

**Behind Human Error:
Human Factors Research to Improve Patient Safety**

David Woods

Past President

Human Factors and Ergonomics Society

Introduction

Throughout the patient safety movement, health care leaders have consistently referred to the potential value of the Human Factors research on human performance and system failure (Leape et al., 1998). The patient safety movement has been based on three ideas derived from results of research on human expertise (Feltovich et al., 1997), collaborative work (Rasmussen et al., 1991), and high reliability organizations (Rochlin, 1999) built up through investments by other industries:

- adopt a systems approach to understand how breakdowns can occur and how to support decisions in the increasingly complex worlds of health care,
- move beyond a culture of blame to create open flow of information and learning about vulnerabilities to failure,
- build partnerships across all stakeholders in health care to set aside differences and to make progress on a common overarching goal.

Creating a durable, informative, and useful partnership between health care and those disciplines with core expertise in areas of human performance is critical to advancing patient safety. Through substantive partnerships, we can look ahead to anticipate and block new paths to failure in the changing world of health care.

Parallel Perspectives

“Human error in medicine, and the adverse events which may follow, are problems of psychology and engineering not of medicine.” John Senders, 1993

Health care organizations and the public are wondering how we can reduce injuries to patients that occur in the process of care. From one perspective, preventing adverse events such as medication misadministrations, delayed diagnoses, or wrong site surgeries is concerned with

specific medical issues in specific health care settings. From another perspective adverse events concern different variables that affect human performance in lawful, predictable ways.

The research base about human performance has been built up from studies of how practitioners interact to handle situations in many different contexts such as aviation, industrial process control, space operations. The results capture empirical regularities and provide explanatory concepts and models. These results allow us to go behind the unique features and surface variability of many different settings to see common underlying patterns.

To understand and predict human performance in health care, as in any complex setting, we need to make use of the different languages that describe human performance. To understand patterns in human judgment one needs to understand concepts such as bounded rationality, knowledge calibration, heuristics, oversimplification fallacies (Feltovich et al., 1997). To understand patterns in communication and cooperative work one needs to understand concepts such as supervisory control, common ground, and open versus closed work spaces (Clark and Brennan, 1991; Galegher et al., 1990; Greenbaum and Kyng, 1991; Rasmussen et al., 1991). To understand patterns in human-computer cooperation, one needs to understand concepts such as the representation effect, object displays, inattention blindness, mental models, data overload, mode error (Norman, 1993; LaBerge, 1995; Rensink et al., 1997; Zhang, 1997; Vicente, 1999). These are concepts foreign to health care practitioners just as medical concepts in cardiology or respiratory therapy are foreign to the human factors community.

The same pattern in communication and cooperative work--the effectiveness of cross-checking--plays out in many different health care settings. The same health care issue, medication misadministration, involves very different kinds of human performance issues depending on the context (e.g., internet pharmacies, patient self-managed treatment, administration through computerized infusion devices, or interactions through computer based communication in an order entry system). When we understand the kinds of processes in human performance that play out in the health care setting or situation of interest, we can use the Human Factors knowledge base to guide the development and testing of new ways to support very high levels of human performance in that health care setting.

This is a process of going back and forth between two different perspectives in a partnership where each party steps outside of their own area of expertise to learn by looking at the situation

through the other's perspective. This means research on patient safety is a kind of interdisciplinary synthesis that requires combining technical knowledge in specific health care areas, technical knowledge of general results/concepts about the various aspects of human performance that play out in that area, and a view of practitioners carrying out this kind of work in context.

Human Factors researchers and professionals have been and continue to be eager to join in such partnerships to make progress on patient safety. We are willing to invest the energy to learn from medical colleagues about the pragmatics and technical knowledge in diverse health care settings. Similarly, we hope that our medical colleagues are ready to invest energy in learning about the technical knowledge that captures regularities in the various aspects of human performance that can play. Indeed, partnerships of this kind have been and continue to be the engine of progress in work on patient safety to date (see Bogner, 1994; see programs such as the Annenberg meetings, Hendee, 1999; and in partnerships at various research labs around the US and the world, e.g., Nyssen and De Keyser, 1998; Howard et al., 1992; Howard et al., 1997; Cook and Woods, 1996).

Startling Results from the Science

One of the great values of science is that during the process of discovery conventional beliefs are questioned by putting them in empirical jeopardy. When scientists formulate new ideas and look at the world anew through these conceptual looking glasses the results often startle us. As a result we can innovate new approaches to accomplish our goals.

This process has been going on for over 20 years in the "New Look" at the factors behind the label human error (Reason, 1997; Woods et al., 1994; Rasmussen, 1990; 1999. Driven by surprising failures in different industries, researchers from different disciplinary backgrounds began to re-examine how systems failed and how people in their various roles contributed to both success and failure. The results often deviated from conventional assumptions in startling ways.

The research found that doing things safely, in the course of meeting other goals, is always part of operational practice. As people in their different roles are aware of potential paths to failure, they develop failure sensitive strategies to forestall these possibilities. When failures occurred against this background of usual success, researchers found multiple contributors each necessary but only jointly sufficient and a process of drift toward failure as planned defenses eroded in the face of

production pressures and change. The research revealed systematic, predictable organizational factors at work, not simply erratic individuals. The research also showed that to understand episodes of failure one had to first understand usual success--how people in their various roles learn and adapt to create safety in a world fraught with hazards, tradeoffs, and multiple goals (Cook et al., 2000).

Researchers have studied organizations that manage potentially hazardous technical operations remarkably successfully, and the empirical results also have been quite surprising (Rochlin, 1999). Success was not related to how these organizations avoided risks or reduced errors, but rather that these high reliability organizations created safety by anticipating and planning for unexpected events and future surprises. These organizations did not take past success as a reason for confidence. Instead they continued to invest in anticipating the changing potential for failure because of the deeply held understanding that their knowledge base was fragile in the face of the hazards inherent in their work and the changes omnipresent in their environment. Safety for these organizations was not a commodity, but a value that required continuing reinforcement and investment. The learning activities at the heart of this process depended on open flow of information about the changing face of the potential for failure. "High reliability" organizations valued such information flow, used multiple methods to generate this information, and then used this information to guide constructive changes without waiting for accidents to occur.

