners in complex settings like medicine? Figure 13.1 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:

1. Knowledge factors-factors related to the knowledge that can be drawn on when solving problems in context.
2. Attentional dynamics-factors that govern the control of attention and the management of mental workload as situations evolve and change over time.
3. Strategic factors-the trade-offs between goals that conflict, especially when the practitioners must act under uncertainty, risk, and the pressure of limited resources (e.g., time pressure; opportunity costs).

These three classes are depicted as interlocking rings at the sharp end of the operational system because these functions overlap. Effective system operation depends on their smooth integration within a single practitioner and across teams of practitioners. The figure does not show a single individual because these categories are not assigned to individuals in a one-to-one fashion. Rather, they are distributed and coordinated across multiple people and across the artifacts they use.

These factors govern the expression of error and expertise together with two other classes of factors. First are the demands placed on practitioners by characteristics of the incidents and problems that occur. These demands vary in type and complexity. One incident may present itself as a textbook version of a problem for which a well-practiced plan is available and appropriate. A different incident may appear embedded in a complicated background of interacting factors, creating a substantial cognitive challenge for practitioners. The problem demands shape the cognitive activities of those confronting the incident at the sharp end.

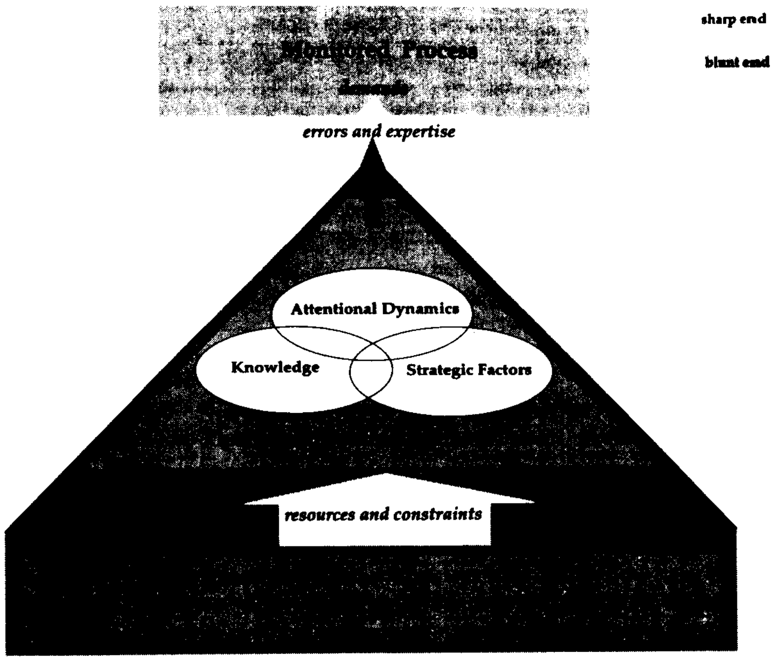

FIG. 13.1. **The sharp and blunt ends of a large complex system. The interplay of problem demands and the resources of practitioners at the sharp end govern the expression of expertise and error. The resources available to meet problem demands are shaped and constrained in large part by the organizational context at the blunt end of the system (from Woods et al., 1994, reprinted by permission).**

The second broad class of factors arises from the blunt end of the system and includes the resources and constraints under which the practitioners function. Recent work on human error has recognized the importance of the ***organizational context*** in system failures (Reason, 1990, chap. 7). This context influences both the physical and cognitive resources available to practitioners as they deal with the system. For example, the knowledge available during system operations is, in part, the result of the organization's investments in training and practice. Similarly, the organizational context influences how easy it is to bring more specialized knowledge to bear as an incident evolves and escalates. Finally, organizational context tends to set up or sharpen the strategic dilemmas practitioners face. Thus, organizational (blunt end) factors provide the context in which the practitioners'

TABLE 13.1
Categories of Cognitive Factors

Category	*Exemplar Incident*	*Cognitive Issues*	*Examples of Conflicts Present*
Knowledge Factors	Myocardial infarction in a vascular surgery patient	• Buggy knowledge • Mental models • Knowledge calibration • Inert knowledge • Simplifications and 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	Goal trade-offs Procedural rules that do not apply to all cases Organizational double binds

(sharp end) knowledge factors, attentional dynamics, and strategic factors function.

This chapter begins with three exemplar incidents. Each was chosen to highlight one of the classes of cognitive factors that are important in human performance as indicated in Table 13.1. Each incident could be judged to contain one or more human errors; this judgment is usually the end point for most investigators who then tabulate the incident frequency in some sort of reporting scheme. Here, however, we take the analysis much further and reveal the complex interplay of the multiple factors sketched in Fig. 13.1 that contributed to the evolution of each incident.

Knowledge Factors: Incident #1 -Myocardial Infarction

> An elderly patient presented with a painful, pulseless, blue arm indicating a blood clot (embolus) in one of the major arteries that threatened loss of that limb. Emergency surgery to perform removal of the clot (embolectomy) was clearly indicated. 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, that is, shortness of breath, dyspnea on exertion and leg swelling (pedal edema). Electrocardiogram changes included inverted T waves. Chest X-ray suggested pulmonary edema. The arterial blood gas showed markedly low oxygen in the arterial blood (P.02 of 56 on unknown FO_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 nitroglycerin drip was started and increased in an effort to reduce the blood pressure. The arterial oxygen saturation (SO_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 abdominal pain and received morphine. Urine output was high in the operating room. The blood pressure continued about 200/100. Nifedipine was given sublingually and the pressure fell over 10 minutes to 90 systolic. The nitroglycerin infusion rate 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 myocardial infarction (heart attack) sometime in the period including the time in the emergency room and the operating room. The patient survived.[1]

This incident raises a host of issues regarding the nature of knowledge and its use during the evolution of the incident. Knowledge factors include those related to the knowledge available for solving problems. Especially important are those factors that conditionalize knowledge toward its use, that is, those that "call knowledge to mind." In Incident #1, it is clear that the participant was employing a great deal of knowledge. In fact, the description of just a few of the relevant aspects of knowledge important to the incident occupies several pages.

There is evidence that the participant was missing or misunderstanding important, but less obvious features of the case. It seems (and seemed to peer experts who evaluated the incident at the time; cf., Cook, Woods, & McDonald, 1991) that the practitioner misunderstood the nature of the patient's intravascular volume, believing the volume was high rather than low. This increased volume is often present in patients with the signs of congestive heart failure. In this case, however, other factors (including the high blood glucose and the prior treatment with a diuretic) were present that indicated that the patient should be treated differently. In retrospect, other practitioners argued that the patient probably should have received more intravenous fluid to replenish the low intravascular volume. They also felt that the patient should have been monitored invasively to allow precise determination of when enough fluid had been given (e.g., a catheter that goes through the heart and into the pulmonary artery).

It is also apparent that many of the practitioner's actions were appropri-

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

ate in the context of the case as it evolved. For example, the level of oxygen in the blood was low and the anesthetist pursued several different means of increasing the blood oxygen level, including the use of oxygen by mask. 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 previous diuretic and the high serum glucose level. It is this last point that appears to have been unappreciated, at first by the physicians who saw the patient initially, and then by the anesthetist.

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. The inability to answer those questions would then have prompted the use of particular monitoring techniques before and during the surgical procedure.

Bringing knowledge to bear effectively in problem solving is a process that involves issues of knowledge *content,* knowledge *organization,* and knowledge *activation.* 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, & 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 knowledge use by humans in complex systems. These include (a) the role of mental models and of knowledge flaws (sometimes called "buggy" knowledge); (b) the issue of knowledge calibration; (c) the problem of inert knowledge; and (d) the use of heuristics, simplifications, and approximations. In many incidents, going behind the label "human error" demands investigating how knowledge was or could have been brought to bear in the evolving incident. Any of the previously mentioned factors could contribute to failures to activate relevant knowledge in context.

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. The term *mental model* has been used to describe the collection of knowledge used by a practitioner. When the mental model is inaccurate or incomplete, its use can give rise to inappropriate actions. These mental models are described as "buggy" (see Chi, Glaser, & Farr,1988; Gentner & Stevens, 1983; and Rouse & Morris, 1986, for some of the basic results on mental models). Studies of practitioners' mental models have 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, & McDonald, 1991). Significantly, the design of devices, particularly the interface between the device and human practitioners, can either aid or impede the development of useful mental models by practitioners. The presence of a buggy mental model of a device is more likely to indicate poor device design than it is some inadequacy of the user's mental machinery (Norman, 1988).

It is possible to design experiments that reveal specific bugs in practitioners' mental models. By forcing pilots to deal with various nonnormal situations, 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. It can be quite difficult to determine whether a buggy mental model was, indeed, involved in an incident. In the exemplar incident, for example, the combination of congestive heart failure (normally improved by reducing the amount of fluid in the circulation) with high urine output from high blood glucose and a diuretic drug (furosemide) was unusual. It is not clear whether the practitioner had a buggy mental model of the relationship between these factors or if the demands of attention to the low oxygen saturation and blood pressure prevented him from examining the model closely enough to discover the relationship. Alternatively, the mental model and associated knowledge may simply have been inert (see the section on inert knowledge). The inability to distinguish between these alternatives is due, in large part, to the limitations of the data about the incident.

Knowledge Calibration. Results from several studies (Cook, Potter, Woods, & McDonald, 1991; Moll van Charante, Cook, Woods, Yue, & Howie, 1993; Sarter & Woods, 1994) indicate that practitioners may be unaware of gaps or bugs in their model of a device or system. This raises the question of knowledge calibration (Wagenaar & Keren, 1986). Everyone has some areas where their knowledge is more complete and accurate than others. 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?

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 effect knowledge calibration (e.g., Wagenaar & Keren, 1986).

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

---

[1] One physician was recently heard to describe another as being "often wrong but never in doubt," an indication that practitioners may recognize the presence of a calibration problem.

