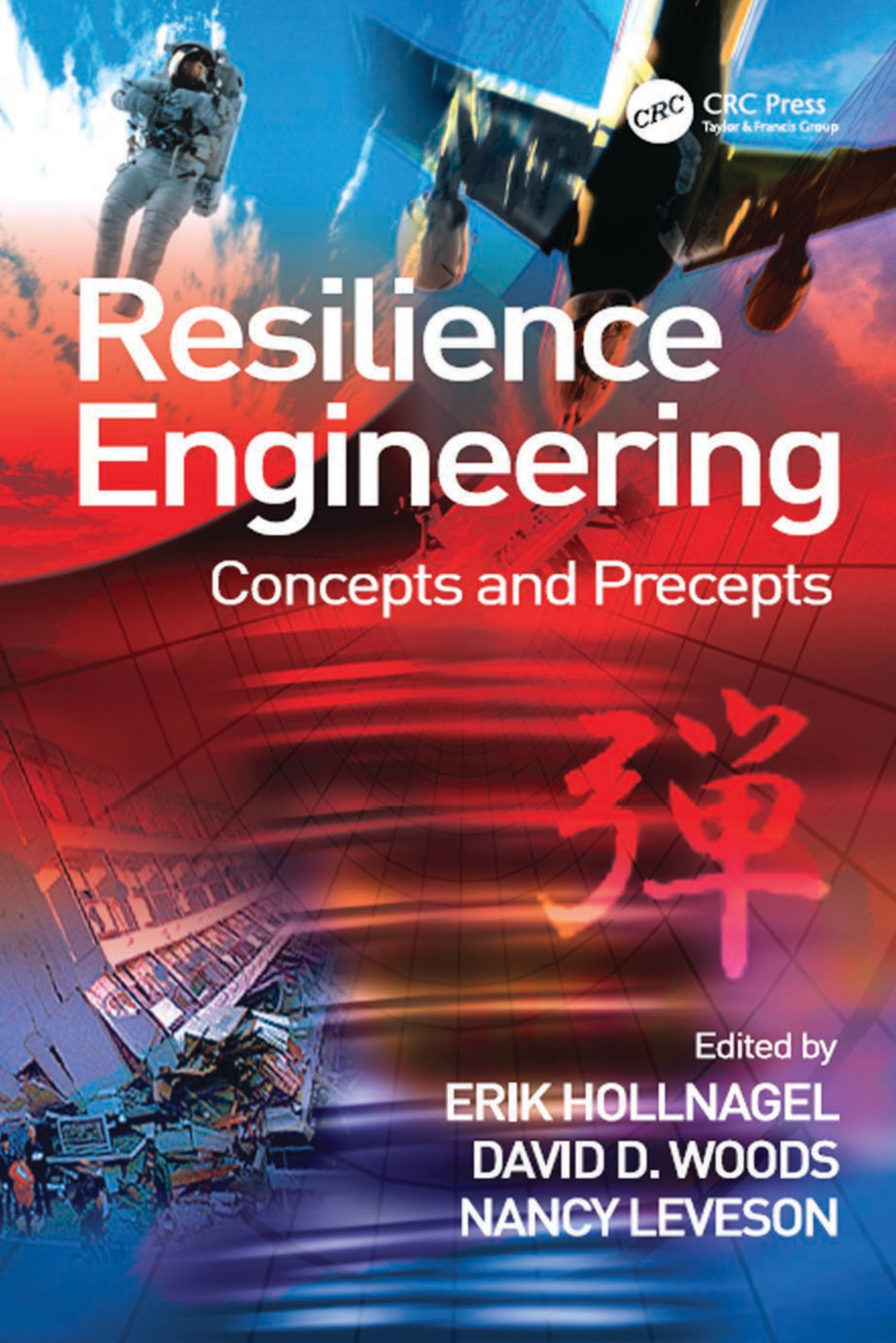
CRC
CRC Press
Taylor & Francis Group
Resilience Engineering
Concepts and Precepts
予弾
Edited by
ERIK HOLLNAGEL
DAVID D. WOODS
NANCY LEVESON

RESILIENCE ENGINEERING

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Resilience Engineering
Concepts and Precepts

Edited by
ERIK HOLLNAGEL
Linköping University, Sweden
DAVID D. WOODS
Ohio State University, USA
NANCY LEVESON
Massachusetts Institute of Technology, USA

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Printed on acid-free paper
Version Date: 20160226

International Standard Book Number-13: 978-0-7546-4641-9 (Hardback) 978-0-7546-4904-5 (Paperback)

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Contents

Preface

Decades of efforts aimed at understanding what safety is and why accidents happen have led to several significant insights. One is that untoward events more often are due to an unfortunate combination of a number of conditions, than to the failure of a single function or component. Another is that failures are the flip side of successes, meaning that there is no need to evoke special failure mechanisms to explain the former. Instead, they both have their origin in performance variability on the individual and systemic levels, the difference being how well the system was controlled.

It follows that successes, rather than being the result of careful planning, also owe their occurrence to a combination of a number of conditions. While we like to think of successes as the result of skills and competence rather than of luck, this view is just as partial as the view of failures as due to incompetence or error. Even successes are not always planned to happen exactly as they do, although they of course usually are desired – just as the untoward events are dreaded.

A case in point is the symposium behind this book. As several of the chapters make clear, the notion of resilience had gradually emerged as the logical way to overcome the limitations of existing approaches to risk assessment and system safety. Ideas about resilience had been circulated more or less formally among several of the participants and the need of a more concerted effort was becoming obvious. Concurrently, a number of individuals and groups in the international community had begun to focus on a similar class of problems, sometimes talking directly about resilience and sometimes using related terms. In the USA, the Santa Fe Institute had begun programmes on robustness in natural and engineering systems and on robustness in social processes. Within the school of high-reliability organizations, the term *resilience* appeared in paper titles, e.g., Sutcliffe & Vogus (2003). A related concept was the proposal of a conceptual framework, named Highly Optimised Tolerance (HOT), to study fundamental aspects of complexity, including robust behaviour (e.g., Carlson & Doyle, 2000 & 2002). In Europe, a research organization of scientists and practitioners from many disciplines collaborated to explore the dynamics of social-

ecological systems under the name of the Resilience Alliance, while another group was called the Information Systems for Crisis Response and Management or ISCRAM community.