Perhaps most startling in this work is the result that the source of failure was not those others who are less careful or motivated than we are. Instead, the process of investing in safety begins with each of us being willing to question our beliefs to learn surprising things about how *we* can contribute to the potential for failure in a changing and limited resource world.

Systems Issues For Research to Improve Safety

Search for Underlying Patterns to Gain Leverage

From past work, progress has come from going beyond the surface descriptions (the phenotypes of failures) to discover underlying patterns of systemic factors (genotypical patterns). Genotypes are patterns about how people, teams, and organizations coordinate activities, information, and problem solving to cope with the complexities of problems that arise (Hollnagel, 1993).

The surface characteristics of a near miss or adverse event are unique to a particular setting and people. Genotypical patterns re-appear in many specific situations. Research in Human Factors has revealed a wealth of genotypical patterns, for example,

- Garden path problems and the potential to fixate on one point of view or hypothesis in problem solving (De Keyser and Woods, 1990).
- Missing side effects of an action or change to a plan in highly coupled systems (Rasmussen, 1986).
- Hindsight bias from knowledge of outcome.
- Alarm overload and high false alarm rates leading to missed or ignored warnings (Stanton, 1991).
- Mode errors in computerized devices with multiple modes and poor feedback about device state (Norman, 1988).

A great deal of leverage for improvements is gained by identifying the genotypical patterns at work in a particular situation of interest. For example, we can sample the kinds of difficult situations that can occur in a health care setting and recognize the presence of garden path problems (e.g., in anesthetic management; Gaba, 1987). We may review a corpus of near misses and note that in several cases a practitioner became fixated on one view of the situation (Cook et al., 1989). Or we may analyze how people handle simulated problems and see the potential for fixating in certain situations (e.g., as has occurred in crisis training via anesthesia simulators; Howard et al., 1992).

Previous work on aiding human and team situation assessment can now seed and guide the development of interventions. To overcome fixation in a garden path problem, one can bring to bear techniques that may break up frozen mindsets (such as new kinds of pattern-based displays or new team structures that help broaden the issues under consideration; De Keyser and Woods, 1990).

Each of the genotypes listed above was identified and studied in aerospace and process control settings, but they all also play out in multiple health care settings:

- Fixation as a danger in anesthetic management during surgery.
- Missed warnings due to high nuisance and false alarm rates in intensive care units.
- Mode errors in computerized infusion devices.
- Hindsight bias in incident review teams.

This list is short and only exemplifies some of the results available to jump start research in health care settings.

Research on patient safety should be using and expanding the set of genotypical patterns related to breakdowns in human performance that occur on health care settings. Research should focus on developing and testing interventions to reduce these problems. In many cases previous work has identified the interventions needed (e.g., in the case of mode errors). In other cases, seed ideas exist which to be further developed given the unique pressures of health care.

Tame Complexity

In the final analysis, the enemy of safety is complexity. In nuclear power and aviation, we have learned at great cost that often it is the underlying complexity of operations that contributes to the human performance problems. Simplifying the operation of the system can do wonders to improve its reliability, by making it possible for the humans in the system to operate effectively. Often, we have found that proposals to improve systems founder when they increase the complexity of practice. Adding new complexity to already complex systems rarely helps and can often make things worse. This applies to system improvements justified on grounds of safety as well.

The search for simplicity of operation has a severe catch however. The very nature of health care delivery includes, creates, or exacerbates many forms of complexity. Success and progress occur through monitoring, managing, taming, and coping with the changing forms of complexity.

This has proven true particularly with respect to efforts to introduce new forms and levels of computerization. Improper computerization can simply exacerbate or create new forms of complexity to plague operations. The situation is complicated by the fact the new technology often has benefits at the same time that it creates new vulnerabilities.

Again the science startles us. To help people in their various roles create safety, research needs to

- Search out the sources of complexity
- Understand the strategies people, teams and organizations use to cope with complexity.
- Devise better ways to help people cope with complexity to achieve success.

This one of the most basic lessons that has come out of the New Look research about error.

Adopt Methods for User Centered Design of Information Technology

When Human Factors practitioners and researchers examine the typical human interface of computer information systems and computerized devices in health care, they are often shocked. What we take for granted as the least common denominator in user centered design and testing of computer systems in other high risk industries (and even in commercial software development houses that produce desktop educational and games software) seems to be far too rare in medical devices and computer systems. The devices are too complex and require too much training to use given typical workload pressures (e.g., Cook et al., 1992; Obradovich and Woods, 1996; Lin et al., 1998).

Computer displays, interfaces, and devices in health care exhibit “classic” human-computer interaction deficiencies. By “classic” we mean that we see these design “errors” in many devices in many settings of use, that these design problems are well understood (e.g., they appear in our textbooks and popular writings, e.g., Norman and Draper, 1986; Norman, 1988), and that the means to avoid these problems are readily available.

We are concerned that the calls for more use of integrated computerized information systems to reduce error could introduce new and predictable forms of error unless there is a significant investment in use-centered design (Winograd and Woods, 1998).

The concepts and methods for use centered design are available and are being used everyday in software houses (Nielsen, 1993; Carroll and Rosson, 1992; Flach and Dominguez, 1995).

Usability testing should be a standard, not an exceptional part of product development practices. Health care delivery organizations also need to understand how they can use these techniques in their own testing processes and as informed consumers of computer information systems.

Building the partnerships, creating demonstration projects, and disseminating the techniques for health care organizations is a significant and rewarding investment to ensure we receive the benefits of computer technology while avoiding designs that induce new errors.

But there is much more to human-computer interaction than adopting basic techniques like usability testing. Much of the work in Human Factors concerns how to use the potential of computers to enhance expertise and performance. We will consider only a few of these issues here.

Study Human Expertise to Develop the Basis for Computerization

The key to skillful as opposed to clumsy use of technological possibilities lies in understanding both the factors that lead to expert performance and the factors that challenge expert performance (Feltovich, Ford, and Hoffman, 1997). Once we understand the factors that contribute to expertise and to breakdown, we then will understand how to use the powers of the computer to enhance expertise. This is an example of a more general rule—to understand failure and success, begin by understanding what makes some problems difficult.