Activating Relevant Knowledge in Context The Problem of Inert Knowledge. Lack of knowledge or buggy knowledge may be one part of the puzzle, but the more critical question may be factors that affect whether relevant knowledge is activated for use in the actual problem-solving context (e.g., Bransford et al, 1986). The question is not just whether the problem solver knows some particular piece of domain knowledge, but whether he or she calls it to mind when it is relevant to the problem at hand and whether he or she knows how to use this knowledge in problem solving. We tend to assume that if a person can be shown to possess a piece of knowledge in any circumstance, then this knowledge should be accessible under all conditions where it might be useful. In contrast, a variety of research has revealed dissociation effects where knowledge accessed in one context remains inert in another (Gentner & Stevens, 1983; Perkins & 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, that is, it was not summoned up during the incident.

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 as might be determined by standardized tests. Rather, the more stringent criterion is that situation-relevant knowledge is accessible under the conditions in which the task is actually performed. Thus, *inert knowledge* is knowledge accessible only in a restricted set of contexts, which may not include contents of relevance to actual practice. Inert knowledge may be related to cases that are difficult not because problem solvers do not know the individual pieces of knowledge needed to build a solution, but because they have not previously confronted the need to join the pieces together. 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 *not to 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 clearly possessed knowledge because they were 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 knowledge 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. Alternatively, 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 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."

Heuristics, Simplifications and the Imprecision of Knowledge. During the past decade, there has been much written about medical decision making, and a large portion of it is highly critical of the decision processes of practitioners. People tend to cope with complexity through simplifying heuristics, that is, through rules of thumb and simplifications. Heuristics are useful because they are usually relatively easy to apply and minimize the cognitive effort required to produce decisions. Heuristics can readily be shown to be incorrect under some circumstances (Tversky & Kahneman, 1974)[1] and, in theory, are less desirable as decision rules than precise computations, at least if the decision maker is considered to have infinite mental resources for computation. However, these simplifications may also be useful approximations that allow limited-resource practitioners to function robustly over a variety of problem demand factors (Woods, 1988).

At issue is whether a simplification is (a) generally useful because it reduces mental workload without sacrificing accuracy, or (b) a distortion or misconception that appears to work satisfactorily under some conditions but leads to error in others. The latter class is described by Feltovich et al. (1989) as an *oversimplification.* In studying the acquisition and representation of complex concepts in biomedicine, Feltovich et al. 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:

1. Seeing different entities as more similar than they actually are.
2. Treating dynamic phenomena statically.
3. Assuming that some general principle accounts for all of a phenomenon.
4. Treating multidimensional phenomena as unidimensional or according to a subset of the dimensions.
5. Treating continuous variables as discrete.

6. Treating highly interconnected concepts as separable.
7. Treating the whole as the sum of its parts (see Feltovich, Spiro, & Coulson, 1993).

These oversimplifying tendencies may occur because of requirements for cognitive effort 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)

Criticisms of practitioner decision making based on simplified or oversimplified knowledge are often used to show that practitioners make bad decisions and that their decision making would be improved by adopting a more mathematically rigorous, probabilistic reasoning approach. It can be shown mathematically, for example, that a particular strategy for contingent choice using strict criteria would be preferable to many other strategies. Such demonstrations are usually sterile exercises, however, for several reasons. First, the effort required to perform such calculations 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 and colleagues (Payne, Bettman, & Johnson, 1988; Payne, Johnson, Bettman, & Coupey, 1990) who demonstrated that simplified methods will produce a higher proportion of correct choices between multiple alternatives under conditions of time pressure. Put simply, if the time and effort required to arrive at a decision is important, it may be possible to have an overall higher quality performance using heuristics than using a "mathematically ideal" approach.

The second reason that it is difficult to rely on formal decision making methods is that medical knowledge is so heterogeneous and imprecise. Much medical research data are drawn from small or only marginally representative samples; drug tests rarely include pregnant women, for example, so the effects of many drugs on pregnant women and fetuses are unknown. Much patient data are derived from coarse measurements at widely spaced intervals, whereas others (for example, the effects of exposure to anesthetic agents) are known precisely but only for a limited period of time. Thus it is possible to have quite precise knowledge about the effect of a disease or a treatment on a specific subset of patients and also to have a great deal of uncertainty about the extent to which that knowledge is useful for a given patient both because the knowledge is derived from a specific subgroup and because the patient is poorly characterized. Many important physiologic

---

[1] Indeed, if a rule of thumb is not inaccurate in some circumstance then it is a robust rule and not a heuristic at all.

variables can be measured only indirectly with poor precision and are known to fluctuate widely even in the healthy population. Physicians often must rely on comparatively remote or indirect measures of critical variables. The precise effect of a therapy is usually only predictable for a group of patients; for example, a preoperative antibiotic will reduce the risk of postoperative infection by a small amount, but the actual benefit to an individual patient coming to the operating room for a specific procedure is extraordinarily difficult to define. All these factors tend to lead medical practitioners toward an empirically based approach to diagnosis and therapy in which successive treatments are applied until the desired result is achieved.

There are also inherent conflicts in the knowledge base that need to be resolved in each individual case by the practitioner. In Incident #1, for example, there are conflicts between the need to keep the blood pressure high and the need to keep the blood pressure low (Fig. 13.2). The heart depends on blood pressure for its own blood supply, but increasing the blood pressure also increases the work it is required to perform. The practitioner must decide what blood pressure is acceptable. Many factors enter into this decision process: What is the patient's normal blood pressure? How labile is the blood pressure now? How will attempts to reduce blood pressure affect other physiological variables? How is the pressure likely to change without therapy? How long will the surgery last? What is the level of surgical skill being employed? As is often the case in this and similar domains, the locus of conflict may vary from case to case and from moment to moment. It is impossible to create algorithms that adequately capture the variety of patient characteristics and risks in a highly uncertain world. These conflicts are a normal part of the medical domain and practitioners are so comfortable with them that it is hard to get the participants in an incident to be explicit about the trade-offs involved in the decisions they made.

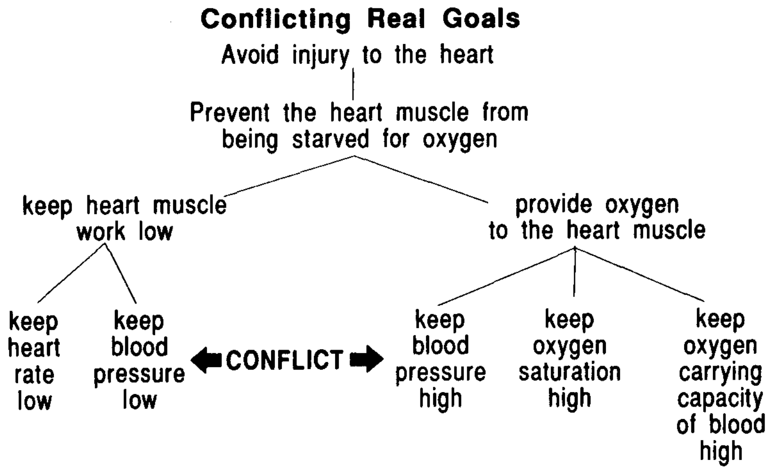

FIG. 13.2. Conflicting domain knowledge. For a cardiac surgery patient the blood pressure should be kept low to minimize the work of the heart, but the blood pressure should be kept high to maximize the blood flow to heart muscle. Flow practitioners at the sharp end resolve this conflict depends on several factors (from Cook, Woods, & McDonald, 1991, reprinted by permission).

In summary, heuristics may represent effective and necessary adaptations to the demands of real workplaces (Rasmussen, 1986). The problem, if there is one, may not always be the shortcut or simplification itself, but whether practitioners know the limits of the shortcuts, can recognize situations where the simplification is no longer relevant, and have the ability to use more complex concepts, methods, or models (or the ability to integrate help from specialist knowledge sources) when the situation they face demands it. 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 anesthetists commenting on the practitioner's behavior in the first incident were critical of his performance:

> 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.

This critique is not quite correct. In fact, each problem was addressed in some way at some time. But the comment about "coming in from each side" identifies what the practitioner was missing in the incident, that is, the interactions between normally separate factors that here were closely linked. Being able to discover that link and appreciate its implications is intimately bound up with knowledge factors including mental models, heuristics, and inert knowledge.

Attentional Dynamics: Incident #2-Hypotension

> During a coronary artery bypass graft procedure, an infusion controller device delivered a large volume of a potent drug to the patient 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 tort) 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, hypertension gave way to hypotension (systolic pressure <60 tort). The hypotension was unresponsive to fluid challenge but did respond to repeated boluses 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 the 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 a closed downstream stopcock. The device was started, but the machine did not detect any flow of drug (the stopcock was closed), triggering visual and auditory alarms. When the stopcock was opened, free flow of fluid containing drug began. The controller was restarted, but the machine again detected no drops because the flow was wide open and no individual drops were formed. The controller alarmed again, with the same message, which appeared to indicate that no flow had occurred. Between the opening of the stopcock and the generation of the error message, sufficient drug was delivered to substantially reduce the blood pressure. The operator saw the reduced blood 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'

This incident is used as an exemplar for the discussion of attentional dynamics, although it also involves a number of issues relevant to knowledge factors. Attentional dynamics refers to those factors affecting cognitive function in dynamic evolving situations, especially those involving the management of workload in time and the control of 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, especially with

---

This case is described more fully in Cook, Woods, and Howie (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).

respect to error. In dynamic, event-driven environments like the operating room, attentional factors are often crucial in the evolution of incidents (cf. Gopher, 1991; Hollister, 1986; Woods, 1992).

In Incident #2, the data are strong enough to support a reconstruction of some of the actual changes in focus of attention of the participants during the incident. A collection of infusion devices like those involved in the incident are shown in Fig. 13.3. The free flow of the drug began when one of the physicians opened the stopcock downstream of the affected device, but this source of the hypotension was not identified until the bag of fluid was nearly empty. There are a number of factors in the environment that contributed to the failure to observe (i.e., attend to) the unintended flow of drug via the infusion device, including: (a) the drip chamber being obscured by the machine's sensor, making visual inspection difficult, (b) presence of an aluminum shield around the fluid bag, hiding its decreasing size, (c) misleading alarm messages from the device, and (d) presence of multiple devices, making it difficult to trace the tubing pathways.