The intention of getting a group of international experts together for an extended period of time to discuss resilience was, however, just one component. Some of the others were that one of the protagonists (David D. Woods) was going to spend some time in Europe, that initial inquiries indicated that the basic funding would be available, and that most of the people whom we had in mind were able and willing to interrupt their otherwise busy schedules to attend the symposium.

The symposium itself was organized as a loosely structured set of discussions with a common theme, best characterized as long discussions interrupted by short presentations – prepared as well as *ad hoc*. The objective of the symposium was to provide an opportunity for experts to meet and debate the present and future of Resilience Engineering as well as to provide a tentative definition of organizational resilience. Readers are invited to judge for themselves whether these goals were achieved and whether the result is, indeed, a success. If so, the credit goes to the participants both for their willingness to take part in a process of creative chaos during one October week in Söderköping, and for their discipline in producing the written contributions afterwards.

We would also like to thank the two main sponsors, the Swedish Nuclear Power Inspectorate (SKI) and the Swedish Civil Aviation Administration (LFV), who were willing to support something that is not yet an established discipline. Thanks are also due to Kyla Steele and Rogier Woltjer for practical and invaluable assistance both during the symposium and the editing process.

Linköping, July 2005

Erik Hollnagel David D. Woods Nancy G. Leveson

Prologue: Resilience Engineering Concepts

David D. Woods
Erik Hollnagel

Hindsight and Safety

Efforts to improve the safety of systems have often – some might say always – been dominated by hindsight. This is so both in research and in practice, perhaps more surprising in the former than in the latter. The practical concern for safety is usually driven by events that have happened, either in one's own company or in the industry as such. There is a natural motivation to prevent such events from happening again, in concrete cases because they may incur severe losses – of equipment and/or of life – and in general cases because they may lead to new demands for safety from regulatory bodies, such as national and international administrations and agencies. New demands are invariably seen as translating into increased costs for companies and are for that reason alone undesirable. (This is, however, not an inevitable consequence, especially if the company takes a longer time perspective. Indeed, for some businesses it makes sense to invest proactively in safety, although cases of that are uncommon. The reason for this is that sacrificing decisions usually are considered over a short time horizon, in terms of months rather than years or in terms of years rather than decades.)

In the case of research, i.e., activities that take place at academic institutions rather than in industries and are driven by intellectual rather than economic motives, the effects of hindsight ought to be less marked. Research should by its very nature be looking to problems that go beyond the immediate practical needs, and hence address issues that are of a more principal nature. Yet even research – or perhaps one

should say: researchers – are prone to the effects of hindsight, as pointed out by Fischhoff (1975). It is practically a characteristic of human nature – and an inescapable one at that – to try to make sense of what has happened, to try to make the perceived world comprehensible. We are consequently constrained to look at the future in the light of the past. In this way our experience or understanding of what has happened inevitably colours our anticipation and preparation for what could go wrong and thereby holds back the requisite imagination that is so essential for safety (Adamski & Westrum, 2003). Approaches to safety and risk prediction furthermore develop in an incremental manner, i.e., the tried and trusted approaches are only changed when they fail and then usually by adding one more factor or element to account for the unexplained variability. Examples are easy to find such as 'human error', 'organisational failures', 'safety culture', 'complacency', etc. The general principle seems to be that we add or change just enough to be able to explain that which defies the established framework of explanations. In contrast, resilience engineering tries to take a major step forward, not by adding one more concept to the existing vocabulary, but by proposing a completely new vocabulary, and therefore also a completely new way of thinking about safety. With the risk of appearing overly pretentious, it may be compared to a paradigm shift in the Kuhnian sense (Kuhn, 1970).

When research escapes from hindsight and from trying merely to explain what *has* happened, studies reveal the sources of resilience that usually allow people to produce success when failure threatens. Methods to understand the basis for technical work shows how workers are struggling to anticipate paths that may lead to failure, actively creating and sustaining failure-sensitive strategies, and working to maintain margins in the face of pressures to do more and to do it faster (Woods & Cook, 2002). In other words, doing things safely always has been and always will be part of operational practices – on the individual as well as the organisational level. It is, indeed, almost a biological law that organisms or systems (including organisations) that spend all efforts at the task at hand and thereby neglect to look out for the unexpected, run a high risk of being obliterated, of meeting a speedy and unpleasant demise. (To realise that, you only need to look at how wild birds strike a balance between head-down and head-up time when eating.) People in their different roles within an organisation are aware of potential paths to failure and therefore develop failure-

sensitive strategies to forestall these possibilities. Failing to do that brings them into a reactive mode, a condition of constant fire-fighting. But fires, whether real or metaphorical, can only be effectively quelled if the fire-fighters are proactive and able to make the necessary sacrifices (McLennan et al., 2005).

Against this background, failures occur when multiple contributors – each necessary but only jointly sufficient – combine. Work processes or people do not choose failure, but the likelihood of failures grow when production pressures do not allow sufficient time – and effort – to develop and maintain the precautions that normally keep failure at bay. Prime among these precautions is to check all necessary conditions and to take nothing important for granted. Being thorough as well as efficient is the hallmark of success. Being efficient without being thorough may gradually or abruptly create conditions where even small variations can have serious consequences. Being thorough without being efficient rarely lasts long, as organisations are pressured to meet new demands on resources. To understand how failure sometimes happens one must first understand how success is obtained – how people learn and adapt to create safety in a world fraught with gaps, hazards, trade-offs, and multiple goals (Cook et al., 2000).

The thesis that leaps out from these results is that failure, as individual failure or performance failure on the system level, represents the temporary inability to cope effectively with complexity. Success belongs to organisations, groups and individuals who are resilient in the sense that they recognise, adapt to and absorb variations, changes, disturbances, disruptions, and surprises – especially disruptions that fall outside of the set of disturbances the system is designed to handle (Rasmussen, 1990; Rochlin, 1999; Weick et al., 1999; Sutcliffe & Vogus, 2003).