The areas of research on human performance in medicine explored in the monograph *A Tale of Two Stories* (Cook, Woods and Miller, 1998) illustrate this process. In these cases, progress depended on investigations that (a) identified the factors that made certain situations more difficult to handle and then (b) explored the individual and team strategies used to handle these situations. As the researchers began to understand what made certain kinds of problems difficult, how expert strategies were tailored to these demands, and how other strategies were poor or brittle, new concepts were identified to support and broaden the application of successful strategies. In each of these cases, the introduction of new technology helped create new dilemmas and difficult judgments. On the other hand, once the basis for human expertise and the threats to that expertise had been studied, new technology was an important means to achieve enhanced performance.

We can achieve substantial gains by understanding the factors that lead to expert performance and the factors that challenge expert performance. This provides the basis to change the system, for example, through new computer support systems and other ways to enhance expertise in practice.

Make Machine Advisors and Automation Team Players

New levels of automation have had many effects in operational settings. There have been positive effects from both an economic and a safety point of view. Unfortunately, operational experience, research investigations, incidents, and occasionally accidents have shown that new and surprising problems have arisen as well. Computer agents can be brittle and only able to handle a portion of the situations that could arise. Breakdowns in the interaction between operators and computer-based automated systems can also contribute to near misses and failures in these complex work environments.

Over the years, Human Factors investigators have studied many of the “natural experiments” in human-automation cooperation – observing the consequences in cases where an organization or industry shifted levels and kinds of automation. One notable example has been the many studies of the surprising consequences of new levels and types of automation on the flight deck in commercial transport aircraft (Billings, 1996).

The overarching result from the research is that for automation concerned with information processing and decision making to be successful, the key requirement is to design for fluent, coordinated interaction between the human and machine elements of the system. In other words, automated and intelligent systems must be designed to be “team players” (Malin et al., 1991; Roth et al., 1997). When automated systems increase autonomy or authority of machines without new tools to support cooperation with people, we find automation surprises contributing to incidents and accidents (Sarter, Woods and Billings, 1997).

Human Factors research has abstracted many patterns and lessons about how to make automated systems team players. One example is that increased automation requires new forms of feedback and display to show human users what the automated agents are doing and what they will do next relative to the state of the process (Norman, 1990). People can be more willing to accept even *poor* advice when it comes from a computer (e.g., Layton et al., 1994). More successful designs reverse the relationship--instead of having people check the computer, critiquing software can be used relatively unobtrusively to remind, suggest, and broaden the factors considered by the human decision maker and improve performance, even for cases where the computer is unable to generate a good solution on its own (Guerlain et al., 1999).

Many of the developments in computer information systems across health care delivery systems (be they intended to enhance safety or productivity) include embedded forms of automation. Using the lessons from past research to guide the design of automated information processing systems will help avoid new paths to failure and increase the benefits to be obtained from these investments in new technology.

Invest in Collaborative Technologies

“When I order a medication, I think the patient gets the medication directly, but there are many other steps, computer systems, and hands that intervene in the process.”

A Physician, 2000

As health care delivery advances and also tries to become more efficient, it has become more distributed over multiple practitioners, groups, locations, and organizations. This change challenges us to coordinate care over these disparate players.

We often think that if these different players are connected through new information technology, effective coordination will follow automatically. The exploding field of computer supported cooperative work or CSCW tells us that achieving high levels of coordination is a special form of expertise requiring significant investment in experience or practice (Grudin, 1994; Galegher et al., 1990). It also tells us that making cooperative work through a computer effective is a difficult challenge. For example, a critical part of effective collaboration is how it helps broaden deliberations and cross-checks judgments and actions to detect and recover from incipient failures. The design of the communication channel and information exchanged can degrade the cross-check function or enhance it, depending on how the channel is designed. A great deal of work is underway to try to identify what factors are important to support good collaborative work to guide the investment in technology (e.g., Nyssen and Javaux, 1996; Heath and Luff, 1992; Jones, 1995).

The field of CSCW is exploding because of the advances in the technology of connectivity, the Internet and telecommunications in general, and because of the potential benefits of being connected. This leads designers to need information about the basis for high levels of skill at coordinated work.

The new technologies of connectivity will transform the face of health care practices, and CSCW should become a core area of expertise, research and development in health care. Relationships between practitioners will change and new relationships will be introduced. The changes can be designed primarily to support efficiency, or they can be designed primarily to enhance safety. Side effects of these coming changes can also create new paths to failure while they block other paths. Research to understand and direct this wave of change to enhance safety will need to be a critical research priority in health care as it is in other high performance fields.

The development and impact of tele-medicine is one example of this process. Achieving continuity of care in the new world of computer connectivity across health care practitioners and providers is another.

Manage the Side Effects of Change

Health care systems exist in a changing world. The environment, organization, economics, capabilities, technology, and regulatory context all change over time. Even the current window of opportunity to improve patient safety is a mechanism for change. And new waves of change are beginning to swell up and move into play (e.g., tele-medicine). Waves of change are due in part to resource pressures and in part to new capabilities. Uncertainties created by these changes have given rise to the public pressure for improving patient safety.

This backdrop of continuous systemic change ensures that hazards and how they are managed are constantly changing. This is important because many of these changes can easily increase complexities of health care delivery. Again, increasing complexity often challenges safety and rarely produces safety benefits without some other investments.

The general lesson is that as capabilities, tools, organizations and economics change, vulnerabilities to failure change as well--some decay but new forms appear. The state of safety in any system is always dynamic, and stakeholder beliefs about safety and hazard also change. Progress on safety depends on anticipating how these kinds of changes will create new vulnerabilities and paths to failure even as they provide benefits on other scores.

For example, new computerization is often seen as a solution to human performance problems. Instead, consider potential new computerization as another source of change. Examine how this change will affect roles, judgments, coordination, what makes problems difficult. This information will help reveal side effects of the change that could create new systemic vulnerabilities.

Armed with this knowledge we can address these new vulnerabilities at a time when intervention is less difficult and less expensive (because the system is already in the process of change). In addition, these points of change are opportunities to learn how the system actually functions and sometimes mal-functions.