There are also extra-environmental factors that contributed 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 been 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 been revisited had the situation been less demanding, but the fall in blood 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 drug 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 bags on the infusion support tree until the decision was made to start an infusion of the antagonist drug and a bag for that drug was being 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 surgery patient. In this case, successful diagnosis of the cause was less important than successful treatment of the consequences of the problem. The practitioners were quick to correct the physiologic, systemic threat even though they were unable to diagnose its source. This shift from diagnosis to what Woods (1988, 1994)

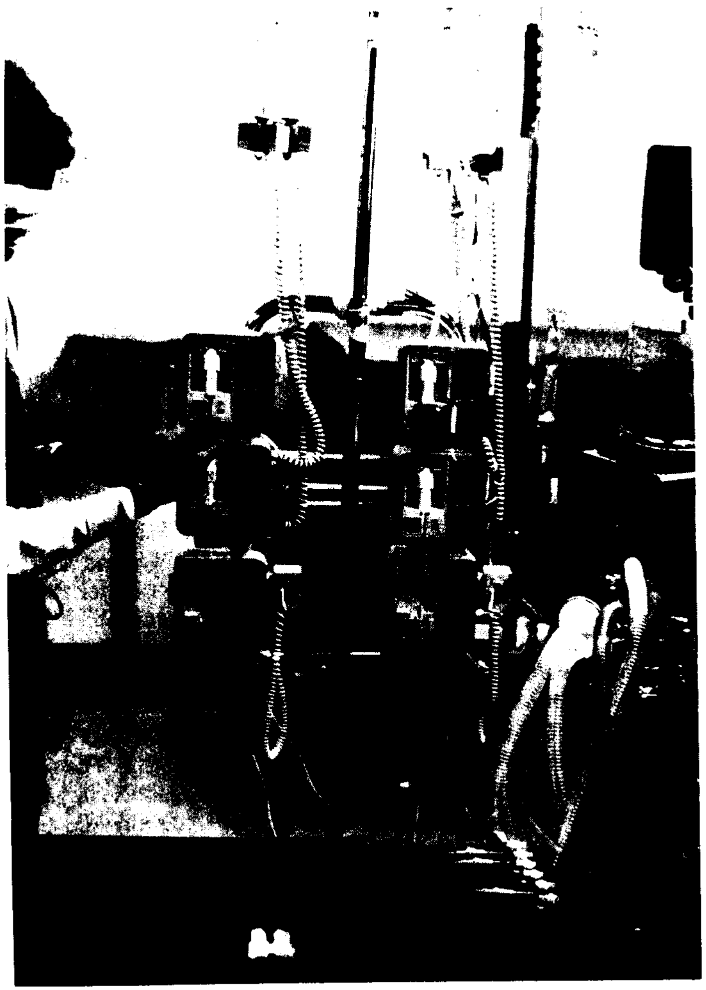

FIG.13.3. A set-up of multiple drug Infusion devices in the heart room. Drugs to raise and lower blood pressure and other cardiovascular system parameters are In the fluid bags above. The controller boxes regulate flow through the tubing based on the detection of fluid drops in drip chambers connected to the bags. The Individual flows are joined together by a series of stopcocks to a single piece of tubing, which is then connected to the patient. (See Moll van Charente et al., 1993, for additional details.)

calls *disturbance management* is crucial in the operating room and in other domains to maintaining the system in a stable configuration to permit later diagnosis and correction of the underlying faults.

The control of attention is an important issue for those trying to understand human performance, especially in event-rich domains such as the operating room. 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 *cognitive cycle* (Neisser, 1976; Tenney, Jager Adams, Pew, Huggins, & Rogers, 1992). 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 major human performance problems can arise from alterations in 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 attentional dynamics (Sarter & Woods, 1991; Tenney et al., 1992). Just a few of the cognitive processes that may pass under the label of situation awareness are: *control of attention* (Gopher, 1991), *mental simulation* (Klein & Crandall, in press), *directed attention* (Woods, 1992), and *contingency planning* (Orasanu, 1990). Because the concept involves tracking processes in time, it has also been described as *mental bookkeeping* to track multiple threads of different but interacting subproblems (Cook, Woods, & McDonald, 1991; Dorner, 1983). These terms refer to tracking the shifting pattern of interactions in the system under control. For example, the state of chemical paralysis of the patient and the "depth" of anesthesia are two different threads. Normally these may be treated independently, but under some circumstances they may interact in ways that have implications for the future course of the patient.

Maintaining situation awareness necessarily requires shifts of attention between the various threads. It also requires more than attention alone, for the object 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 requires 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 & Rousseau, 1989). In each case what is being lost is some of 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.

Obtaining a clear, empirically testable model for situation awareness is difficult. For example, Hollister (1986) presented 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 & Woods, 1991).

Failures to Revise Situation Assessments: Fixation or Cognitive Lockup. The results of several studies (Cook, McDonald, & Smallhout, 1989; De Keyser & Woods, 1990; Gaba & DeAnda, 1989; Johnson, Moen, & Thompson, 1988; Johnson & Thompson, 1981; Woods, O'Brien, & Hanes, 1987) strongly suggest that one source of error in dynamic domains is *a failure to revise* situation assessment as new evidence comes in. Evidence discrepant from the agent's or team's current assessment is missed or discounted or rationalized as not really being discrepant with the current assessment. In addition, it seems that several major accidents involved a similar pattern of behavior from the operational teams involved; 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 based on 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 (Klein, Orasanu, Calderwood, & Zsambok, 1993; Woods & Roth, 1988).

In psychological fixations, 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. *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 lack of knowledge or other problems that impair error detection and recovery (Cook et al., 1989)[1] 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 practitioners seem to have many hypotheses in mind, but never entertain the correct one. The external behavior looks incoherent because they are jumping from one action to another without any success. The second pattern of behavior 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. The persistence in practitioner behavior can be remarkable. For example, practitioners may repeat the same action or recheck the same data channels several times. This pattern is easily identified because of the unusual number of repetitions despite an absence

---

[1] 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.

of results. The practitioners often detect the absence of results themselves but without any change in strategy. A third pattern is "everything is OK" (Perrow, 1984). Here the practitioners do not react to the change in their environment. Even if there are a lot of cues and evidence that something is going wrong, they do not seem to pay much attention to them. The practitioners seem to discount or rationalize away indications that are discrepant with their model of the situation.

There are certain types of problems that may encourage fixations by mimicking other situations. This, in effect, **leads practitioners down *a garden path.* 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 (locally rational strategies given difficult problems and limited resources), rather than to the any particular bias or problem in the strategies used. Simply put, going down a garden path is not an "error" per se. It is how the problem presents to practitioners that makes it easy to entertain plausible but erroneous possibilities. Anesthesiology and similar domains have inherent uncertainties in diagnostic problems, and it may be necessary for practitioners to entertain and evaluate what turn out to be erroneous assessments. Problems arise when the revision process breaks down and the practitioner becomes fixated on an erroneous assessment, missing, discounting or reinterpreting discrepant evidence (see Johnson et al., 1988; Roth, Woods, & 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 in order to keep up with a changing incident.

Fixation is a characteristic of practitioners in an incident. There are several cognitive processes involved in attentional dynamics that may give rise to fixation:

1. Breakdowns in shifting or scheduling attention as the incident unfolds.
2. Factors of knowledge organization and access that make critical knowledge inert.
3. Difficulties calling to mind alternative hypotheses that could account for observed anomalies-problems in the processes underlying hypothesis generation.
4. Problems in strategies for situation assessment (diagnosis) given **the probability of *multiple* factors, for example, how to value parsimony** (single-factor assessments) versus multifactor interpretations.

Fixation may represent the down side of normally efficient and reliable cognitive processes involved in diagnosis and disturbance management in dynamic contexts. Although fixation is fundamentally about problems in attentional dynamics, it may also involve inert knowledge (calling to mind potentially relevant knowledge such as alternative hypotheses) or strategic factors (trade-offs about what kinds of explanations to prefer).

It is clear that in demanding situations where the condition of the patient and the operating room system 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.

Strategic Factors: Incident **#3-Busy** Weekend Operating Schedule

On a weekend in a large tertiary care hospital, the anesthesiology team (consisting of four physicians of whom three are 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 10 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.

The third incident is remarkable in part because 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 intubate patients on the floor, requests for pain control, and emergency room trauma cases. The in-charge person also serves as a backup resource for the operations in progress. In this incident, the anesthetist in charge committed all of her available resources, including herself, to doing anesthesia. This effectively eliminated the in-charge person's ability to act as a buffer 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 the 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 are part of the normal playing field of the specialty.

The conflicts and their resolution presented in Incident #3 and the trade-offs between highly unlikely but highly undesirable events and highly likely but less catastrophic ones are examples of strategic factors. People have to make trade-offs between different but interacting or conflicting goals, between values or costs placed on different possible outcomes or courses of

action, and between the risks of different errors. People make these trade-offs when acting under uncertainty, risk, and the pressure of limited resources (e.g., time pressure, opportunity costs). One may think of these trade-offs in terms of simplistic global examples like safety versus economy. Trade-offs also occur on other dimensions. In dynamic fault management, for example, there is a trade-off 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. 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 trade-offs can also involve coordination among agents in the distributed human-machine cognitive system (Roth, Bennett, & Woods, 1987). A machine expert recommends a particular diagnosis or action, but 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? The pulse oximeter may provide an unreliable reading, especially when perfusion is poor and the oxygen saturation is low. Is the current reading of 80% indicative of an artifact or an accurate representation of the patient's oxygen saturation?

Criterion setting on these different trade-offs may not be a conscious process or a decision made by individuals. More likely, it may be an emergent property of systems of people, either of 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 trade-offs 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 look to be simple blunders. Indeed, most of the media reports of "human error in medicine" focus on such cases. But a more careful assessment of the distributed system including the patient, physicians, and the larger institutions comprising the hospital may reveal strategic factors at work. Behavior in the specific incident derives from how the practitioners set their trade-off criteria across different kinds of risks from different kinds of incidents that could occur. Because incidents are evaluated as isolated events, such trade-offs 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 later discussion of hindsight). There are many strategic factors that 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, 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 trade-offs being made, and the pressures present that shift the operating points for these trade-offs.

