Creating Foresight:
Lessons for Enhancing Resilience from Columbia

David Woods
Cognitive Systems Engineering Laboratory
Institute for Ergonomics
The Ohio State University

12-14-04
Draft

To appear in
Learning from the Columbia Accident
Bill Starbuck and Moshe Farjoun (Editors)
Blackwell

The past seems incredible, the future implausible.[1]

Introduction

To look forward and envision how organizations can achieve very high reliability and resilience, one first must look back with clarity unobscured by hindsight bias. The Columbia accident, as a highly visible event investigated in depth by a distinguished and independent panel, provides an opportunity to review generic patterns seen across multiple accidents and across studies in multiple fields of practice (Hollnagel, 1993). This chapter examines patterns present in the Columbia accident (STS-107) in order to consider how organizations in general can learn and change *before* dramatic failures occur.

From the point of view of learning and change, the Columbia accident investigation is important because the independent investigating board (CAIB) found the hole in the wing of the shuttle was produced not simply by debris, but by holes in organizational decision making. Furthermore, the factors that produced the holes in this organization's decision making are not unique to today's NASA or limited to the Shuttle program, but are generic vulnerabilities that have contributed to other failures and tragedies across other complex industrial settings.

CAIB's investigation revealed how NASA failed to balance safety risks with intense production pressure. As a result, this accident matches a classic pattern—a drift

[1] Woods and Cook, 2002.

toward failure as defenses erode in the face of production pressure. When this pattern is combined with a fragmented distributed problem solving process that is missing cross checks and unable to see the big picture, the result is an organization that cannot see its own blind spots about risks. Further, NASA was unable to revise its assessment of the risks it faced and the effectiveness of its countermeasures against those risks as new evidence accumulated. What makes safety/production tradeoffs so insidious is that evidence of risks become invisible to people working hard to produce under pressure so that safety margins erode over time.

As an organizational accident Columbia shows the need for organizations to monitor their own practices and decision processes to detect when they are beginning to drift toward safety boundaries. The critical role for the safety group within the organization is to monitor the organization itself—to measure organizational risk—the risk that the organization is operating nearer to safety boundaries than it realizes. This process of monitoring the organization's model is an important part of the emerging research on how to help organizations monitor and manage resilience (e.g., Woods and Shattuck, 2000; Cook et al., 2000; Sutcliffe and Vogel, 2003; and see Brown, 2005 for examples of breakdowns in resilience).

In studying tragedies such as Columbia, experience indicates that failure challenges organizations' model of how they are vulnerable to failure and thus creates windows for rapid learning and improvement (Lanir, 1986; Woods et al., 1994, chapter 6). Seizing the opportunity to learn is the responsibility leaders owe to the people and families whose sacrifice and suffering was required to make the holes in the organization's decision making visible to all. Just as Columbia led NASA and Congress to begin to transform the culture and operation of all of NASA, the generic patterns can serve as lessons for transformation of other high risk operations, before failures occur.

This is the topic of the newly emerging field of Resilience Engineering and Management which uses the insights from research on failures in complex systems, especially the organizational contributors to risk, and the factors that affect human performance to provide practical systems engineering tools to manage risk proactively (Hollnagel, Leveson and Woods, 2005). Organizations can use the emerging techniques of Resilience Engineering to balance the competing demands for very high safety with real time pressures for efficiency and production. NASA, as it follows through on the recommendations of the Columbia Accident Investigation Board (CAIB), will serve as a model for others on how to thoroughly re-design a safety organization and provide for independent technical voices in organizational decision making.

Escaping Hindsight

Hindsight bias is a psychological effect that leads people to misinterpret the conclusions of accident investigations.[2] Often the first question people ask about the

[2] The hindsight bias is a well reproduced research finding 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 decision making leading up to an accident such as Columbia takes the form of: “why did NASA continue flying the Shuttle with a known problem…?” (the ‘known problem’ refers to the dangers of debris striking and damaging the Shuttle wing during takeoff which the CAIB identified as the physical, proximal cause of the accident.)

As soon as the question is posed in this way, it is easy to be trapped into oversimplifying the situation and the uncertainties involved before the outcome is known (Dekker, 2002). After-the-fact “the past seems incredible,” hence NASA managers sound irrational or negligent in their approach to obvious risks. However, before any accident has occurred and while the organization is under pressure to meet schedule or increase efficiency, potential warning flags are overlooked or re-interpreted since those potential “futures look implausible.” For example, the signs of Shuttle tile damage became an issue of orbiter turn around time and not a flight risk. Because it is difficult to disregard knowledge of outcome, it is easy to play the classic blame game, define a “bad” individual, group, or organization as the culprit, and stop. When this occurs, the same difficulties that led to the Columbia accident will go unrecognized in other programs and in other organizations.