Another reason to study change is that health care systems are under severe resource and performance pressures from stakeholders. First, change under these circumstances tends to increase coupling, that is, the interconnections between parts and activities, in order to achieve

greater efficiency and productivity. However, research has found that increasing coupling also increases operational complexity and increases the difficulty of the problems practitioners can face. Second, when change is undertaken to improve systems under pressure, the benefits of change may be consumed in the form of increased productivity and efficiency and not in the form of a more resilient, robust and therefore safer system.

Thus, one linchpin of future success on safety is the ability to anticipate and assess the impact of change to forestall new paths to failure.

In addition, investments in safety are best timed to coincide with windows of opportunity where change is happening for other reasons as well. A great deal of leverage may result from projects designed to show health care organizations how to take advantage of change points as windows of opportunity where they can re-think processes, work flow, and new modes of collaboration to reduce the potential for breakdowns.

The System Of Health Care Delivery Is Changing To Include Patients In New Ways.

Patients are becoming involved in their own care (or their family members's care) in new ways. Patients have new access to information so that they can take an active role in treatment decisions. Technology and other factors are shifting care from in patient settings to home settings where patients become a provider of their own treatment (e.g., Obradovich and Woods, 1996; Lin et al., 1998).

Patient self-managed treatment represents a large and growing part of the health care system. Human factors can support patient safety in self-managed treatment. Errors can occur when information and interfaces do not fit the patient's capacities, past experiences or the demands of daily life. Human factors can focus attention on underlying mechanisms, behavioral patterns, and contextual contributions and provide the methods and tools to design the support systems that match patient needs.

Studying Human Performance, Human-Machine Systems, Collaboration, and Organizational Dynamics Requires Methods Unfamiliar to Health Care

Ultimately, the study of human performance is in one sense or another the study of problem solving. Since its origins over one hundred years ago, understanding problem solving, be it a person, human-machine system, distributed teams, or organizations, has always been the study of the processes that lead

up to outcomes--what is seen as a problem-to-be-solved, how to search for relevant data, anticipating future events, generating hypotheses, evaluating candidate hypotheses, modifying plans to handle disruptions.

The base data is the process or the *story* of the particular episode – how multiple factors came together to produce that outcome (Klein et al., 1995; Klein, 1998; Dekker, in press). Patterns abstracted from these processes are aggregated, compared, and contrasted under different conditions—different problem demands (scenarios), different human-human and human-machine teams, different levels of expertise, different external tools.

The fields that study one or another type of problem solving have developed sophisticated methods tailored to meet the uncertainties of studying and modeling these processes. These are deeply foreign to medical research communities, but they are the lifeblood of coming to understand human performance in any complex setting including health care. Ethnography (Hutchins, 1995), interaction analysis (Jordan and Henderson, 1995), protocol analysis (Ericsson and Simon, 1984), critical incident techniques (Flanagan, 1954; Klein, 1998), work analysis (Vicente, 1999) are just a few of the techniques to be mastered by the student of human problem solving at work.

One critical resource for the study of problem solving is mechanisms to build or obtain access to simulation environments at different scopes and degrees of fidelity. Much of the progress in aviation safety has depended on researchers access to full scope training simulators to study issues such as effective human-human and human-machine cooperation (e.g., Layton et al., 1994). This has occurred through research simulators at NASA Ames and Langley Research Centers, through partnerships with pilot training centers (when research and training goals can be synchronized), and through the use of rapid prototyping tools to create simulation test beds. We have already begun to see how the availability of simulator resources in health care (notably for anesthesia and the operating room) can be a catalyst to learning about the factors that lead to effective or ineffective human performance (Howard et al., 1992; Nyssen, and De Keyser, 1998; Guerlain et al., 1999).

Using these resources to understand human performance depends on special skills such as problem or scenario design (the design of the problems study participants attempt to solve) and in analysis techniques such as interaction and protocol analysis (De Keyser and Samercay, 1998). Health care research organizations will need to create, modify, and use simulation resources to

provide critical pieces of evidence in the process of finding effective ways to improve safety for patients.

Technology Evaluation: Beyond Verification and Validation (V&V)

Evaluating changes intended to improve some aspect of human performance is a difficult problem. Human Factors has worked with many industries to assess the impact of technology and other interventions (e.g., training systems) designed to aid human performance. Stakeholders have frequently asked us to give them a simple up or down result—does this particular system or technology help significantly or not? We refer to such studies as verification and validation evaluations or V&V.

Despite the surface appeal of such efforts and the desire to provide definitive answers to guide investments, V&V has proved to be a limited tool in other high risk domains. The short summary of the lessons is that such studies provide too little information, too late in the design process, at too great a cost.

They provide too little information in a variety of ways. There are multiple degrees of freedom in using new technology to design systems, but V&V studies are not able to tell developers how to use those degrees of freedom to create useful and usable systems. The problem in design today is not can we build it, but rather what would be useful to build given the wide array of possibilities new technology provides.

Measurement problems loom large because V&V studies usually try to capture overall outcomes. However, the systems are intended to influence aspects of the processes (human expertise, cooperative work, a culture of safety) that are important to outcomes in particular kinds of situations that could arise (Woods et al., 1995). As a result, global outcome measures tend to be insensitive to the operative factors in the processes of interest or wash out differences that are significant in restricted kinds of situations.

New systems and technology are not unidimensional, but multi-faceted, so that problems of credit assignment become overwhelming. Introducing new technology is not manipulating a single variable, but a change that reverberates throughout a system transforming judgments, roles,

relationships, and weightings on different goals (Carroll and Rosson, 1992).¹ This process, called the task-artifact cycle, creates the envisioned world problem for research and design (Dekker and Woods, 1999; Hoffman and Woods, 2000): how do the results of studies and analyses that characterize cognitive and cooperative activities in the current field of practice inform or apply to the design process, since the introduction of new technology will transform the nature of practice, what it means to be an expert, and the paths to failure? Health care specialists need only consider the case of laparoscopic surgery to see these processes play out (e.g., Dominguez et al., in press; Cook et al., 1998).

V&V studies occur too late in the design process, especially given their great cost, to provide useful input. By the time that the V&V results are available, the design process has committed to certain design concepts and implementation directions. These sunk costs make it extremely difficult to act on what is learned from evaluation studies late in the process.