The impact of potential conflicts may be quite difficult to assess. Consider the anesthetist. 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 (Fig. 13.4).

In a given circumstance, the relationships between these goals can produce conflicts. In the daily routine, for example, maximizing patient safety and avoiding lawsuits creates the need to maximize information about the patient through preoperative workup. The anesthetist may find some hint of a potentially problematic condition and consider 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 (Fig. 13.4). 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 surgeons' 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. Increasing external economic pressure, in particular, can generate sharp conflicts in anesthesiology and in other areas of medicine (Eddy, 1993a, 1993b).

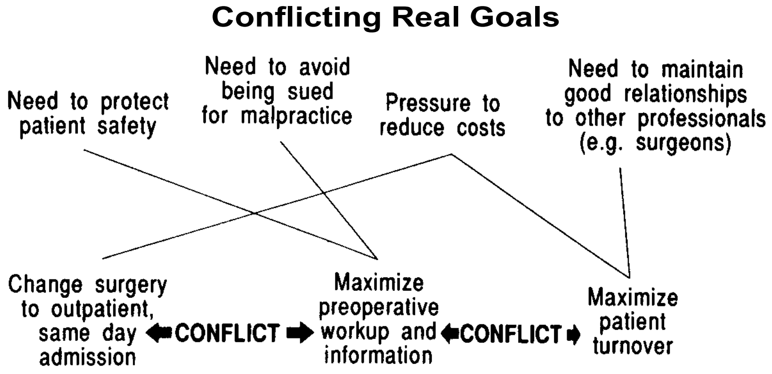

FIG. 13.4. Conflicting goals in anesthesiology. Maximizing patient safety and avoiding lawsuits creates the need to maximize information about the patient through preoperative workup. The cost-reduction goal provides an incentive for the use of same-day surgery and limits preoperative workup. The anesthetist may be squeezed in this conflict (from Cook, Woods, & McDonald, 1991, reprinted by permission). Compare this conflict with the one shown in Fig. 13.2.

For an example from outside of medicine, consider the task of en route flight planning in commercial aviation. Pilots sometimes need to modify their flight plans en route when conditions change (e.g., weather). Some of the goals that need to be considered 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 trade-offs. Layton, Smith, and McCoy (in press) 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 deiced and then enters the queue for takeoff. After the aircraft has been deiced, the effectiveness of the deicing 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 deiced again will cause additional delays, plus the aircraft will have to re-enter the takeoff queue again. Thus, the organization of activities (where deicing occurs relative to queuing 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 circumstances), 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 involve economic pressures but also intrinsic characteristics of the field of activity. 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 (Fig. 13.2). Specific actions will depend on details of the context. The appropriate blood pressure target adopted by the anesthetist depends in part on the individual's strategy, the nature of the patient, kind of surgical procedure, circumstances within the case that may change (e.g., the risk of major bleeding), and negotiation 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 the organizational or social context represent another source of goal competition. Some of the organizational factors producing goals include management policies, legal liability, regulatory guidelines, and economic 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 (National Transportation Safety Board, 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., career advancement, avoiding conflicts with other groups).

It should not be thought that the organizational goals are necessarily simply the written policies and procedures of the institution. Indeed, the messages received by practitioners about the nature of the institution's goals may be quite different from those that management acknowledges. Many goals are indirect and implicit. Some of the organizational influences on how practitioners will negotiate their way through conflicting goals may not be explicitly stated or written anywhere. These covert factors are especially insidious because they affect behavior and yet are unacknowledged. For example, the Navy sent a clear message to its commanders by the differential treatment it accorded to the commander of the *Stark* following that incident (U.S. House of Representatives Committee on Armed Services, 1987) as opposed to the *Vincennes* following that incident (Rochlin, 1991; U.S. Department of Defense, 1988).

In Incident #3, economic factors, intrinsic characteristics of the domain of practice, and organizational factors all contributed to the goal conflicts the practitioner faced.

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 adequately. Some practitioners will not follow up hints about some aspect of the patient's history because to do so would impact the usual practices 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 stance and order tests for minor indications, even though the yield is low, in order to be on the safe side. This generates increased costs and incurs the wrath of their surgical colleagues for the delays thus generated. In either case, the nature of the goals and pressures on the practitioner are seldom made explicit and rarely examined critically.

In postincident analysis, in hindsight, the consequences will be apparent. It should be clear, however, that the external pressures for highly efficient performance are strongly in favor of limiting the preoperative workup of patients and omitting tests that are unlikely to yield important findings. Assessments after the incident will always identify factors that if changed would have produced a more favorable result; large, complex systems always have many such factors available for scrutiny. Thus, if those practitioner actions that are shaped by the goal conflict contribute to a bad outcome in a specific case, then it is easy for postincident evaluations to say that a human error occurred-the practitioners should have delayed the surgical procedure in order to investigate the hint. The role of the goal conflict may never be noted.

To evaluate the behavior of the practitioners involved in an incident, it is important to elucidate the relevant goals, the interactions between 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 helping 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 trade-off 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 medical practice, 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 medical practice 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 a trade-off, from hindsight they are vulnerable to charges of and penalties for error. In particular, control via regimentation and bureaucratically derived policies (just follow the procedures) or the introduction of machine-cognitive agents that automatically diagnose and plan responses, can undermine the effective authority of the practitioners on the scene. However, these same people may still be responsible and held accountable both formally and informally for bad outcomes. The results of research on the role of responsibility and authority are limited but consistent-splitting authority and responsibility appears to have bad consequences for the ability of operational systems to handle variability and surprises that go beyond preplanned routines (Hirschhorn, 1993; Roth et al., 1987).

There is one important investigation of the effects of responsibility-authority double binds in the industrial literature. Hirschhorn (1993) examined an organization's (i.e., the managers) 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. In part this is the result of the regulatory climate that believes that absolute adherence to procedures is the means to achieve safe operations and avoid "human error." This creates conflicts in some situations and generates dilemmas for the people involved. If they follow the standard procedures strictly, the job will not be accomplished adequately; if they always wait for formal permission to deviate from standard procedures, throughput and productivity will degrade substantially. If they deviate and it later turns out that there is a problem with what they did (e.g., they did not adapt adequately), they may create safety or economic problems. The double bind arises because they are held responsible for the outcome (the bad outcome, the lost productivity, the erroneous adaptation) but do not have authority for the work practices because they are expected to comply exactly with the written procedures. Notice the similarity to the evolving nature of medical practice today, with the introduction of increasing regulation and so- called "practice parameters" (Arens, 1993).

After the Three Mile Island accident, utility managers were encouraged by the Nuclear Regulatory Commission to develop detailed and comprehensive work procedures to reduce the likelihood of another major disaster. The management at a particular nuclear power plant instituted a policy of verbatim compliance with the procedures developed at the blunt end of the system. 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, contradictory, and buggy, and (b) novel circumstances arose that were not anticipated in the written 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. As Hirschhorn (1993) said:

> They had much responsibility, indeed as licensed professionals many could be personally fined for errors, but were uncertain of their authority. What freedom of action did they have, what were they responsible for? This gap between responsibility and authority meant that operators and their supervisors felt accountable for events and actions they could neither influence nor control.

Workers coped 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 the procedures, but the process was cumbersome and time-consuming. This is not surprising: If modifications are easy and liberally granted, then it may be seen as undermining the policy of strict procedure following. The increasingly complex and bureaucratic policies and procedures of U.S. hospitals seems likely to generate a situation similar to that described by Hirschhorn.

The n-Tuple Bind

The three incidents that have been described 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, and strategic factors) plays a role in the conduct of anesthesia 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 (Table 13.1).

Together the conflicts produce a situation for the practitioner that appears to be a maze of potential pitfalls. This combination of pressures and goals that produce a conflicted environment for work is what we call *the n-tuple bind.*[1] The practitioner is confronted with the need to follow a single course of action from a myriad of possible courses. The choice of 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 overconstrained 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. 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, because of the hindsight bias, 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 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 the patient's "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 (in the sense of the path that minimizes risk to the patient) misrepresents the dilemmas that confront the practitioner. The safest anesthetic is the one that is not conducted, just as 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 of cause and effect by retrospective reviewers is discussed next.

SYSTEM FAILURES AND HUMAN PERFORMANCE

Large, Complex System Failures: The Latent Failure Model

The spectacular failures of large, semantically complex, time-pressured, tightly coupled, high consequence, high-reliability systems[1] have prompted the study of how such systems fail and the role of human operators in successful and unsuccessful operation. The complexity of these systems arises in large part from the need to make them reliable. All such complex systems include potentially disastrous failure modes and are carefully crafted to reduce the risk of such failures. Significantly, these systems usually have multiple redundant mechanisms, "safety" systems, and elaborate policies

---

[1] 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, an extension of the notion of a *double bind.*

[1] These failures include the explosion of *Apollo 13,* the destruction of the space shuttle *Challenger,* the *Herald of Free Enterprise* ferry capsizing, the Clapham Junction railroad disaster, the grounding of the tanker *Exxon Valdez,* a number of airplane crashes, the reactor explosion at Chernobyl, and a host of other nuclear power incidents, most particularly the destruction of the reactor at Three Mile Island. Some of these are reviewed in Perrow (1984) and Reason (1990).

and procedures to keep them from failing in ways that produce bad outcomes.

The results of combined operational and technical measures make systems relatively safe from single-point failures; that is, they are protected against the failure of a single component or procedure. For example, the routine oxygen and nitrous oxide supply for anesthesia machines is derived from a hospital-wide pipeline. Each machine, however, has its own supply tanks available as a backup should the hospital supply fail, as well as elaborate valving mechanisms to insure automatic switch over to the cylinder supply. There are even special backups designed to shut off the flow of nitrous oxide (which will not support life) if the oxygen pressure falls below a preset level. In addition, the machines are gas powered and will operate even if external electrical supplies are lost. Of course, there are components and procedures that cannot be protected through redundancy. An example of such a component is the nuclear power plant's reactor containment building. The building is critical to plant safety and there is only one, but it is lavishly constructed to withstand extreme events. Similarly, the anesthesia machine has internal piping and mechanisms that make the machine vulnerable to single-point failures, although these failures are few and the components are conservatively designed (Andrews, 1990).