Interestingly, the CAIB worked hard to overcome hindsight bias and uncover the breakdown in organizational decision making that led to the accident. All organizations can misbalance safety risks with pressure for efficiency. It is difficult to sacrifice today’s real production goals to consider uncertain evidence of possible future risks. The heart of the difficulty is that it is most critical to invest resources to follow up on potential safety risks when the organization is least able to afford the diversion of resources due to pressure for efficiency or throughput.

To escape hindsight bias in understanding how a specific case of drift toward failure developed, one charts the evolution of the mindset of the groups involved (Woods et al., 1994). Dekker’s 2002 book, “The Field Guide to Human Error Investigations,” provides a basic guide on how to carry out this analysis process. To set the stage for a discussion of the generic patterns present in the lead up to the Columbia tragedy and the implications of these general patterns for the future, this section examines several critical points in the evolution of mindset prior to STS-107.[3] The Board’s analysis reveals shift points where opportunities to re-direct the evolution away from failure boundaries were missed. Identifying these points and the contributing factors highlights several basic patterns of failure that have been abstracted from past

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 judgment.

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.

[3] The discussion is based only on the material available in chapters 6 to 8 of the CAIB report; charting the evolution of mindset across the teams identifies areas where further information would be very valuable.

accidents and studies. These generic patterns provide insights guide organizational change (see Hollnagel, 1993 for the general concept and Woods and Shattuck, 2000 or Patterson et al., 2004 for examples of how analysis of accidents can reveal a general pattern in distributed cognition).

Charting the Drift toward Failure: I Foam events are not in-flight anomalies.

To start charting the evolution of mindset in this accident, consider how different groups evaluated foam events against a backdrop of the risks of various kinds of debris strikes, including the risks of damage to different structures on Shuttle. The data available in the CAIB report helps us see several points where the evaluation of these risks shifted or could have shifted.

Shift 1:

The first critical shift is the re-classification of foam events from in-flight anomalies to maintenance and turn around issues (STS-113 Flight Readiness Review, CAIB, p. 125-126). Closely related is the shift to see foam loss as an accepted risk or even as one pre-launch briefing put it -- “not a safety of flight issue” (CAIB, p. 126 1st column to top of 2nd column).

This shift in the status of foam events is a critical part of explaining the limited and fragmented evaluation of the STS-107 foam strike and how analysis of that foam event never reached the problem solving groups that were practiced at investigating anomalies, their significance and consequences, i.e., mission control (Patterson et al., 1999).

The data collected by the CAIB imply several contributors to the change in status of foam events:
(a) Pressure on schedule issues produced a mindset centered on production goals (CAIB, p. 125, 2nd column). There are several way that this could have played a role: first, schedule pressure magnifies the importance of activities that affect turnaround; second, when events are classified as in-flight anomalies a variety of formal work steps and checks are invoked; third, the work to assess anomalies diverts resources from the tasks to be accomplished to meet turn around pressures.
(b) A breakdown or absence of cross-checks on the rationale for classifying previous foam loss events as **not** an in-flight safety issue (CAIB, p. 125, Figure 6.1-5; 126 top). In fact the rationale for the re-classification was quite thin, weak and flawed. The CAIB’s examination reveals that no cross checks were in place to detect, question or challenge the specific flaws in the rationale.
(c) The use of what on the surface looked on the surface like technical analyses to justify previously reached conclusions, rather than using technical analyses to test tentative hypotheses (CAIB, p. 126 1st column).

It would be very important to know more about what the mindset and stance of different groups toward this shift in classification. For example, one would want to consider: Was the shift due to the salience of the need to improve maintenance and turn around? Was this a organizational structure issue (which organization focuses on what aspects of problems)? What was mission control’s reaction to the re-classification? Was it heard about by other groups? Did reactions to this shift remain underground relative to formal channels of communication?

Interestingly, the organization had in principle 3 categories of risk: in flight anomalies, accepted risks, and non-safety issues. As the organization began to view foam events as an accepted risk, there was no formal means for follow up with a re-evaluation of an "accepted" risk to assess if it is in fact acceptable as new evidence builds up or as situations change. For all practical purposes, there was no difference in how the organization was handling non-safety issues and accepted risks (i.e., accepted risks were being thought of and acted on no differently than non-safety issues). Yet the organization acted as if items placed in the accepted risk category were being evaluated and handled appropriately (i.e., as if the assessment of the hazard was accurate and up-to-date and as if the countermeasures deployed were still shown to be effective).

Shift 2:
Another component in the drift process is the interpretation of past "success" (this doesn't occur at any one point but is a general background to the evolution prior to the launch). The absence of failure is taken as positive indication that hazards are not present or that countermeasures are effective. In this context, it is very difficult to gather or see if evidence is building up that should trigger a re-evaluation and revision of the organization's model of vulnerabilities.