From Reactive to Proactive Safety

This book marks the maturation of a new approach to safety management. In a world of finite resources, of irreducible uncertainty, and of multiple conflicting goals, safety is created through proactive resilient processes rather than through reactive barriers and defences. The chapters in this book explore different facets of resilience as the

ability of systems to anticipate and adapt to the potential for surprise and failure.

Until recently, the dominant safety paradigm was based on searching for ways in which limited or erratic human performance could degrade an otherwise well designed and 'safe system'. Techniques from many areas such as reliability engineering and management theory were used to develop 'demonstrably safe' systems. The assumption seemed to be that safety, once established, could be maintained by requiring that human performance stayed within prescribed boundaries or norms. Since 'safe' systems needed to include mechanisms that guarded against people as unreliable components, understanding how human performance could stray outside these boundaries became important.

According to this paradigm, 'error' was something that could be categorised and counted. This led to numerous proposals for taxonomies, estimation procedures, and ways to provide the much needed data for error tabulation and extrapolation. Studies of human limits became important to guide the creation of remedial or prosthetic systems that would make up for the deficiencies of people. Since humans, as unreliable and limited system components, were assumed to degrade what would otherwise be flawless system performance, this paradigm often prescribed automation as a means to safeguard the system from the people in it. In other words, in the 'error counting' paradigm, work on safety comprised protecting the system from unreliable, erratic, and limited human components (or, more clearly, protecting the people at the blunt end – in their roles as managers, regulators and consumers of systems – from unreliable 'other' people at the sharp end – who operate and maintain those systems).

When researchers in the early 1980s began to re-examine human error and collect data on how complex systems had failed, it soon became apparent that people actually provided a positive contribution to safety through their ability to adapt to changes, gaps in system design, and unplanned for situations. Hollnagel (1983), for instance, argued for the need of a theory of action, including an account of performance variability, rather than a theory of 'error', while Rasmussen (1983) noted that 'the operator's role is to make up for holes in designers' work.' Many studies of how complex systems succeeded and sometimes failed found that the formal descriptions of work embodied in policies, regulations, procedures, and automation were incomplete as

models of expertise and success. Analyses of the gap between formal work prescriptions and actual work practices revealed how people in their various roles throughout systems always struggled to anticipate paths toward failure, to create and sustain failure-sensitive strategies, and to maintain margins in the face of pressures to increase efficiency (e.g., Cook et al, 2000). Overall, analysis of such 'second stories' taught us that failures represented breakdowns in adaptations directed at coping with complexity while success was usually obtained as people learned and adapted to create safety in a world fraught with hazards, trade-offs, and multiple goals (Rasmussen, 1997). In summary, these studies revealed:

- How workers and organisations continually revise their approach to work in an effort to remain sensitive to the possibility for failure.
- How distant observers of work, and the workers themselves, are only partially aware of the current potential for failure.
- How 'improvements' and changes create new paths to failure and new demands on workers, despite or because of new capabilities.
- How the strategies for coping with these potential paths can be either strong and resilient or weak and mistaken.
- How missing the side effects of change is the most common form of failure for organisations and individuals.
- How a culture of safety depends on remaining dynamically engaged in new assessments and avoiding stale, narrow, or static representations of the changing paths (revising or reframing the understanding of paths toward failure over time).
- How overconfident people can be that they have already anticipated the types and mechanisms of failure, and that the strategies they have devised are effective and will remain so.
- How continual effort after success in a world of changing pressures and hazards is fundamental to creating safety.

In the final analysis, safety is not a commodity that can be tabulated. It is rather a chronic value 'under our feet' that infuses all aspects of practice. Safety is, in the words of Karl Weick, a dynamic non-event. Progress on safety therefore ultimately depends on providing workers and managers with information about changing vulnerabilities and the ability to develop new means for meeting these.

Resilience

Resilience engineering is a paradigm for safety management that focuses on how to help people cope with complexity under pressure to achieve success. It strongly contrasts with what is typical today – a paradigm of tabulating error as if it were a thing, followed by interventions to reduce this count. A resilient organisation treats safety as a core value, not a commodity that can be counted. Indeed, safety shows itself only by the events that do not happen! Rather than view past success as a reason to ramp down investments, such organisations continue to invest in anticipating the changing potential for failure because they appreciate that their knowledge of the gaps is imperfect and that their environment constantly changes. One measure of resilience is therefore the ability to create foresight – to anticipate the changing shape of risk, before failure and harm occurs (Woods, 2005a).

The initial steps in developing a practice of Resilience Engineering have focused on methods and tools:

- to analyse, measure and monitor the resilience of organisations in their operating environment.
- to improve an organisation's resilience vis-à-vis the environment.
- to model and predict the short- and long-term effects of change and line management decisions on resilience and therefore on risk.

This book charts the efforts being made by researchers, practitioners and safety managers to enhance resilience by looking for ways to understand the changing vulnerabilities and pathways to failure. These efforts begin with studies of how people cope with complexity – usually successfully. Analyses of successes, incidents, and breakdowns reveal the normal sources of resilience that allow systems to produce success when failure threatens. These events and other measures indicate the level and kinds of brittleness/resilience the system in question exhibits. Such indicators will allow organisations to develop the mechanisms to create foresight, to recognise, anticipate, and defend against paths to failure that arise as organisations and technology change.

Part I: Emergence

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 1

Resilience – the Challenge of the Unstable

Erik Hollnagel

Safety is the sum of the accidents that do not occur. While accident research has focused on accidents that occurred and tried to understand why, safety research should focus on the accidents that did not occur and try to understand why.

Understanding Accidents

Research into system safety is faced with the conundrum that while there have been significant developments in the understanding of how accidents occur, there has been no comparable developments in the understanding of how we can adequately assess and reduce risks. A system is safe if it is impervious and resilient to perturbations and the identification and assessment of possible risks is therefore an essential prerequisite for system safety. Since accidents and risk assessment furthermore are two sides of the same coin, so to speak, and since both are constrained in equal measure by the underlying models and theories, it would be reasonable to assume that developments in system safety had matched developments in accident analysis. Just as we need to have an aetiology of accidents, a study of possible causes or origins of accidents, we also need to have an aetiology of safety – more specifically of what safety is and of how it may be endangered. This is essential for work on system safety in general and for resilience engineering in particular. Yet for reasons that are not entirely clear, such a development has been lacking.