When large system failures do occur, they are the result of multiple, apparently innocuous faults that occur together (Perrow, 1984; Reason, 1990; Turner, 1978). All complex systems have many such apparently minor faults. These can include such simple items as a burned out indicator bulb on a seldom-used control panel, a checklist that is out of date because the corresponding equipment has been modified, or an apparently minor failure of some backup system (for example, an emergency generator). For the most part, the minor faults are inactive, play no role in system operation, and are therefore described as *latent failures* (Reason, 1990). These latent failures may be found at any level within an organization from the corporate boardroom to the individual physical components of the system. System failures occur when a particular collection of latent failures are present together in a combination that allows the system to fail. Rather than being derived from the massive failure of a single component, system failures arise from the insidious accumulation of individual faults, each of which seems too small or unimportant to threaten the larger system. Thus *Challenger* failed because of the combination of the brittle O-ring seals *and* the unexpectedly cold weather *and* the reliance on the seals in the design of the boosters *and* the change in roles between the contractor and the NASA engineers *and* other factors. None of these conditions was individually able to provoke a system failure, but together they were able to disrupt an extraordinarily safety-oriented system in a catastrophic way. In the field of aviation, a combination of factors were responsible for the simultaneous failure of all three engines of an L-1011 jumbo jet (Norman, 1992). In medicine, a similar case can be found in the failure of the Therac-25 radiation therapy machine. This device would, under certain highly unusual circumstances, deliver huge doses of radiation to patients. These circumstances, although unlikely, did arise and injure several patients. Review of the design and use of the Therac-25 showed that multiple small failures in software, in testing, and in operator procedures were required to generate the system failure (Leveson & Turner, 1993). It is important to note that the latent failures can involve technology, operational practices, and even organizational elements: Latent failures can occur in any part of the larger system (Reason, 1990).

These large system failures have several notable characteristics. First, they are *complex*. Large system failures are comprised of multiple failed components or procedures. Predicting this combination is likely to be difficult or impossible for human operators; the failure mode is hard to foresee and prevent. Second, failures are likely to be *catastrophic* rather than minor. The multiple redundancies and robust design characteristics of a large system tend to limit small-scale failures and to minimize their consequences. In addition, the cost of the so-called safety systems and redundancies generally encourages the development of ever larger and more economically efficient systems in order to reduce the average cost of each unit of performance. Thus, it is not an oil tanker accident but supertanker accident, not a plane crash but a jumbo jet crash, not an overdose of radiation but a massive overdose.

Third, the potential for catastrophic failure encourages the employment of *human skill* and expertise at the final few links in the causal chain of events. The more delicate the system, the more important its function, the more often a person will be charged with protecting the system's integrity or accomplishing some critical goal. Moreover, the systems are so complex and are operated under such variable conditions that only human operators can be expected to have both the flexibility and judgment necessary to control them. Fourth, because disasters are composed of a collection of latent failures occurring together, large system failures appear in retrospect to be unique. After-accident reviews will show that the failure depended on having a particular pattern of small faults. As the number of latent failures required to produce the system failure increases (i.e., as the system becomes, in some sense, "safer"), the odds against repeating a precise pattern of failure become astronomical. Paradoxically, this will make *any* future system failure seem extremely unlikely, even though the accumulation of latent failures actually makes the system more failure prone.

Gaba's group (Gaba, Maxwell, DeAnda, 1987) at Stanford suggested that this model of large system behavior and failure might apply to anesthesia practice in particular and, by extension, to medical practice in general. He noted that anesthesia practice includes many of the characteristics of complex systems and that the infrequent anesthesia mishaps appeared to be similar to the disasters studied in other complex systems. The anesthesiologist works in a highly complex, technologically intensive environment. The conduct of an anesthetic is a critical process that is severely time pressured, and the elements of the system are tightly coupled together in ways that do not provide much slack.[1] Cooper's group (Cooper et al., 1978; Cooper, Newbower, & Kitz, 1984) at Massachusetts General Hospital noted that anesthesia incidents appeared to be unique and were difficult to analyze, exactly as one would expect in a system that had been refined to eliminate single-point failures. Significantly, the loci of single-point failures in the conduct of the anesthetic have been studied and largely removed or buffered by redundant components or safety procedures. Although their study predated the latent failure model of large system failure, Cooper's data also indicated that critical incidents that progressed toward bad outcomes required multiple, simultaneous failures. Thus, there are reasons to consider that anesthesia practice and, by extension, modem medical practice, has the characteristics of a large, complex system and may be expected to fail in similar ways. A recent study supports this view (Cook, Woods, & McDonald, 1991).

One consequence of the latent failure model of large system failures is that efforts to improve the overall system performance by "fixing" particular latent failures that contributed to a past mishap are unlikely to markedly reduce the accident rate. Because that particular pattern of contributors is so unlikely to recur and because there are many unrecognized latent failures that remain in the system, the correction of one set of specific flaws is of limited value. The usual response to a system failure is to attempt to make certain that it doesn't happen again by producing new rules and regulations, new equipment, and new training of key personnel. Because the exact set of flaws is unlikely to recur, these attempts will add more cost and make the system even more complex and brittle than it was before the accident. Because the system is already highly reliable, some time will pass between instituting these changes and the next accident. This leads those who promulgated the changes to believe that they have significantly improved the system safety ("after all, it hasn't failed like that since we instituted our program X"). After some time, however, another accident occurs, but this time with a different sequence of events derived from a different collection of latent failures. This apparently unique accident is seen in isolation and the cycle is repeated (see Fig. 13.5).

---

[1] For a detailed discussion of coupling, see Perrow,1984, chap. 3 and especially Table 3.2. It is interesting to note that Perrow does not include the operating room and anesthesia in his Fig. 3.1, although based on our studies, it would lie somewhere between aircraft and nuclear plants.

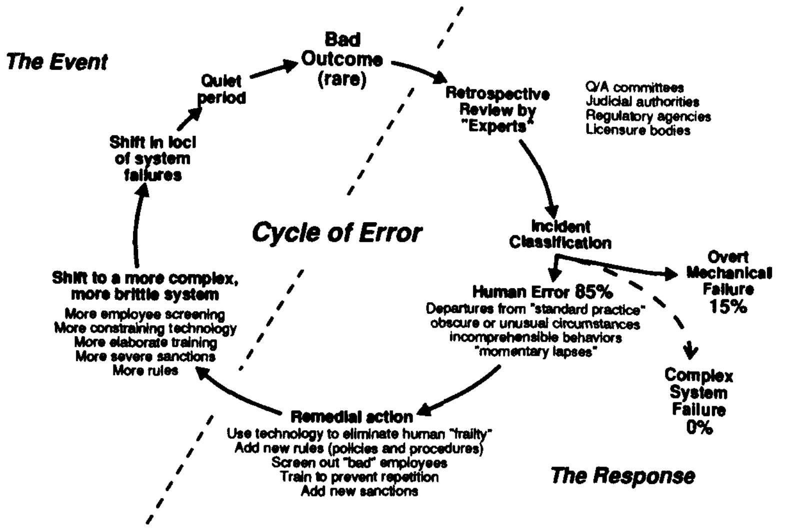

FIG.13.5. The cycle of error. Attributing system failures to human operators generates demands for more rules, automation, and polkaing. But these actions do not significantly reduce the number of latent failures In the system. Because overt failures are rare, a quiet period follows Institution of these new policies, convincing administrators that the changes have been effective. When a new overt failure occurs, it seems to be unique and unconnected to prior failures (except In the label human am), and the cycle repeats. With each pass through the cycle, more rules, policies, and sanctions make the system more complicated, conflicted, and brittle, increasing the opportunities for latent failures to contribute to disasters. (01993, R.I. Cook, reprinted by permission)

Retrospective Evaluations of Human Performance in System Transients

Attributing System Failures to Practitioners. System failures, near failures, and critical incidents are the usual source for investigations of human performance. When critical incidents do occur, operator failure or human error will almost always be indicted as a major cause of any bad 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 70% or 75%. Incident rates attributed to human error are the same in several domains including aviation, nuclear power, shipping, and, most recently, in anesthesia and medicine (cf., Hollnagel, 1993). Cooper et al. (1978) found that anesthesiologists were contributors in 82% of critical incidents. Wright, Mackenzie, Buchan, Cairns, and Price (1991) and Chopra et al. (1992) found similar rates in the operating room and intensive care unit, respectively. The repeated finding of about three fourths of incidents arising from human error has built confidence in the notion that there is a problem with human error in these domains. Indeed, it is the belief that fallible humans are responsible for large system failures that 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 errors.

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 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 (Woods et al., 1994).

For at least four reasons, it is actually not surprising that human operators are blamed for bad outcomes. First, operators are available to blame. These large and intrinsically dangerous systems have a few well-identified humans at the sharp end. Those humans are closely identified with the system function, and so it is unlikely that a bad outcome will occur without having them present. Moreover, these individuals are charged, often formally and institutionally, with maintaining the system's 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) and available to blame.

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 that led to the system failure (Rasmussen, 1986). It is particularly difficult to construct a sequence that passes back through humans in the chain. To construct such a sequence requires the ability to reconstruct, in detail, the cognitive processing of operators during the events that preceded the bad outcome. There are few tools for doing this in any but the most simple laboratory settings. The environment of the large system makes these sorts of reconstructions extremely difficult. Indeed, a major area of research is the development of tools to help investigators trace the cognitive processing of operators as they deal with normal situations, situations at the edges of normality, and system faults and failures. The incidents described in the first part of this chapter are unusual in that substantial detail about what happened and what the participants saw and did was available to researchers. In general, most traces of causality will begin with the outcome and work backward in time until they encounter a human whose actions seem to be, in hindsight, inappropriate or

suboptimal. Because so little is known about how human operators process a multitude of conflicting demands of large, complex systems (e.g., avoid delays in the train schedule but also keep the trains from colliding), incident analyses rarely demonstrate the ways in which the actions of the operator made sense at the time and from their perspective.