The advent of rapid prototyping technology has revolutionized evaluation studies. While late V&V studies still have a role, the emphasis has shifted completely in many different work domains to early, generative techniques such as ethnography, envisioning techniques, and participatory design (Greenbaum and Kyng, 1991; Carroll and Rosson, 1992; Smith et al., 1998; Sanders, 2000). Health care needs to build on this experience and learn the use of these new techniques.

High Reliability Organizations and Reactions to Failure

One of the most productive areas of work on human error in the last 10 years has been about reactions to failure (Woods et al., 1994; Rochlin, 1999). In this work we have come to understand more about what characterizes high reliability organizations (Rochlin et al., 1987; Weick and Roberts, 1993; Westrum, 1993; Grabowski and Roberts, 1997) and about the common oversimplifications or fallacies about error that block forward progress (Cook et al., 1998).

High reliability organizations create safety by anticipating and planning for unexpected events and future surprises. These organizations continue to invest in anticipating the changing potential

¹ This is one way to view Human Factors as a field—the body of work that describes how technology and organizational change transforms work in systems.

for failure, regardless of past success, because they appreciate that their knowledge is imperfect and that their environment continues to change. The heart of this process is learning activities which depend on open flow of information about the changing threats and about the changing effectiveness of their failure sensitive strategies. For these organizations, safety is a value, not a commodity.

Escape from Hindsight Bias

There are a variety of factors that block or inhibit the learning processes central to a high reliability culture. One is the hindsight bias (Fischhoff, 1975; Woods et al., 1994; Woods and Cook, 1999). The hindsight bias is one of the most reproduced research findings relevant to accident analysis and reactions to failure. Knowledge of outcome biases our judgment about the processes that led up to that outcome.

In the typical study, two groups of judges are asked to evaluate the performance of an individual or team. Both groups are shown the same behavior; the only difference is that one group of judges are told the episode ended in a poor outcome; while other groups of judges are told that the outcome was successful or neutral. Judges in the group told of the negative outcome consistently assess the performance of humans in the story as being flawed in contrast with the group told that the outcome was successful. Surprisingly, this hindsight bias is present even if the judges are told beforehand that the outcome knowledge may influence their judgement.

Hindsight is not foresight. After an accident, we know all of the critical information and knowledge needed to understand what happened. But that knowledge is not available to the participants before the fact. In looking back we tend to oversimplify the situation the actual practitioners faced, and this tends to block our ability to see the deeper story behind the label human error.

Researchers use methods designed to remove hindsight bias to see the multiple factors and contributors to incidents, to see how people usually make safety in the face of hazard, and to see systemic vulnerabilities before they contribute to failures. Research has developed a variety of techniques to reduce hindsight bias. These are available to use and modify as necessary in health care research on safety.

Despite widespread communication about this bias, it continues to plague the literature on medical error. The studies of injury or death rates as a result of error and virtually all incident review procedures used in health care today fail to control for hindsight bias. This should not be considered acceptable by any one interested in improving safety. It is time to stop repeating this “error” in the study of error.

Innovate New Models of Accountability

As the new messages about a systems approach to safety circulated and became more visible in health care, they collided with the common belief that practitioners and managers should be “accountable” to patients and to other stakeholders. But accountability in this common belief is operationalized in terms of pursuit of culprits, threats of disciplinary actions, and threats of stigmatization.

These two trends collide in a basic double bind for the patient safety movement: blame, even if disguised as accountability, drives out information about systemic vulnerabilities, stops learning, and undermines the potential for improvement (Billings, 2000). The challenge for research is to find a way out of the double bind—How do we create a safe environment for learning about the potential for failure in a publicly accountable system of health care delivery?

Accountability is emphasized in the debate on patient safety because how decision makers are held accountable is presumed to influence how they make decisions and the quality of those decisions. Social links such as accountability can indeed be powerful forces influencing human decision making, and these relationships have been studied in organizational dynamics, social cognition, and human-machine interaction (e.g., Hirschhorn, 1993; Lerner and Tetlock, 1999; Tetlock, 2000) This information can be integrated with ethical and legal scholarship to stimulate the innovation of new systems for managing accountability relationships (Sharpe, 2000; Palmer et al., 2000).

Practitioner decision making always occurs in a context of expectations that one may be called to give accounts for those decisions to different parties. How and to whom people expect to be called to account affects their performance in implicit and explicit ways. The expectations for what are considered adequate accounts and the consequences for people when their accounts are judged inadequate are critical parts of the cycle of giving accounts and being called to account.

Interestingly, different factors in this reciprocating cycle can support or undermine practitioner performance and systems learning in predictable ways. Note that from a behavioral science point of view, accountability is a neutral term that only points to the processes in this cycle of giving and being called to give accounts or reasons for a decision.

First, past research shows that there is a complex set of factors, relationships and effects at work in the reciprocating cycle of calling on and giving of accounts. Second, the empirical regularities and relationships are not consistent with motivational accounts, i.e., that accountability creates general improvements by increasing task motivation. Third, and most startling, the research demonstrates that some factors in the reciprocating cycles of accountability may degrade decisions, performance, cooperation and learning, while other relationships in the cycle may enhance these cognitive processes.

For example, under some conditions the need to give an account for a decision to others can increase critical thinking and attenuate commitment (presumably ways to enhance the decision), while other conditions can increase self justification, bolster an initial attitude and commitment (presumably ways that reduce the quality of a decision). Some forms of accountability can increase defensive behavior, create adversarial relationships among parties who need to cooperate, or lead people to prefer options that are easier to justify given knowledge of the standards others impose for giving of suitable accounts.

These results allow us to see the label “culture of blame” in a new way. It is a kind of system of accountability, but only one way to design and manage such systems. If one analyzes a “culture of blame” in terms of the dynamics of cycles of accountability, we find many of the factors that have been implicated in degrading performance, cooperation, and learning.

The social cognition research on cycles of accountability clearly captures the complexity of the effects and demonstrates the naïveté of the belief that improving safety only requires holding others accountable. The slogan of “moving beyond a culture of blame” is a call to abandon poor systems of accountability, not an absence of accountability. The systems approach does not mean abandoning accountability. It is a necessary part of our life as social creatures that we explain our actions to others. The systems approach examines the reciprocating cycle of giving accounts and calling to accounts between the sharp end of practice and the blunt end of organizational context to determine the lawful effects of different systems for accountability.