The value or, indeed, the necessity of having an accident model has been recognised for many years, such as when Benner (1978) noted that:

> Practical difficulties arise during the investigation and reporting of most accidents. These difficulties include the determination of the scope of the phenomenon to investigate, the identification of the data required, documentation of the findings, development of recommendations based on the accident findings, and preparation of the deliverables at the end of the investigation. These difficulties reflect differences in the purposes for the investigations, which in turn reflect different perceptions of the accident phenomenon.

The 'different perceptions of the accident phenomenon' are what in present day terminology are called the accident models. Accident models seem to have started by relatively uncomplicated single-factor models of, e.g., accident proneness (Greenwood & Woods, 1919) and developed via simple and complex linear causation models to present-day systemic or functional models (for a recent overview of accident models, see Hollnagel, 2004.)

The archetype of a simple linear model is Heinrich's (1931) Domino model, which explains accidents as the linear propagation of a chain of causes and effects (Figure 1.1). This model was associated with one of the earliest attempts of formulating a complete theory of safety, expressed in terms of ten axioms of industrial safety (Heinrich et al., 1980, p. 21). The first of these axioms reads as follows:

> The occurrence of an injury invariably results from a completed sequence of factors – the last one of these being the accident itself. The accident in turn is invariably caused or permitted directly by the unsafe act of a person and/or a mechanical or physical hazard.

According to this view, an accident is basically a disturbance inflicted on an otherwise stable system. Although the domino model has been highly useful by providing a concrete approach to understanding accidents, it has unfortunately also reinforced the misunderstanding that accidents have a root cause and that this root cause can be found by searching backwards from the event through the chain of causes that preceded it. More importantly, the domino model suggests that system safety can be enhanced by disrupting the linear sequence, either by 'removing' a 'domino' or by 'spacing' the 'dominos'

further apart. (The problems in providing a translation from model components to the world of practice are discussed further in Chapter 17.)

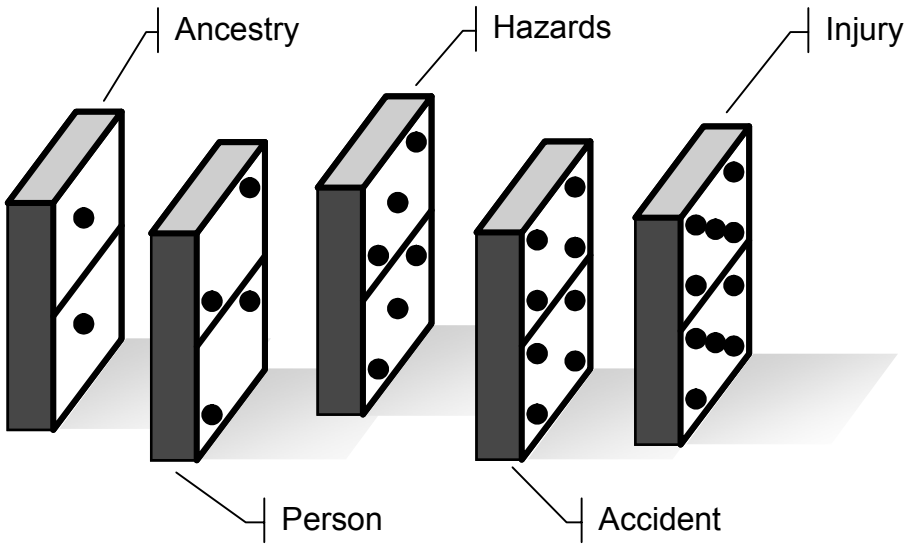

Figure 1.1: Simple linear accident model (Domino model)

The comparable archetype of a complex linear model is the well-known Swiss cheese model first proposed by Reason (1990). According to this, accidents can be seen as the result of interrelations between real time 'unsafe acts' by front-line operators and latent conditions such as weakened barriers and defences, represented by the holes in the slices of 'cheese', cf. Figure 1.2. (Models that describe accidents as a result of interactions among agents, defences and hosts are also known as epidemiological accident models.) Although this model is technically more complex than the domino model, the focus remains on structures or components and the functions associated with these, rather than on the functions of the overall system as such. The Swiss cheese model comprises a number of identifiable components where failures (and risks) are seen as due to failures of the components, most conspicuously as the breakdown of defences. Although causality is no longer a *single* linear propagation of effects, an accident is still the result of a relatively clean combination of events, and the failure of a barrier is still the failure of an individual component. While the whole idea of a

complex linear model such as this is to describe how coincidences occur, it cannot detach itself from a structural perspective involving the fixed relations between agents, hosts, barriers and environments.

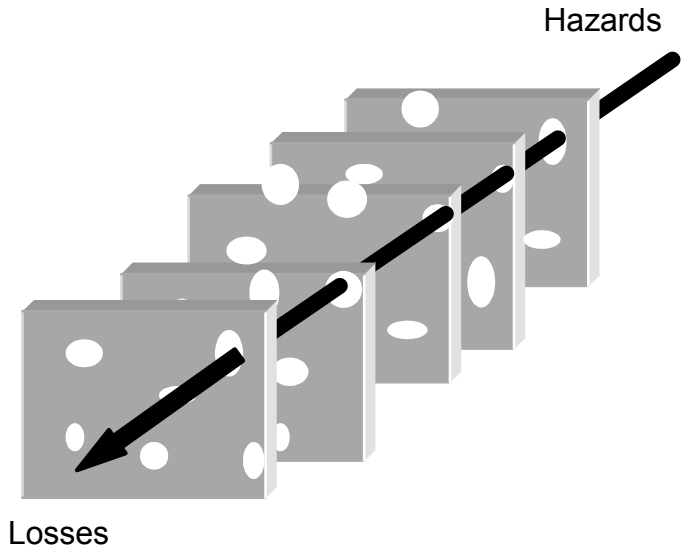

Figure 1.2: Complex linear accident model (Swiss cheese model)

Since some accidents defy the explanatory power of even complex linear models, alternative explanations are needed. Many authors have pointed out that accidents can be seen as due to an unexpected combination or aggregation of conditions or events (e.g., Perrow, 1984). A practical term for this is *concurrence*, meaning the temporal property of two (or more) things happening at the same time and thereby affecting each other. This has led to the view of accidents as non-linear phenomena that emerge in a complex system, and the models are therefore often called systemic accident models, cf., Chapter 4.