The third reason that human error is often the verdict is paradoxical. Human error is the attributed 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 improvement in some underlying quality of the system itself, rather than the honing of the operator skills and expertise to a fine edge. The studies of aircraft carrier flight operations by Rochlin, La Porte, and Roberts (1987) point out that the qualities of human operators are crucial to maintaining system performance goals and that, by most measures, failures should be occurring much more often than they do. As consumers of these systems' products (health care, transportation, defense) society is lulled by success into the belief 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 are many incidents that could become overt failures. Thorough after-accident analyses often indicate that there were numerous incidents or "dress rehearsals" that preceded an accident, as has been reported for the mode error at the heart of the crash of an advanced commercial aircraft at Strasbourg (Woods et al., 1994).

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. Hindsight bias, as Fischhoff (1975) puts it, is the tendency for people to "consistently exaggerate what could have been anticipated in foresight." Studies have shown consistently that 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. After a system failure, knowledge of the outcome biases the reviewer toward attributing failures to system operators. During postevent reviews, knowledge of the outcome will give reviewers the sense that participants ignored presumably obvious or important factors and that the participants therefore erred. Indeed, this effect is present even when those making the judgments have been warned about the phenomenon and been advised to guard against it (Fischhoff, 1975,1982). Fischhoff (1982) wrote:

> 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. (p. 343)

In effect, reviewers 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 trade-offs, the attentional demands and double binds, all may be underemphasized when an incident is viewed in hindsight. In complex, uncertain, highly conflicted settings, such as anesthesia practice and the other similar disciplines such as military situations (Lipshitz, 1989), critics will be unable to disconnect their knowledge of the outcome in order to be able to make unbiased judgments about the performance of human operators during the incident (Baron & Hershey, 1988).

Interestingly, although the phenomenon of *hindsight bias is* well known in psychology, medical practice has had to rediscover it *de novo.* More than a decade after Fischhoff's seminal papers, a study demonstrated the phenomenon in physician judgment. Caplan, Posner, and Cheney (1990) asked two groups of anesthesiologists to evaluate human performance in sets of cases with the same descriptive facts but with the outcomes randomly assigned to be either bad or neutral. The professionals consistently rated the care in cases with bad outcomes as substandard, whereas they viewed the same behaviors with neutral outcomes as being up to standard even though the care (i.e., the preceding human acts) were identical. Typically, hindsight bias in evaluations makes it seem that participants failed to account for information or conditions that "should have been obvious" or behaved in ways that were inconsistent with the (now known to be) significant information. Thus, the judgment of whether or not a human error occurred is 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 after bad outcomes.*

Outcome Failures and Process Defects. It is reasonable to ask if there are means for any evaluation of human performance in complex systems. Indeed, the preceding argument seems a little disingenuous. On the one hand, it is claimed that human performance is critical to the operations of complex systems, and on the other hand, it is argued that there is no scientific way to describe something as a human error and therefore that it is necessarily impossible to distinguish between expert and inexpert performance.

One resolution of this apparent paradox is to distinguish between outcome failures and process defects. Outcome failures are defined in terms of a categorical shift in consequences on some performance dimension. Note that outcome failures are necessarily defined in terms of the language of the domain, for example, sequelae such as neurological deficit, reintubation, myocardial infarction within 48 hours, or an unplanned ICU admission. Process defects are departures from a 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, insufficient intravenous access, insufficient monitoring, regional versus general anesthetic, decisions about canceling a case, or problematic plans or actions with regard to the anesthetic agent of choice. They may also be defined psychologically in terms of deficiencies in some cognitive or information-processing function-for example, activation of knowledge in context, situation awareness, diagnostic search, goal trade-offs.

The distinction between outcome and process is important because the relationship between them is not fixed. Not all process defects are associated with bad outcomes. The defect may be insufficient to create the bad outcome by itself. In addition, as Edwards (1984) said, "a good decision cannot guarantee a good outcome," that is, bad outcomes may result even if there are no defects in process. This is especially true for domains such as anesthesiology where bad outcomes can occur despite the exercise of nearly flawless expertise by the medical personnel involved (cf. Keats,1979,1990).

The rate of process defects may be frequent when compared with 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 (a process of error detection and recovery). Just such a situation can be seen in Incident #2. Evaluating human performance by examining the process of human problem solving in a complex system depends on specifying a start-

---

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

dard about how problems should be handled. There are several categories of standards that can be used to evaluate defects in the process of solving a problem.

One standard is *a normative model of task performance.* This method requires detailed knowledge about precisely how problems should be solved, that is, nearly complete and exhaustive knowledge of the way 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. Unfortunately, normative models rarely exist or are not applicable to complex situations like anesthesia practice. Those models are largely limited to mathematically precise situations such as games or artificial tasks in bounded worlds.

Another standard is the comparison of actual behavior to *standard operating practices (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, the guidelines for management of the difficult airway), 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). These standards may be of limited value because they are either codified in ways that ignore the real nature of the domain[10] or because the coding is too vague to use for evaluation. For example, one senior anesthesiologist, when asked about the policy of the institution regarding the care for emergent Cesarean sections replied, "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., Roth et al., 1987; Suchman, 1987). The set of rules is necessarily incomplete and sometimes

contradictory. It is the role of the human at the sharp end to resolve the apparent contradictions and conflicts in order to satisfy the goals of the system.

In general, procedural rules are too vague to be used for evaluation if they are not specific enough to determine the adequacy of performance before the fact. Thus, a procedural rule such as "the anesthetic shall not begin until the patient has been properly prepared for surgery" is imprecise, whereas another such as "flammable anesthetic agents shall not be used" is specific. When the rules are codified as written policies, imprecise rules usually function simply to provide administrative hierarchies the opportunity to assign blame to operators after accidents and to finesse the larger institutional responsibility for creating the circumstances that lead to accidents (see the report on the aircraft accident at Dryden, in Moshansky, 1992). Significantly, the value of both the normative and standard practices methods of evaluating the problem-solving process of human operators is limited to the most simple systems and generally fails as system size and complexity increase.

A third approach is called the *neutral observer criterion* by De Keyser and Woods (1990). The neutral observer criterion is an empirical approach that compares practitioner behavior during the incident in question to the behavior of similar practitioners at various points in the evolving incident. 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 (in this domain, the patient and related monitors and equipment), 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 hindsight bias; knowledge about the later outcome may alter the neutral observers' judgment about the propriety of earlier responses. The function of the neutral observer is to help define the envelope of appropriate responses given the information available to the practitioner at each point in the incident.

The writers' 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 were taken by individuals to those of other experts placed in the same situation. Note that this is a strong criterion for comparison and it necessarily requires that the evaluators possess 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 pace of system activity is

---

[10] *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 nonfunctional (e.g., Hirschhorn, 1993).

high (i.e., in large, complex systems). The writers have used this technique in examining human performance in incidents from several different sources in anesthesia and in other domains. The technique is complex, as the descriptions and discussions of the three exemplar incidents indicate, but the complexity simply matches that of the domain and the human behaviors being evaluated.

Did the **Practitioners** Commit Errors?

The three exemplar incidents in this chapter are not remarkable or unusual; rather they reflect the normal, day-to-day operations that characterize busy, urban tertiary care hospitals. In each incident, human performance is closely tied to system performance and to eventual outcome, although the performance of the practitioners is not the sole determinant of outcome. The myocardial infarction following the events of Incident #1 may well have happened irrespective of any actions taken by practitioners. That patient was likely to have an infarction, and it is not possible to say whether the anesthetist's actions caused the infarction. The incidents and the analysis of human performance that they prompt (including the role of latent failures in system transients) may make us change our notion of what constitutes a human error.

Arguably, the performance in each exemplar incident is flawed. In retrospect, things can be identified that might have been done differently and that would have forestalled or minimized the incident or its effect. In the myocardial infarction incident, intravascular volume was misassessed, and treatment for several simultaneous problems was poorly coordinated. In the hypotension incident (#2), the device set-up by practitioners probably contributed 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 in order to meet certain demands, but simultaneously exposed the larger system to a rare but important variety of failure. In some sense, each of the 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.

Especially in the hypotension incident (#2), the resiliency of human performance in an evolving incident is demonstrated. The practitioners were willing to abandon their efforts at diagnosis and shift to *a disturbance management* mode of response in order to preserve the patient's life pending resolution 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 an infinite list of shoulds for practitioners in anesthesiology and other large systems, but these lists quickly become unwieldy and, in any case, focus only on the most salient failures from the most recent disaster. It is easy to slip into the "cycle of error" (Fig. 13.5), focusing on error out of context, increasing the complexity of the larger system, exacerbating conflicts, and creating more opportunities for latent failures to accumulate and come together in an accident.

The scheduling incident (#3) is somewhat different. In that incident, it is clear how knowledge of the outcome biases the evaluations of practitioner performance. As Abraham Lincoln said, "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 will make no difference." 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 in order 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?

From this discussion, we are being forced to conclude that the human error is a judgment made in hindsight. 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. As the incidents suggest, human performance is not simply either adequate or inadequate. Neither 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 eliminate the richness of detail that might help to show how the activities of the practitioners were locally rational and miss the bottlenecks and dilemmas that challenge practitioner expertise and skill. The term human error should not represent the concluding point but rather the starting point for studies of accident evolution in large systems.

The schema of knowledge factors, attentional dynamics, and strategic factors provide one means of categorizing the activities of teams of practitioners." The model of large system failure arising from the concatenation of multiple small latent failures provides an explanation for the mysteriously unique appearance of failures. That model also explains the limited success achieved by the pursuit of first causes in the cycle of error. It also suggests that the human practitioner's role in large systems may be in part to uncouple elements of the system in order to minimize the propagation of the latent failures resident in the system (Perrow, 1984).