New work is needed to model and describe systems of accountability and their effects in health care. The results from this work should be used to engage all stakeholders in a process to explore new designs of systems of accountability that will produce the desired effects and advance our common goals.

Create and Share Learning Tools

The research on high reliability organizations emphasizes continued learning about risks and mitigation strategies. But in observations of health care, direct learning and improvement from experience with accidents and incidents has proven to be very limited and narrow. This appears to be partly because of the fear of blame and litigation. In addition, there are few organizational structures that promote learning about paths to failure.

An important area for new work is creating learning tools that function throughout health organizations. Some believe that expanded requirements for notification of adverse events and near misses will accomplish this. But lessons from aviation indicate that much more is needed in the analysis of cases and in feedback mechanisms to health care practitioners to complete a cycle of learning based on incident reporting (Billings, 1999). New incident reporting systems will need to encompass much more than new forms to fill out and new notification requirements to be effective. Serious questions remain about whether these programs will be effective including: the independence of the review teams, the need to move beyond looking at cases one at a time to analyzing sets of cases, how to generate meaningful sustained improvements, and how to provide feedback to practitioners about how what was learned is relevant to them in their portion of the world of health care.

One mechanism to enhance learning about risks of adverse events is the creation of regional centers like the Veterans Health Administration's Centers for Inquiry to collaborate with multiple health care systems. These centers can be incubators to develop new techniques for learning and sharing the lessons from cases including role play simulations, walkthroughs of cases, and other techniques. These centers can provide the infrastructure to build partnerships between human performance and health care experts to promote the information flow and learning processes central to a culture of safety.

Interdisciplinary Partnerships

The success of NASA programs in aviation safety were built on interdisciplinary partnerships. The success of safety initiatives in health care are very likely to depend on building the same kinds of partnerships across quite different technical disciplines. Considering patient safety inevitably leads to technical issues about human performance, human-computer cooperation, organizational dynamics, software engineering and other fields normally considered outside health care. The new research efforts will need to be structured to build up these partnerships between technical areas concerned with different aspects of human performance and different medical and practitioner specialties. One important activity will be to develop a new cadre of experts who are skilled at these interdisciplinary projects (for example, programs that invest in supporting dissertation research).

The success of NASA programs was based on a portfolio of research and development activities that included more basic work on human performance (e.g., mental workload), innovation of new directions for aiding human performance (e.g., cockpit resource management), advanced development, and technology transfer projects. In this process different kinds of work were carried out and sponsored including field research, simulator studies, concept development, evaluation studies. The infrastructure at NASA helped cross-stimulate all of these activities around the goal of developing new design directions that would be useful in improving aviation safety. The structure at NASA and NASA's role in the aviation industry provided another essential ingredient for success—independence. The issues underlying safety are potentially controversial. A technically grounded, independent organization whose only purpose is advancing safety can develop a reservoir of technical results and organizational confidence to ride through these controversies with their substantive efforts for safety intact.

Health care would do well to study the formal and informal organizational basis of the successes of the NASA R&D program and to model their own research efforts on patient safety on NASA's Human Factors programs.

Conclusion

We have a window of opportunity for improving safety for patients, but there are many false trails that will consume the energy and resources available. To take advantage of this window we must be prepared to question conventional wisdom and assumptions by building a partnership between different health care specialties and different human performance specialties that

intersect at the label human error. These human performance specialties are substantive, deep, and unfamiliar to health care. I have not been able to cover all of them in this piece. But they are the wellspring for techniques, concepts, and systems that will improve human performance in health care as has been the case in other high risk domains.

From the past work in Human Factors a simple standard emerges for judging success in research on error and safety. Research is successful to the degree that it helps recognize, anticipate, and defend against paths to failure that arise as organizations and technology change, **before any patient is injured.**

References

- Barley, S. and Orr, J., Eds. (1997). *Between Craft and Science: Technical Work in US Settings*,
- Billings, C. E. (1996). *Aviation automation: The search for a human-centered approach*. Mahwah, NJ: Erlbaum.
- Billings, C. E. (1999). The NASA Aviation Safety Reporting System: Lessons Learned from Voluntary Incident Reporting. In Proceedings of *Enhancing Patient Safety and Reducing Errors in Health Care*. National Patient Safety Foundation, Chicago IL (held at Annenberg Center for Health Sciences, Rancho Mirage CA Nov. 8–10, 1998).
- Billings, C. E. (2000). Human Error in Perspective.
- Bogner, M.S., Ed. (1994). *Human Error in Medicine*, Mahwah, NJ: Erlbaum.
- Carroll, J.M. and Rosson, M. B. (1992). Getting around the task-artifact cycle: How to make claims and design by scenario. *ACM Transactions on Information Systems*. 10, 181-212.
- Clark, H. H. and Brennan, S. E. (1991). Grounding in communication. In L. Resnick, J. Levine, and S. Teasley. (Eds.) *Socially Shared Cognition*. Washington DC: American Psychological Association.
- Cook, R. I., McDonald, J. S., and Smallhout, R. (1989). *Human error in the operating room: Identifying cognitive lock up*. Cognitive Systems Engineering Laboratory Technical Report 89-TR-07, Columbus, OH: Department of Industrial and Systems Engineering, The Ohio State University.
- Cook, R. I., Woods, D. D, and Howie. M. B. (1992). Unintentional delivery of vasoactive drugs with an electromechanical infusion device. *Journal of Cardiothoracic and Vascular Anesthesia*, 6:1--7, 1992.
- Cook, R. I. and Woods D. D. (1996). Implications of automation surprises in aviation for the future of total intravenous anesthesia (TIVA). *Journal of Clinical Anesthesia*, 8:29s-37s.

Cook, R. I., Woods, D. D., Miller, C. (1998) *A Tale of Two Stories: Contrasting Views on Patient Safety*. National Patient Safety Foundation, Chicago IL, April 1998 (www.npsf.org/exec/report.html).