If an organization is only able to change its model of itself unless and until completely clear cut evidence accumulates, that organization will tend to learn late, i.e., revise its model of vulnerabilities, only after serious events occur. On the other hand, learning organizations assume their model of risks and countermeasures is fragile and even seek out evidence about the need to revise and update this model (Rochlin, 1999). They do not assume their model is correct and then wait for evidence to come to their attention for to do so will guarantee an organization that acts riskier than it desires.

Feynman's famous appendix in the Challenger accident report captures how relying on past success blocks perception of warning signs, and changes what even counts as a warning, before outcome is known (see also Weick, Sutcliffe, and Obstfeld, 1999). Consider how the larger organization and other stakeholders would react if, prior to an accident and when the organization is under acute pressure to achieve production goals, a group watching for warning signs decides to sacrifice a tangible acute production goal to invest time, money and energy of personnel in an issue that <u>might</u> contribute to increased risk. I daresay most organizations do not reward such monitoring for warnings and decisions to sacrifice schedule/cost without very strong evidence that the sacrifice is necessary. Yet such behavior guarantees such organizations are acting much riskier than they claim or want to be operating.

Shift 3:
Several opportunities to revise the status of foam events, the hazard they represented, and how these events were to be handled in-flight, occurred prior to the launch of STS-107. These missed opportunities are represented by the damage suffered on shuttle flights STS-27R, STS-45 (and similarly on other flights).

Foam events are only one source of debris strikes that threaten different aspects of the orbiter structure. Debris strikes carry very different risks depending on where and what they strike. The hinge in considering the response to the foam strike on STS-107 is that the debris struck the leading edge structure (RCC panels and seals) and not the tiles. Did concern and progress on improving tiles block the ability to see risks to other structures? Did NASA regard the leading edge as much less vulnerable to damage than tiles (e.g., CAIB, p. 31; memo on p. 141; p. 145 paragraph 3)? Chapter 6 of the CAIB report only provides a few tantalizing cues about how various groups regarded the vulnerability of leading edge structures.

This is important because the damage in STS-45 provided an opportunity to focus on the leading edge structure and re-consider the margins to failure of that structure given strikes by various kinds of debris. Did this mission create a sense that the leading edge structure was less vulnerable than tiles? Did this mission fail to revise a widely held belief that the RCC leading edge panels were more robust to debris strikes than they really are (e.g., CAIB, p. 145)? Who followed up the damage to RCC panel and what did they conclude? Who received the results? How were risks to non-tile structures evaluated and considered – including landing gear door structures? More information about the follow up to leading edge damage in STS-45 would shed light on how this opportunity was missed.

Charting the Drift toward Failure: II An anomaly in limbo.

Once the foam strike was detected on the launch of Columbia in STS-107, a variety or groups played a role in the evaluation—and lack of evaluation—of this anomaly. This is an example of a problem solving process distributed over interacting groups, or more generally, an example of distributed cognition which has been studies in related settings (Hutchins, 1995) and, interestingly, specifically in NASA shuttle mission control (Patterson et al., 1999; Patterson and Woods, 2001; Watts et al., 1996; Chow et al., 2000). This section makes a few observations about distributed cognition in the case of STS-107 relative to the general research findings.

A management *stance* emerged early which downplayed significance of the strike. The initial and very preliminary assessments of the foam strike created a stance toward further analysis that this was not a critical or important issue for the mission. The stance developed and took hold before there were results from any technical analyses. This indicates that preliminary judgments were biasing data evaluation, instead of following a proper engineering evaluation process where data evaluation points teams and management to conclusions.

Indications that the event was outside of boundary conditions for NASA's understanding of the risks of debris strikes seemed to go unrecognized (CAIB, p. 143, notes the limits of the modeling tool Crater with respect to the analysis needed; also p. 160 bottom). When events fall outside of boundaries of past data and analysis tools and when the data available includes large uncertainties, the event is by definition anomalous and of high risk. While personnel noted the specific indications in themselves, no one was able to use these indicators to trigger any deeper or wider recognition of the nature of the anomaly in this situation (for example, the email on CAIB

p. 151-152). This pattern of seeing the details but being unable to recognize the big picture is commonplace in accidents (Woods et al., 1987).

As the Debris Assessment Team was formed after the strike was detected and began to work, the question arose: Is the size of the debris strike “out-of-family” or “in-family” given past experience? While the team looked at past experience, they were unable to get a consistent or informative read on how past events indicated risk for this event. It appears no other groups or representatives of other technical areas were brought into the picture. This absence of any cross-checks is quite notable and inconsistent with how mission control groups evaluate in-flight anomalies (e.g., Watts et al., 1996). Past studies indicate that a review or interaction with another group would have provided broadening checks which help uncover inconsistencies and gaps as people need to focus their analysis, conclusions and justifications for consideration and discussion with others.