This view recognises that complex system performance always is variable, both because of the variability of the environment and the variability of the constituent subsystems. The former may appropriately be called exogenous variability, and the latter endogenous variability. The endogenous variability is to a large extent attributable to the people in the system, as individuals and/or groups. This should nevertheless not be taken to imply that human performance is wrong or erroneous

in any way. On the contrary, performance variability is necessary if a joint cognitive system, meaning a human-machine system or a socio-technical system, is successfully to cope with the complexity of the real world (Hollnagel & Woods, 2005). The essence of the systemic view can be expressed by the following four points:

- Normal performance and as well as failures are emergent phenomena. Neither can therefore be attributed to or explained by referring to the (mal)functions of specific components or parts. Normal performance furthermore differs from *normative* performance: it is not what is prescribed by rules and regulation but rather what takes place as a result of the adjustments required by a partly unpredictable environment. Technically speaking, normal performance represents the equilibrium that reflects the regularity of the work environment.
- The outcomes of actions may sometimes differ from what was intended, expected or required. When this happens it is more often due to variability of context and conditions than to the failures of actions (or the failure of components or functions). On the level of individual human performance, local optimisation or adjustment is the norm rather than the exception as shown by the numerous shortcuts and heuristics that people rely on in their work.
- The adaptability and flexibility of human work is the reason for its efficiency. Normal actions are successful because people adjust to local conditions, to shortcomings or quirks of technology, and to predictable changes in resources and demands. In particular, people quickly learn correctly to anticipate recurring variations; this enables them to be proactive, hence to save the time otherwise needed to assess a situation.
- The adaptability and flexibility of human work is, however, also the reason for the failures that occur, although it is rarely the cause of such failures. Actions and responses are almost always based on a limited rather than complete analysis of the current conditions, i.e., a trade-off of thoroughness for efficiency. Yet since this is the normal mode of acting, normal actions can, by definition, not be wrong. Failures occur when this adjustment goes awry, but both the actions and the principles of adjustment are technically correct.

Accepting a specific model does not only have consequences for how accidents are understood, but also for how resilience is seen. In a simple linear model, resilience is the same as being impervious to specific causes; using the domino analogy, the pieces either cannot fall or are so far apart that the fall of one cannot affect its neighbours. In a complex linear model, resilience is the ability to maintain effective barriers that can withstand the impact of harmful agents and the erosion that is a result of latent conditions. In both cases the transition from a safe to an unsafe state is tantamount to the failure of some component or subsystem and resilience is the ability to endure harmful influences. In contrast to that, a systemic model adopts a functional point of view in which resilience is an organisation's ability efficiently to adjust to harmful influences rather than to shun or resist them. An unsafe state may arise because system adjustments are insufficient or inappropriate rather than because something fails. In this view failure is the flip side of success, and therefore a normal phenomenon.

Anticipating Risks

Going from accident analysis to risk assessment, i.e., from understanding what *has* happened to the identification of events or conditions that in the future *may* endanger system safety, it is also possible to find a number of different models of risks, although the development has been less noticeable. Just as there are single-factor accident models, there are risk assessment models that consider the failure of individual components, such as Failure Mode and Effects Analysis. Going one step further, the basic model to describe a sequence of actions is the event tree, corresponding to the simple linear accident model. The event tree represents a future accident as a result of possible failures in a pre-determined sequence of events organised as a binary branching tree. The 'root' is the initiating event and the 'leaves' are the set of possible outcomes – either successes or failures. In a similar manner, a fault tree corresponds to a complex linear model or an epidemiological model. The fault tree describes the accident as the result of a series of logical combinations of conditions, which are necessary and sufficient to produce the top event, i.e., the accident (cf. Figure 1.3).

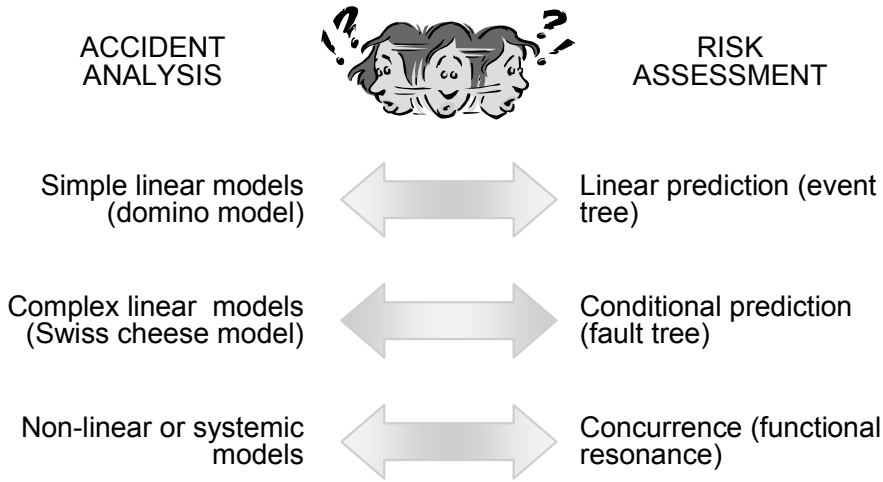

Figure 1.3: Models for accident analysis and risk assessment

Event trees and fault trees may be adequate for risk assessment when the outcomes range from incidents to smaller accidents, since these often need less elaborate explanations and may be due to relatively simple combinations of factors. Most major accidents, however, are due to complex concurrences of multiple factors, some of which have no apparent *a priori* relations. Event and fault trees are therefore unable fully to describe them – although this does not prevent event trees from being the favourite tool for Probabilistic Safety Assessment methods in general. It is, indeed, a consequence of the systemic view that the potential for (complex) accidents cannot be described by a fixed structure such as a tree, graph or network, but must invoke some way of representing dynamic bindings or couplings, for instance as in the functional resonance accident model (Hollnagel, 2004). Indeed, the problems of risk assessment may to a large degree arise from a reliance on graphical representations, which – as long as they focus on descriptions of links between parts – are unable adequately to account for concurrence and for how a stable system slowly or abruptly may become unstable.