Cook, R. I., Render, M. L. and Woods, D.D. (2000). Gaps in the continuity of care and progress on patient safety. *British Medical Journal*, 320, 791-794, March 18, 2000.

Dekker, S. and Woods, D. D. (1999). Extracting Data from the Future: Assessment and Certification of Envisioned Systems. In S. Dekker and E. Hollnagel (Eds.), *Coping with Computers in the Cockpit*. Ashgate, p. 7-27.

Dekker, S. (in press). *The Field Guide to Human Error*. Cranfield University Press.

De Keyser, V. (1992). Why field studies? In M.G. Helander & M. Nagamachi (eds.) *Design for Manufacturability*, London: Taylor & Francis.

De Keyser, V. and Woods, D.D. (1990). Fixation errors: Failures to revise situation assessment in dynamic and risky systems. In A.G. Colombo and A. Saiz de Bustamante, editors, *Systems Reliability Assessment*, Kluwer Academic Publishers, Dordrechts, The Netherlands, p.231—252.

De Keyser, V. and Samercay, R. (1998). Activity theory, situated action and simulators. *Le Travail Humain*, 61(4), 305—312.

Dominguez, C., Flach, J., Lake, P., McKellar, D., & Dunn, M. (in press). The conversion decision in laparoscopic surgery: Knowing your limits and limiting your risks. In J. Shanteau, K. Smith, & P. Johnson (Eds.), *Psychological explorations of competent decision making*. New York: Cambridge University Press.

Ericsson, K. A. and Simon, H. A. *Protocol Analysis: Verbal Reports as Data*. MIT Press Cambridge, Mass., 1984.

Feltovich, P., Ford, K. and Hoffman, R., eds. (1997). *Expertise in Context*. Cambridge MA: MIT Press.

Fischhoff, B. (1975). Hindsight – foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, 1(3), 288-299.

Flach, J. M. and Dominguez, C. O. (1995). Use-Centered Design: Integrating the User, Instrument, and Goal. *Ergonomics in Design*, July.

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, 51(4), 327—358.

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

Gaba, D. M., Howard, S. K., and Jump, B. (1994). Production pressure in the work environment: California anesthesiologists' attitudes and experiences. *Anesthesiology* 81:488-500.

Galegher, J., Kraut, R. and Egido, C., Eds. 1990 *Intellectual teamwork: Social and technical bases of cooperative work*, Hillsdale, NJ: Erlbaum.

Grabowski, M. and Roberts, K. H. (1997). Risk mitigation in large-scale systems: Lessons from high reliability organizations. *California Management Review*, 39(4), Summer.

Greenbaum, J. and Kyng, M., Eds. (1991) *Design at Work.: Cooperative Design of Computer Systems*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Grudin, J. (1994). Computer-supported cooperative work: History and focus. *IEEE Computer*, 27(5), 19-27.

Guerlain, S., Smith, P.J., Obradovich, J.H., Rudmann, S., Strohm, P., Smith, J., and Svirebely, J. (1996). Dealing with brittleness in the design of expert systems for immunohematology. *Immunohematology*, 12, 101--107.

Guerlain, S., Smith, P.J., Obradovich, J. Heintz, Rudmann, S., Strohm, P. Smith, J.W., Svirbely, J., and Sachs, L. (1999). Interactive Critiquing as a Form of Decision Support: An Empirical Evaluation. *Human Factors* 41(1), 72-89.

Heath, C. and Luff, P. (1992) Collaboration and Control: Crisis Management and Multimedia Technology in London Underground Line Control Rooms. *Computer-Supported Cognitive Work*, 1, 69-94.

Hendee, W. (ed.). (1999). Proceedings of *Enhancing Patient Safety and Reducing Errors in Health Care*. National Patient Safety Foundation, Chicago IL (held at Annenberg Center for Health Sciences, Rancho Mirage CA Nov. 8–10, 1998).

Hirschhorn, L. (1993). Hierarchy vs. bureaucracy: The case of a nuclear reactor. In K. H. Roberts (Ed.), *New challenges to understanding organizations*. New York: McMillan.

Hoffman R. and Woods. D.D. (2000). Special Section on Studying Cognitive Systems in Context. *Human Factors*, 42(1).

Hollnagel. E. (1993). *Human Reliability Analysis: Context and Control*. London: Academic Press.

Howard, S. K., Gaba, D. M., Fish, K. J., Yang, G.S., and Sarnquist, F. H. (1992). Anesthesia crisis resource management training: teaching anesthesiologists to handle critical incidents. *Aviation, Space, and Environmental Medicine* 63:763-770.

Howard, S. K., Smith, B. E., Gaba, D. M., and Rosekind, M.R. (1997). Performance of well-rested vs. highly-fatigued residents: A simulator study. *Anesthesiology*; A-981.

Hutchins, E. (1995). How a cockpit remembers its speeds. *Cognitive Science*, 19, 265--288.

Jones, P. M. (1995). Cooperative work in mission operations: Analysis and implications for computer support. *Computer Supported Cooperative Work*, 3, 103-145.

Jordan, B. and Henderson, A. (1995) Interaction Analysis: Foundations and Practice. *The Journal for the Learning Sciences*. 4(1), 39-103.

Klein, G. (1998). *Sources of Power: How People Make Decisions*. . MIT Press, Cambridge MA.

Klein, G., Orasanu, J., Calderwood, R. and Zsombok, C., Eds., (1995). *Decision Making in Action: Models and Methods*. Norwood, NJ: Ablex.

LaBerge, D. (1995). *Attentional Processing: The Brain's Art of Mindfulness*. Cambridge, MA: Harvard University Press.

Layton, C., Smith, P. J. and McCoy, C. E. (1994). Design of a Cooperative Problem-Solving System for En-Route Flight Planning: An Empirical Evaluation. *Human Factors*, 36:94--119.

Leape, L. L. Woods, D. D. Hatlie, M. J. Kizer, K. W. Schroeder, S. A. and Lundberg, G. A. (1998). Promoting Patient Safety by Reducing Medical Errors, *JAMA*, 280, 1444-1447, October 28, 1998 (editorial)

Lerner, J. S. and Tetlock, P. E. (1999). Accounting for the Effects of Accountability. *Psychological Bulletin*, 125, 255--275.