Together, the exemplar incidents and their analyses imply that many of the changes occurring in medical practice may make the system more brittle and increase the apparent contribution of human error. In response to incidents, organizations generate more rules, regulations, policies, and procedures that make it more likely that medical practitioners will be found to have erred by postincident analyses (Fig. 13.5). Emphasis on cutting costs and increasing efficiency generates more pressure on practitioners, making scenarios like that of the scheduling incident more likely. Increased use of technology such as the computer-based infusion devices in the hypotension incident (#2) raises the complexity of incidents and creates new modes of failure. Even the burgeoning volume of medical knowledge plays a role, making the likelihood of the sort of inert knowledge problems of the myocardial infarction incident more probable (Feltovich 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 that, in turn, become part of the cycle of error and its consequences.

"The practitioners need not be 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.

ENHANCING HUMAN PERFORMANCE

Training

If human performance is critical to system performance, then it seems reasonable to try to enhance it. One method of improving human performance is retraining. Unfortunately, most retraining is predicated on the presence of a human error problem, that is, that flawed human performance is the root cause of system failures and that eliminating this failure mode is the key to success. Under this assumption, many training programs consist merely of routinization of tasks according to a rote method. This approach is sometimes called blame and train, because it begins with the concept that human error is the source of the problem and that this error arises from capriciousness or inattentiveness by practitioners. This was exactly what happened following the Three Mile Island accident in 1979. The regulatory agencies and organizations responded in part with an emphasis in training on rote performance of compiled procedures, and the result was that operational personnel confronted a variety of dilemmas about whether to depart from the standard procedures in more complicated incidents (Woods et al., 1987).

There are several methods in use in the aviation and anesthesia domains that represent contrasting approaches to training. Cockpit resource management (CRM) is a tool used by several major air carriers in an effort to improve crew performance (Wiener, Kanki, & Helmreich, 1993). CRM acknowledges that air crews are a resource for solving problems and attempts to give the crews more experience in working together in crisis situations where coordination is critical. Unlike blame and train methods that seek to regiment human performance to eliminate human error, CRM implicitly views human performance as the critical resource in dealing with novel, threatening situations and focuses on developing in pilots and engineers the ability to work together as a coordinated team. In anesthesia, Gaba's group at Stanford developed a similar tool called Crisis Resource Management (CRM) that provides anesthetists with opportunities to see themselves act under the pressure of simulated disasters (Gaba & DeAnda, 1989). Gaba's group uses material from the aviation CRM as well. Again, the implicit view of this training method is that the human practitioner is a resource and the only likely source of system resilience in the event of an incident. The anesthesia CRM concentrates on infrequently experienced but quite realistic scenarios (e.g., a complete power outage) as a test bed for improving human performance. Both CRMs are qualitatively different from the majority of training approaches generally in use. Both make extensive use of elaborate simulators and large bodies of domain knowledge and can be quite expensive.

Technology

All of the large systems with which we are concerned (anesthesia, nuclear power operations, aviation) are intensely technological, so much so that they do not exist apart from their technology. During the past decade, each of these domains has seen the introduction of automation, the purpose of which is to eliminate human activity as the source of errors. The record of these systems is mixed and controversial (Woods et al., 1994).

Much technological innovation is supposed to reduce human error by reducing human workload. The introduction of microprocessor-controlled infusion pumps, for example, can eliminate the cumbersome process of adjusting a manual valve to regulate drips. However, these same devices create other demands, including set-up, operation, and fault diagnosis as seen in the hypotension incident. Similar equipment in the cockpit and the operating room are actually examples of *clumsy automation* (Cook et al., 1990; Wiener, 1989), where the workload reduction is limited to the times when the operator is not busy at any rate (e.g., mid-flight), and the cost is a substantial increase in workload at peak workload times (e.g., takeoff and landing). Such systems are poor amplifiers of human performance and are likely to degrade performance at critical times." Clumsy automation also includes technologies that provide great increases in precision but demand equally precise operation, such as the newer generation of drug infusion pumps. One such device has a library of drug concentration data and is able, given suitable input parameters, to compute and deliver precisely metered doses of potent drugs based on patient weight. The set-up procedure for this device is more complicated and time consuming than its predecessors and increases the potential for large (i.e., order of magnitude) errors in dosing. Practitioners may also encounter black box systems whose behavior is extraordinarily difficult to understand even when the function performed is relatively simple (Cook, Potter, Woods, & McDonald 1991).

Another use of technology is to eliminate human decision making as a source of error by eliminating the decision making entirely. This trend is most advanced in the commercial aircraft cockpit, but it has also been demonstrated in the operating room and the nuclear plant control center. One effect of attempting to automate decision making can be to increase the intensity of the responsibility-authority double bind. In the larger context, practitioners faced with such devices confront a double problem: Not only do they have to understand the situation confronting them, but they must also understand how the machine sees that situation and be able to evaluate the machine's proposed responses to the situation (Roth et al., 1987; Sarter & Woods, in press; Woods et al., 1994).

"This may be one reason practitioners are sometimes reluctant to embrace such technologies.

Eliminating Human Error Versus Aiding Human Performance

Clearly, those strategies that derive from a desire to minimize human error are different from those that seek to aid human performance. Rules, regulations, sanctions, policies, and procedures are largely predicated on the belief that human error is at the heart of large system failures and that a combination of restrictions and punishments will transform human behavior from error to an error-free state. The same basis exists for some training and technology programs, for example, blame and train and automated decision systems, whereas others (notably CRM) regard human performance as the primary means for dealing with system transients and look for ways to produce more effective human performance. The distinction is an important one and not simply a matter of degree; the choice of path depends critically on the validity of the whole notion of human error.

CONCLUSION

Human operator performance in large systems and the failures of these systems are closely linked. The demands that large, complex systems operations place on human performance are mostly cognitive. The difference between expert and inexpert human performance depends on the timely and appropriate action that in turn is shaped by knowledge factors, attentional dynamics, and strategic factors. A brief examination of a few incidents occurring in anesthesia practice 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 human decision making and activity. The schema of knowledge factors, attentional dynamics, and strategic factors can provide a framework for laying out the issues confronting practitioners at the sharp end.

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 be 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.

This chapter has presented a second approach, one that views human performance as the means for resolving the uncertainties, conflicts, and competing demands inherent in large, complex systems (Hollnagel, 1993). This view acknowledges the presence of both blunt and sharp ends of the system. The blunt end, including regulatory bodies, administrative entities, economic policies, and technology development organizations, can affect sharp-end practitioners by creating and amplifying conflicts and by determining the resources available for resolving those conflicts. The analyses guided by this approach explicitly avoid the term *human error* because it obscures more than it reveals.

Human error is not a distinct category of human performance. After the outcome is clear, any attribution of error is a social and psychological judgment process, not a narrow, purely technical or objective analysis. Outcome knowledge biases retrospective evaluations. Different judges with different background knowledge of the events and context or with different goals will judge the performance of human practitioners differently. Recognizing the limits of the label *human error* can lead us in new, more fruitful directions for improving the performance of complex systems (Woods et al., 1994).

So how should we view a large, complex system failure? If a bad outcome is seen as yet another incident containing one or more human errors by some practitioners, that is, 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 sorts of defective people from becoming practitioners. We can try to police practitioner activities more closely.

This chapter suggests quite a different approach. It proposes that system failures are 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 analysis of the decisions and actions of individuals and groups embedded in the larger system that provides resources and constraints. To study human performance and system failure requires studying the function of the system in which practitioners are embedded. Failures tell us about situations where knowledge is not brought to bear effectively, where the attentional demands are extreme, where the *n-tuple* bind is created. Knowledge of these systemic 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 reconstruct what the view was like or would have been like had we stood on the same road.

The few examples in this chapter give the flavor of what operating at the sharp end really demands of practitioners. It is not surprising that human operators occasionally should be unable to extract good outcomes from the conflicted and contradictory circumstances in which they work. The surprise is that they are able to produce good outcomes as often as they do.

ACKNOWLEDGMENTS

This preparation of this chapter was supported in part by grants from the Anesthesia Patient Safety Foundation and from the National Aeronautics and Space Administration.

REFERENCES

Andrews, J. J. (1990). Inhaled anesthesia delivery systems. In R. D. Miller (Ed.), *Anesthesia* (3rd ed., pp. 171-223). New York: Churchill Livingstone.

Arens, J. F. (1993). A practice parameters overview. *Anesthesiology, 78,* 229-230.

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

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

Bransford, J., Sherwood, R., Vye, N., & Rieser J. (1986). Teaching and problem solving: Research foundations. *American Psychologist, 41,* 1078-1089.

Caplan, R. A., Posner, K. L., Cheney, F. W. (1991). Effect of outcome on physician judgments of appropriateness of care. *Journal of the American Medical Association, 265,* 1957-1960.

Chi, M., Glaser, R., & Farr, M. (1988). *The nature of expertise.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Chopra, V., Bovill, J. G., Spierdijk, J., & 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., & Smalhout, R. (1989). *Human error in the operating room: Identifying cognitive lock up.* (Cognitive Systems Engineering Laboratory Tech. Rep. 89-TR-07). Columbus, OH: The Ohio State University.

Cook, R. I., Potter, S. S., Woods, D. D., & 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., & Howie, M. B. (1990). The natural history of introducing new information technology into a dynamic high-risk environment. *Proceedings of the 34th Annual Human Factors Society Meeting* (pp. 429-433).

Cook, R. I., Woods, D. D., & 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., Woods, D. D., & McDonald, J. S. (1991). *Human performance in anesthesia: A corpus of cases. Report to the Anesthesia Patient Safety Foundation.* (Cognitive Systems Engineering Laboratory Technical Report 91-TR-03). Columbus, OH: The Ohio State University.

Cooper, J. B., Newbower, R. S., & Kitz, R. J. (1984). An analysis of major errors and equipment failures in anesthesia management: Conditions for prevention and detection. *Anesthesiology,* 60, 43-42.