The real challenge for system safety, and therefore also for resilience engineering, is to recognise that complex systems are dynamic and that a state of dynamic stability sometimes may change into a state of dynamic instability. This change may be either abrupt, as in an

accident, or slow, as in a gradual erosion of safety margins. Complex systems must perforce be dynamic since they must be able to adjust their performance to the conditions, cf. the four points listed above. These adjustments cannot be pre-programmed or built into the system, because they cannot be anticipated at the time of design – and sometimes not even later. It is practically impossible to design for every little detail or every situation that may arise, something that procedure writers have learned to their dismay. Complex systems must, however, be *dynamically* stable, or constrained, in the sense that the adjustments do not get out of hand but at all times remain under control. Technically this can be expressed by the concept of damping, which denotes the progressive reduction or suppression of deviations or oscillation in a device or system (over time). A system must obviously be able to respond to changes and challenges, but the responses must not lead the system to lose control. The essence of resilience is therefore the intrinsic ability of an organisation (system) to maintain or regain a dynamically stable state, which allows it to continue operations after a major mishap and/or in the presence of a continuous stress.

Dictionaries commonly define resilience as the ability to 'recover quickly from illness, change, or misfortune', one suggestive synonym being buoyancy or a bouncing quality. Using this definition, it stands to reason that it is easier to recover from a potentially destabilising disturbance if it is detected early. The earlier an adjustment is made, the smaller the resulting adjustments are likely to be. This has another beneficial effect, which is a corollary of the substitution myth. According to this, artefacts are value neutral in the sense that the introduction of an artefact into a system only has the intended and no unintended effects (Hollnagel & Woods, 2005, p. 101). Making a response to or recovering from a disturbance requires an adjustment, hence a change to the system. Any such change may have consequences that go beyond the local and intended effects. If the consequences from the recovery are small, i.e., if they can effectively be confined to a subsystem, then the likelihood of negative side-effects is reduced and the resilience is therefore higher. As a result of this, the definition of resilience can be modified to be the ability of a system or an organisation to react to and recover from disturbances at an early stage, with minimal effect on the dynamic stability. The challenges to system safety come from instability, and resilience engineering is an expression of the methods and principles that prevent this from taking place.

For the analytical part, resilience engineering amounts to a systemic accident model as outlined above. Rather than looking for causes we should look for concurrences, and rather than seeing concurrences as exceptions we should see them as normal and therefore also as inevitable. This may at times lead to the conclusion that even though an accident happened nothing really went wrong, in the sense that nothing happened that was out of the ordinary. Instead it is the concurrence of a number of events, just on the border of the ordinary, that constitutes an explanation of the accident or event.

For the predictive part, resilience engineering can be addressed, e.g., by means of a functional risk identification method, such as proposed by the functional resonance accident model (Hollnagel, 2004). To make progress on resilience engineering we have to go beyond failure modes of component (subsystem, human) to concurrences. This underlines the functional view since concurrences take place among events and functions rather than among components. The challenge is understand when a system may lose its dynamic stability and become unstable. To do so requires powerful methods combined with plenty of imagination (Adamski & Westrum, 2003).

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Systems are Ever-Changing

Yushi Fujita

> No system (i.e., combination of artifact and humans) can avoid changes. Changes occur continuously throughout a system's lifetime. This should be regarded as a destiny. The incompleteness of a system is partly attributable to this ever-changing nature. Changes take place because of external drivers (e.g., economic pressure). But changes also take place because of internal drivers. For instance, humans are always motivated to make changes that they think will improve system administration; humans often find unintended ways of utilizing the artifact; leaders are encouraged to introduce new visions in order to stimulate and lead people; … Like these, the system is always subject to changes, hence metamorphosing itself like a living matter. This floating nature often causes mismatches between administrative frameworks and the ways in which the system is actually utilized.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 2

Essential Characteristics of Resilience

David D. Woods

Avoiding the Error of the Third Kind

When one uses the label 'resilience,' the first reaction is to think of resilience as if it were adaptability, i.e., as the ability to absorb or adapt to disturbance, disruption and change. But all systems adapt (though sometimes these processes can be quite slow and difficult to discern) so resilience cannot simply be the adaptive capacity of a system. I want to reserve resilience to refer to the broader capability – how well can a system handle disruptions and variations that fall outside of the base mechanisms/model for being adaptive as defined in that system.

This depends on a distinction between understanding how a system is competent at designed-for-uncertainties, which defines a 'textbook' performance envelope and how a system recognizes when situations challenge or fall outside that envelope – unanticipated variability or perturbations (see parallel analyses in Woods et al., 1990 and Carlson & Doyle, 2000; Csete & Doyle, 2002). Most discussions of definitions of 'robustness' in adaptive systems debate whether resilience refers to first or second order adaptability (Jen, 2003). In the end, the debates tend to settle on emphasizing the system's ability to handle events that fall outside its design envelope and debate what is a design envelope, what events challenge or fall outside that envelope, and how does a system see what it has failed to build into its design (e.g., see url: http://discuss.santafe.edu/robustness/).

The area of textbook competence is in effect a model of variability/uncertainty and a model of how the strategies/plans

/countermeasures in play handle these, mostly successfully. Unanticipated perturbations arise (a) because the model implicit and explicit in the competence envelope is incomplete, limited or wrong and (b) because the environment changes so that new demands, pressures, and vulnerabilities arise that undermine the effectiveness of the competence measures in play.