Lin, L., Isla, R., Doniz, K., Harkness, H., Vicente, K., Doyle, D. (1998). Applying Human Factors to the design of medical equipment: Patient controlled analgesia. *Journal of Clinical Monitoring*, 14, 253—263.

Malin, J. T., Schreckenghost, D. L., Woods, D. D., Potter, S. S., Johannesen, L., Holloway, M., and Forbus, K. D. (1991). *Making Intelligent Systems Team Players*_(NASA Technical Memorandum 104738). NASA Johnson Space Center: Houston, TX.

Nielsen, J. (1993). *Usability Engineering*. Boston Academic Press.

Norman D. A. and Draper S. W. (Eds.) (1986) *User Centered System Design: New Perspectives on Human-Computer Interaction*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

- Norman, D. A. (1988). *The Design of Everyday Things*. New York, Basic Books.
- Norman, D. A. (1990). The 'problem' of automation: Inappropriate feedback and interaction, not 'over-automation.' *Philosophical Transactions of the Royal Society of London*, B 327:585--593.
- Norman, D. A. (1993). *Things that Make us Smart*. Addison-Wesley: Reading, MA.
- Nyssen, A. S. and De Keyser, V. (1998). Improving training in problem solving skills: Analysis of anesthetists' performance in simulated problem situations. *Le Travail Humain*, 61(4), 387—402.
- Nyssen, A. S. and Javaux, D. (1996). Analysis of synchronization constraints and associated errors in collective work environments, *Ergonomics*, 39, 1249—1264.
- Obradovich J.H. and Woods, D.D. (1996). Users as designers: How people cope with poor HCI design in computer-based medical devices. *Human Factors*, 38(4), 574—592.
- Palmer, L., Emanuel, L. and Woods, D. (2000). Managing Systems of Accountability for Patient Safety. National Patient Safety Foundation, in progress.
- Rasmussen J. (1986). *Information Processing and Human-Machine Interaction*. New York North Holland.
- Rasmussen J. (1990). The role of error in organizing behavior. *Ergonomics*. 33: 1185-1199.
- Rasmussen J. (1999). The concept of human error: Is it useful for the design of safe systems in health care? In Vincent C, deMoll B. (Eds.) *Risk and Safety in Medicine*. London: Elsevier, 1999.
- Rasmussen, J., Brehmer, B. and Lepat, J. (Eds.), (1991). *Distributed Decision Making: Cognitive Models for Cooperative Work*. Chichester, England: Wiley.
- Reason J. (1997). *Managing the Risks of Organizational Accidents*. Brookfield, VT: Ashgate.

- Rensink, R. A., O'Regan, J. K. and Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, 8(5), 368-373.
- Rochlin, G. I. (1999). Safe operation as a social construct. *Ergonomics*, 42(11): 1549—1560.
- Rochlin, G. I., La Porte, T. R. and Roberts, K. H. (1987). The self-designing high-reliability organization, Aircraft carrier flight operations at sea. *Naval War College Review*, Autumn, 76-90.
- Roth, E. M., Malin, J. T. and Schreckenghost, D. L. (1997). Paradigms for intelligent interface design. In M. Helander, T. K. Landauer, and P. Prabhu (Eds.), *Handbook of Human-Computer Interaction* (2nd ed.), North-Holland (pp. 1177-1201).
- Sanders, E. B.-N. (2000). Generative Tools for Co-Designing. In Scrivener, Ball and Woodcock (eds.) *Collaborative Design*. London: Springer-Verlag.
- Sarter, N. B., Woods, D. D., and Billings, C. E. (1997). Automation Surprises. In G. Salvendy (Ed.), *Handbook of Human Factors/Ergonomics*, (2nd ed.). New York, NY: Wiley.
- Senders, J. W. On patient injury and death stemming from the design of medical devices. Report, 12 September, 1993
- Sharpe, V. A. (2000). Behind Closed Doors: Accountability and Responsibility in Patient Care. *Journal of Medicine and Philosophy*, 25, 28—47.
- Smith, P., Woods, D., McCoy, E., Billings, C., Sarter, N., Denning R and Dekker, S. (1998). Using forecasts of future incidents to evaluate future ATM system designs. *Air Traffic Control Quarterly*, 6(1), 71—85.
- Stanton, N. (1994). *Human Factors of Alarm Design*, Taylor & Francis, London.
- Tetlock, P. E. (2000). Cognitive biases and organizational correctives: Do both disease and cure depend on the politics of the beholder? *Administrative Science Quarterly*, 45, 293—326.

Vicente, K. (1999). *Cognitive Work Analysis: Toward, Safe, Productive, and Healthy Computer-Based Work*. Mahwah, NJ, Erlbaum.

Weick, K. E. and Roberts, K. H. (1993). Collective mind and organizational reliability: The case of flight operations on an aircraft carrier deck. *Administration Science Quarterly*, 38, 357--381.

Westrum, R. (1993). Cultures with requisite imagination. In J. A. Wise, V. D. Hopkin and P. Stager (Eds.), *Verification and validation of complex systems: Human factors issues*. Springer-Verlag: Berlin. NATO ASI Series.

Winograd, T. and Woods, D. D. (1997). Challenges for Human-Centered Design. In *Human-Centered Systems: Information, Interactivity, and Intelligence*. J. Flanagan, T. Huang, P. Jones, S. Kasif, (Eds.), National Science Foundation, Washington DC, July, 1997.

Woods D. D., Johannesen L., Cook R. I., and Sarter, N. B. (1994). *Behind Human Error: Cognitive Systems, Computers and Hindsight*. Crew Systems Ergonomic Information and Analysis Center, Wright Patterson Air Force Base, Dayton OH.

Woods, D. D., Cook, R. I. and Billings, C.E. (1995). The impact of technology on physician cognition and performance. *Journal of Clinical Monitoring*, 11:92-95.

Woods, D. D. and Cook, R. I. (1999). Perspectives on Human Error: Hindsight Bias and Local Rationality. In F. Durso (Eds.), *Handbook of Applied Cognitive Psychology*, New York, Wiley, p. 141-171.

Zhang, J. (1997). The nature of external representations in problem solving. *Cognitive Science*, 21(2): 179--217.