Cooper, J. B., Newbower, R. S., Long, C. D., & McPeek, B. (1978). Preventable anesthesia mishaps: A study of human factors. *Anesthesiology,* 49, 399-406.

DeKeyser, V., & Woods, D. D. (1990). Fixation errors: Failures to revise situation assessment in dynamic and risky systems. In **A. G.** Colombo & A. Saiz de Bustamante (Eds.), *System reliability assessment* (pp. 231-251). Dordercht, The Netherlands: Kluwer.

Dorner, D. (1983). Heuristics and cognition in complex systems. In R. Groner, M. Groner, & W. F. Bischof (Eds.), *Methods of heuristics* (pp. 89-107). Hillsdale, NJ: Lawrence Erlbaum Associates.

Eddy, D. M. (1993a). Clinical decision making: Broadening the responsibilities of practitioners [Letters]. *Journal of the American Medical Association,* 270,708-710.

Eddy, D. M. (1993b). Redefining the role of practitioners: The team approach. *Journal of the American Medical Association,* 269, 1849-1855.

Edwards, W. (1984). How to make good decisions. *Acta Psycholgica, 56,5-27.*

Feltovich, P. J., Spiro, R. J., & Coulson, R. (1989). The nature of conceptual understanding in biomedicine: The deep structure of complex ideas and the development of misconceptions. In D. Evans & V. Patel (Eds.), *Cognitive science in medicine: Biomedical modeling.* Cambridge, MA: MIT Press.

Feltovich, P. J., Spiro, R. J., & Coulson, R. (1993). Learning, teaching and testing for complex conceptual understanding. In N. Fredericksen, R. Mislevy, & I. Bejar (Eds.), Test *theory for a new generation of tests.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Fischhoff, B. (1975). Hindsight does not equal foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance, 1,* 288-299.

Fischhoff, B. (1982). For those condemned to study the past: Heuristics and biases in hindsight. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 335-354). New York: Cambridge University Press.

Gaba, D. M., & DeAnda, A. (1989). The response of anesthesia trainees to simulated critical incidents. *Anesthesia and Analgesia, 68,* 444-451.

Gaba, D. M., Maxwell, M., & DeAnda, A. (1987). Anesthetic mishaps: Breaking the chain of accident evolution. *Anesthesiology,* 66, 670-676.

Gentner, D., & Stevens, A. L. (Eds.). (1983). *Mental models.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Gopher, D. (1991). The skill of attention control: Acquisition and execution of attention strategies. In *Attention and performance XIV* (pp. 299-322) Hillsdale, NJ: Lawrence Erlbaum Associates.

Hirschhorn, L. (1993). Hierarchy versus bureaucracy: The case of a nuclear reactor. In K. H. Roberts (Ed.), *New challenges to understanding organizations,* (pp. 137-149). New York: Macmillan.

Hollister, W. M. (Ed.). (1986). Improved guidance and control automation at the man-machine interface (AGARD Advisory Report No. 228). AGARD-AR-228.

Hollnagel, E. (1993). *Reliability of cognition: Foundations of human reliability analysis.* London: Academic Press.

Johnson, P. E., Moen, J. B., & Thompson, W. B. (1988). Garden path errors in diagnostic reasoning. In L. Bolec & M. J. Coombs (Eds.), *Expert system applications* (pp. 395-428) New York: Springer-Verlag.

Johnson, P. E., & Thompson, **W. B.** (1981). Strolling down the garden path: Detection and recovery from error in expert problem solving. Paper presented at *the Seventh International Joint Conference on Artificial Intelligence.* Vancouver, British Columbia.

Keats, A. S. (1979). Anesthesia mortality-A new mechanism. *Anesthesiology, 68, 2-* **4.**

Keats, A. S. (1990). Anesthesia mortality in perspective. *Anesthesia and Analgesia, 71,* 113-119.

Kemeny, J. G., et al. (1979) *Report of the President's Commission on the accident at Three Mile Island.* New York: Pergamon.

Klein, G. A., & Crandall, B. (in press). The role of mental simulation in problem solving and decision making. In J. Flach, P. Hancock, J. Caird, & K. Vicente (Eds.), *The ecology of human-machine systems.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Klein, G. A., Orasanu, J., Calderwood, R., & Zsambok, C. (Eds.). (1993). *Decision* making in action: *Models and methods.* Norwood, NJ: Ablex.

Layton, C., Smith, P. J., & McCoy, E. (in press). Design of a cooperative problem solving system for enroute flight planning: An empirical evaluation. *Human Factors.*

Leveson, **N. G.,** & Turner, C. S. (1993, July). *An investigation of the Therac-25 accidents. IEEE Computer, 18-41.*

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, May 21). Review of "The evolution of useful things" by Henry Petroski. *Science, 260,* 1166-1168.

Moll van Charante, E., Cook, R. I., Woods, D. D., Yue, Y., & 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 (pp. 263-274). London: Pergamon.

Moshansky, V. P. (1992). *Final report of the Commission of Inquiry into the Air Ontario crash at Dryden, Ontario.* Ottawa: Minister of Supply and Services, Canada.

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). Springfield, VA: National Technical Information Service.

National Transportation Safety Board (1993). *US Air Flight 405 LaGuardia Airport March 22, 1992* (Report No. AAR-93/02). Springfield, VA: National Technical Information Service.

Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology.* San Francisco, CA: Freeman.

Norman, D. A. (1988). *The psychology of everyday things.* New York: Basic Books.

Norman, D. A. (1992). *Turn signals are the facial expressions of automobiles.* New York: Addison-Wesley.

Orasanu, J. (1990). *Shared mental models and crew decision making* (Tech. Rep. No. 46). Princeton NJ: Princeton University, Cognitive Science Laboratory.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 14,534-552.

Payne, J. W., Johnson, E. J., Bettman, J. R., & Coupey, E. (1990). Understanding contingent choice: A computer simulation approach. *IEEE Transactions on Systems, Man, and Cybernetics,* 20,296-309.

Perrow, C. (1984). *Normal accidents. Living with high-risk technologies.* New York: Basic Books.

Perkins, D., & Martin, F. (1986). Fragile knowledge and neglected strategies in novice programmers. In E. Soloway & S. Iyengar (Eds.), *Empirical studies of programmers (Vol. 1,* pp. 213-279). Norwood, NJ: Ablex.

Rasmussen, J. (1986). *Information processing and human-machine interaction: An approach to cognitive engineering.* New York: North-Holland.

Rasmussen, J., Duncan, K., & Leplat, J. (Eds.). (1987). *New technology and human error.* New York: Wiley.

Reason, J. (1990). *Human error.* Cambridge, England: Cambridge University Press.

Roberts, K. H., & 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. I., La Porte, T. R., & Roberts, K. H. (1987, Autumn). The Self-designing high-reliability organization: Aircraft carrier flight operations at sea. *Naval War College Review, pp.* 76-90.

Rochlin, G. I. (1991). Iran Air Flight 655 and the USS Vincennes. In T. LaPorte (Ed.), *Social responses to large technical systems (pp.* 99-126). Dordrecht, Netherlands: Kluwer Academic.

Rogers, W. P., et al. (1986). *Report of the Presidential Commission on the Space Shuttle Challenger Accident.* Washington, DC: U.S. Government Printing Agency.

Roth, E. M., Bennett, K. B., & Woods, D. D. (1987). Human interaction with an "intelligent" machine. *International Journal of Man-MachineStudies,* 27,479-525.

Roth, E. M., Woods, D. D., & Pople, H. E., Jr. (1992). Cognitive simulation as a tool for cognitive task analysis. *Ergonomics,* 35,1163-1198.

Rouse, W. B., & 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., & Woods, D. D. (1991). Situation awareness: A critical but ill-defined phenomenon. *International Journal of Aviation Psychology, 1,* 43-55.

Sarter, N. B., & 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., & 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, 4,1-28.*

Sarter, N. B.,& Woods, D. D. (in press). "How in the world did we get into that mode?" Mode error and awareness in supervisory control. *Human Factors (SpecialIssue on Situation Awareness).*

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., & 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.

Turner, B. A. (1978). *Man-made disasters.* London: Wykeham.

Tversky, A, & Kahneman, 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.* Springfield, VA: National Technical Information Service.

Wagenaar, W., & Keren, G. (1986). Does the expert know? The reliability of predictions and confidence ratings of experts. Decision making in complex systems. In E. Holinagel, G. Mancini, & D. D. Woods (Eds.), *Intelligent decision making in process control environments (pp.* 87-103). Berlin: Springer-Verlag.

Wiener, E. L. (1989). *Human factors of advanced technology ('glass cockpit') transport aircraft* (Tech. Rep. 117528). Moffett Field, CA: NASA Ames Research Center.

Wiener, E., Kanki, B., & Helmreich, R. (Eds.). (1993). Cockpit resource management. New York: Academic Press.

Woods, D. D. (1988). Coping with complexity: The psychology of human behavior in complex systems. In L. P. Goodstein, H. B. Andersen, & S. E. Olsen, (Eds.), *Tasks, errors, and mental models (pp.* 128-148). New York: Taylor & Francis.

Woods, D. D. (1992). *The alarm problem and directed attention.* (Cognitive Systems

Engineering Laboratory Technical Report 92-TR-01). Columbus, OH: The Ohio State University.

Woods, D. D. (1994). Cognitive demands and activities in dynamic fault management: Abductive reasoning and disturbance management. In N. Stanton (Ed.), *The human factors of alarm design (pp.* 63-92). London: Taylor & Francis.

Woods, D. D., O'Brien, J., & 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., & Roth, E. M. (1988). Cognitive systems engineering. In M. Helander (Ed.), *Handbook of human-computer interaction.* New York: Elsevier.

Woods, D. D., Johannesen, L., Cook, R. I., & Sarter, N. (1994). *Behind human error: Cognitive system, computers and hindsight.* Wright-Patterson AFB, OH: Crew Systems Ergonomics Information Analysis Center.

Wright, D., Mackenzie, S. J., Buchan, I., Cairns, C. S., & Price, L. E. (1991, September 14). Critical incidents in the intensive therapy unit. *Lancet, 338,* 676-678.