Resilience then concerns the ability to recognize and adapt to handle unanticipated perturbations that call into question the model of competence, and demand a shift of processes, strategies and coordination. When evidence of holes in the organization's model builds up, the risk is what Ian Mitroff called many years ago, the error of the third kind, or solving the wrong problem (Mitroff, 1974). This is a kind of under-adaptation failure where people persist in applying textbook plans and activities in the face of evidence of changing circumstances that demand a qualitative shift in assessment, priorities, or response strategy.

This means resilience is concerned with monitoring the boundary conditions of the current model for competence (how strategies are matched to demands) and adjusting or expanding that model to better accommodate changing demands. The focus is on assessing the organization's adaptive capacity relative to challenges to that capacity – what sustains or erodes the organization's adaptive capacities? Is it degrading or lower than the changing demands of its environment? What dynamics challenge or go beyond the boundaries of the competence envelope? Is the organization as well adapted as it thinks it is? Note that boundaries are properties of the model that defines the textbook competence envelope relative to the uncertainties and perturbations it is designed for (Rasmussen, 1990a). Hence, resilience engineering devotes effort to make observable the organization's model of how it creates safety, in order to see when the model is in need of revision.

To do this, Resilience Engineering must monitor organizational decision-making to assess the risk that the organization is operating nearer to safety boundaries than it realizes (Woods, 2005a). Monitoring resilience should lead to interventions to manage and adjust the adaptive capacity as the system faces new forms of variation and challenges.

Monitoring and managing resilience, or its absence, brittleness, is concerned with understanding how the system adapts and to what kinds of disturbances in the environment, including properties such as:

- buffering capacity: the size or kinds of disruptions the system can absorb or adapt to without a fundamental breakdown in performance or in the system's structure;
- flexibility versus stiffness: the system's ability to restructure itself in response to external changes or pressures;
- margin: how closely or how precarious the system is currently operating relative to one or another kind of performance boundary;
- tolerance: how a system behaves near a boundary – whether the system gracefully degrades as stress/pressure increase or collapses quickly when pressure exceeds adaptive capacity.

In addition, cross-scale interactions are critical, as the resilience of a system defined at one scale depends on influences from scales above and below:

- Downward, resilience is affected by how organizational context creates or facilitates resolution of pressures/goal conflicts/dilemmas, for example, mismanaging goal conflicts or poor automation design can create authority-responsibility double binds for operational personnel (Woods et al., 1994; Woods, 2005b).
- Upward, resilience is affected by how adaptations by local actors in the form of workarounds or innovative tactics reverberate and influence more strategic goals and interactions (e.g., workload bottlenecks at the operational scale can lead to practitioner workarounds that make management's attempts to command compliance with broad standards unworkable; Cook et al., 2000).

As illustrated in the cases of resilience or brittleness described or referred to in this book, all systems have some degree of resilience and sources for resilience. Even cases with negative outcomes, when seen as breakdowns in adaptation, reveal the complicating dynamics that stress the textbook envelope and the often hidden sources of resilience used to cope with these complexities.

Accidents have been noted by many analysts as 'fundamentally surprising' events because they call into question the organization's model of the risks they face and the effectiveness of the countermeasure deployed (Lanir, 1986; Woods et al., 1994, chapter 5; Rochlin, 1999; Woods, 2005b). In other words, the organization is unable to recognize or interpret evidence of new vulnerabilities or ineffective countermeasures until a visible accident occurs. At this stage the organization can engage in fundamental learning but this window of opportunity comes at a high price and is fragile given the consequences of the harm and losses. The shift demanded following an accident is a reframing process. In reframing one notices initial signs that call into question ongoing models, plans and routines, and begins processes of inquiry to test if revision is warranted (Klein et al., 2005). Resilience Engineering aims to provide support for the cognitive processes of reframing an organization's model of how safety is created before accidents occur by developing measures and indicators of contributors to resilience such as the properties of buffers, flexibility, precariousness, and tolerance and patterns of interactions across scales such as responsibility-authority double binds.

Monitoring resilience is monitoring for the changing boundary conditions of the textbook competence envelope – how a system is competent at handling designed-for-uncertainties – to recognize forms of unanticipated perturbations – dynamics that challenge or go beyond the envelope. This is a kind of broadening check that identifies when the organization needs to learn and change. Resilience engineering needs to identify the classes of dynamics that undermine resilience and result in organizations that act riskier than they realize. This chapter focuses on dynamics related to safety-production goal conflicts.

Coping with Pressure to be Faster, Better, Cheaper

Consider recent NASA experience, in particular, the consequences of NASA's adoption of a policy called 'faster, better, cheaper' (FBC). Several years later a series of mishaps in space science missions rocked the organization and called into question that policy. In a remarkable 'organizational accident' report, an independent team investigated the organizational factors that spawned the set of mishaps (Spear, 2000).

The investigation realized that FBC was not a policy choice, but the acknowledgement that the organization was under fundamental

pressure from stakeholders. The report and the follow-up, but short-lived, 'Design for Safety' program noted that NASA had to cope with a changing environment with increasing performance demands combined with reduced resources: drive down the cost of launches, meet shorter, more aggressive mission schedules, do work in a new organizational structure that required people to shift roles and coordinate with new partners, eroding levels of personnel experience and skills. Plus, all of these changes were occurring against a backdrop of heightened public and congressional interest that threatened the viability of the space program. The MCO investigation board concluded: NASA, which had a history of 'successfully carrying out some of the most challenging and complex engineering tasks ever faced by this nation,' was being asked to 'sustain this level of success while continually cutting costs, personnel and development time … these demands have stressed the system to the limit' due to 'insufficient time to reflect on unintended consequences of day-to-day decisions, insufficient time and workforce available to provide the levels of checks and balances normally found, breakdowns in inter-group communications, too much emphasis on cost and schedule reduction.' The MCO Board diagnosed the mishaps as indicators of an increasingly brittle system as production pressure eroded sources of resilience and led to decisions that were riskier than anyone wanted or realized. Given this diagnosis, the Board went on to re-conceptualize the issue as how to provide tools for proactively monitoring and managing project risk throughout a project life-cycle and how to use these tools to balance safety with the pressure to be faster, better, cheaper.

The experience of NASA under FBC is an example of the law of stretched systems: every system is stretched to operate at its capacity; as soon as there is some improvement, for example in the form of new technology, it will be exploited to achieve a new intensity and tempo of activity (Woods, 2003). Under pressure from performance and efficiency demands (FBC pressure), advances are consumed to ask operational personnel 'to do more, do it faster or do it in more complex ways', as the Mars Climate Orbiter Mishap Investigation Board report determined. With or without cheerleading from prestigious groups, pressures to be 'faster, better, cheaper' increase. Furthermore, pressures to be 'faster, better, cheaper' introduce changes, some of which are new capabilities (the term does include 'better'), and these changes modify the vulnerabilities or paths toward failure. How conflicts and trade-offs

like these are recognized and handled in the context of vectors of change is an important aspect of managing resilience.

Balancing Acute and Chronic Goals

Problems in the US healthcare delivery system provide another informative case where faster, better, cheaper pressures conflict with safety and other chronic goals. The Institute of Medicine in a calculated strategy to guide national improvements in health care delivery conducted a series of assessments. One of these, Crossing the Quality Chasm: A New Health System for the 21st Century (IOM, 2001), stated six goals needed to be achieved simultaneously: the national health care system should be – Safe, Effective, Patient-centered, Timely, Efficient, Equitable.[1] Each goal is worthy and generates thunderous agreement. The next step seems quite direct and obvious – how to identify and implement quick steps to advance each goal (the classic search for so-called 'low hanging fruit'). But as in the NASA case, this set of goals is not a new policy direction but rather an acknowledgement of demanding pressures already operating on health care practitioners and organizations. Even more difficult, the six goals represent a set of interacting and often conflicting pressures so that in adapting to reach

---

[1] The IOM states the quality goals as –
'Health Care Should Be:

- Safe – avoiding injuries to patients from the care that is intended to help them.
- Effective – providing services based on scientific knowledge to all who could benefit and refraining from providing services to those not likely to benefit (avoiding underuse and overuse, respectively).
- Patient-centered – providing care that is respectful of and responsive to individual patient preferences, needs, and values and ensuring that patient values guide all clinical decisions.
- Timely – reducing waits and sometimes harmful delays for both those who receive and those who give care.
- Efficient – avoiding waste, including waste of equipment, supplies, ideas, and energy.
- Equitable – providing care that does not vary in quality because of personal characteristics such as gender, ethnicity, geographic location, and socioeconomic status.'

for one of these goals it is very easy to undermine or squeeze others. To improve on all simultaneously is quite tricky.

As I have worked on safety in health care, I hear many highly placed voices for change express a basic belief that these six goals can be synergistic. Their agenda is to energize a search for and adoption of specific mechanisms that simultaneously advance multiple goals within the six and that do not conflict with others – 'silver bullets'. For example, much of the patient safety discussion in US health care continues to be a search for specific mechanisms that appear to simultaneously save money and reduce injuries as a result of care. Similarly, NASA senior leaders thought that including 'better' along with faster and cheaper meant that techniques were available to achieve progress on being faster, better, and cheaper together (for almost comic rationalizations of 'faster, better, cheaper' following the series of Mars science mission mishaps and an attempt to protect the reputation of the NASA administrator at the time, see Spear, 2000). The IOM and NASA senior management believed that quality improvements began with the search for these 'silver bullet' mechanisms (sometimes called 'best practices' in health care). Once such practices are identified, the question becomes how to get practitioners and organizations to adopt these practices. Other fields can help provide the means to develop and document new best practices by describing successes from other industries (health care frequently uses aviation and space efforts to justify similar programs in health care organizations). The IOM in particular has had a public strategy to generate this set of silver bullet practices and accompanying justifications (like creating a quality catalog) and then pressure health care delivery decision makers to adopt them all in the firm belief that, as a result, all six goals will be advanced simultaneously and all stakeholders and participants will benefit (one example is computerized physician order entry).

However, the findings of the Columbia accident investigation board (CAIB) report should reveal to all that the silver bullet strategy is a mirage. The heart of the matter is not silver bullets that eliminate conflicts across goals, but developing new mechanisms that balance the inherent tensions and trade-offs across these goals (Woods et al., 1994). The general trade-off occurs between the family of acute goals – timely, efficient, effective (or after NASA's policy, the Faster, Better, Cheaper or FBC goals) and the family of chronic goals, for the health care case consisting of safety, patient-centeredness, and equitable access.

The tension between acute production goals and chronic safety risks is seen dramatically in the Columbia accident which the investigation board found was the result of pressure on acute goals eroding attention, energy and investments on chronic goals related to controlling safety risks (Gehman, 2003). Hollnagel (2004, p. 160) compactly captured the tension between the two sets of goals with the comment that:

> If anything is unreasonable, it is the requirement to be both efficient and thorough at the same time – or rather to be thorough when with hindsight it was wrong to be efficient.

The FBC goal set is acute in the sense that they happen in the short term and can be assessed through pointed data collection that aggregates element counts (shorter hospitals stays, delay times). Note that 'better' is in this set, though better in this family means increasing capabilities in a focused or narrow way, e.g., cardiac patients are treated more consistently with a standard protocol. The development of new therapies and diagnostic capabilities belongs in the acute sense of 'better.'

Safety, access, patient-centeredness are chronic goals in the sense that they are system properties that emerge from the interaction of elements in the system and play out over longer time frames. For example, safety is an emergent system property, arising in the interactions across components, subsystems, software, organizations, and human behavior.

By focusing on the tensions across the two sets, we can better see the current situation in health care. It seems to be lurching from crisis to crisis as efforts to improve or respond in one area are accompanied by new tensions at the intersections of other goals (or the tensions are there all along and the visible crisis point shifts as stakeholders and the press shift their attention to different manifestations of the underlying conflicts). The tensions and trade-offs are seen when improvements or investments in one area contribute to greater squeezes in another area. The conflicts are stirred by the changing background of capabilities and economic pressure. The shifting points of crisis can be seen first in 1995-6 as dramatic well publicized deaths due to care helped create the patient safety crisis (ultimately documented in Kohn et al., 1999). The patient safety movement was energized by patients feeling vulnerable as