Evidence that the strike posed risk of serious damage kept being encountered—RCC panel impacts at angles greater than 15 degrees predicted coating penetration (CAIB, p. 145), foam piece 600 times larger than ice debris previously analyzed (CAIB, p. 143), models predicting tile damage deeper than tile thickness (CAIB, p. 143). Yet, a process of discounting evidence discrepant with the current assessment went on several times (though eventually the Debris Assessment Team concerns seem to focus on the landing gear doors rather than the leading edge structure).

Given the concerns about potential damage that arose in the Debris Assessment Team and given their desire to determine the location more definitively, the question arises: did the team conduct contingency analyses of damage and consequences across the different candidates sites—leading edge, landing gear door seals, tiles? Based on the evidence compiled in the CAIB report, there was no contingency analysis or follow through on the consequences if the leading edge structure (RCC) was the site damaged. This is quite puzzling as this was the team’s first assessment of location and in hindsight there initial estimate proved to be reasonably accurate.

This lack of follow through coupled with the Debris Assessment Team’s growing concerns about the landing gear door seals (e.g., the unsent email on p. 157, CAIB; email p. 163, CAIB) seems to indicate that they may have viewed the leading edge structures as more robust to strikes than other orbiter structures. The CAIB report fails to provide critical information about how different groups viewed the robustness or vulnerability of the leading edge structure to damage from debris strikes (of course, post-accident these beliefs can be quite hard to determine, but various memos/analyses may indicate more about the perception risks to this part of the orbiter). Insufficient data is available to understand why was RCC damage under-pursued by the Debris Assessment Team?

What is striking is how there was a fragmented view of what was known about the strike and its potential implications over time, people and groups. There was no place, artifact, or person who had a complete and coherent view of the analysis of the foam strike event (note a coherent view includes understanding the gaps and uncertainties in the data or analysis to that point). This contrasts dramatically with how mission control works to investigate and handle anomalies where there are clear lines of responsibility to have a complete, coherent view of the evolving analysis vested in the relevant flight controllers and in the flight director. Mission control has mechanisms to keep different people in the loop (via monitoring voice loops, for example) so that all are up to date on the current picture of situation. Mission control also has mechanisms for correcting assessments as analysis proceeds, whereas in this case, the fragmentation and partial views seemed to block re-assessment and freeze the organization on an erroneous assessment (for studies of distributed cognition during anomalies in mission control see Patterson et al., 1999; Watts et al., 1996; Patterson and Woods, 2001; Chow et al., 2000).

As Debris Assessment Team worked at the margins of knowledge and data, their partial assessments did not benefit from cross-checks through interactions with other technical groups with different background and assumptions. There is no report of a technical review process that accompanied their work. Interactions with people or groups with different knowledge and assumptions is one of the best ways to improve assessments and to aid revision of assessments. Mission control anomaly response includes many opportunities for cross-checks to occur. In general, it is quite remarkable that the groups practiced at anomaly response—mission control—never became involved in the process.

The process of analyzing the foam strike by the Debris Assessment Team broke down in many ways. The fact that this group also advocated steps that we now know would have been valuable (the request for imagery to locate the site of the foam strike) leads us to miss the generally fragmented distributed problem solving process. The fragmentation also occurred across organizational levels (Debris Assessment Team to Mission Management Team or MMT). Effective collaborative problem solving requires more direct participation by members of the analysis team in the overall decision making process. This is not sufficient of course; for example, the MMT's stance already defined the situation as, 'show me that the foam strike is an issue' rather than 'convince me the anomaly requires no response or contingencies.'

Overall, the evidence points to a broken distributed problem solving process—playing out in between organizational boundaries. The fragmentation in this case indicates the need for a senior technical focal point to integrate and guide the anomaly analysis process (e.g., the flight director role). And this role requires real authority. The MMT and the MMT chair were in principle in a position to supply this role, but:

~ Was the MMT practiced at providing the integrative problem solving role?

~ Were there other cases where significant analysis for in flight anomalies was guided by the MMT or were they all handled by the mission control team?

The problem solving process in this case has the odd quality of being stuck in limbo. Not dismissed or discounted away completely, yet unable to get traction as an in-flight anomaly to be thoroughly investigated with contingency analyses and re-planning activities. The dynamic appears to be a management stance that puts the event outside of safety of flight (e.g., conclusions drove, or eliminated, the need for analysis and investigation, rather than investigations building the evidence from which one would draw conclusions). Plus, the Debris Assessment Team exhibited a fragmented problem solving process that failed to integrate partial and uncertain data to generate a big picture—i.e., the situation was outside the understood risk boundaries and carried significant uncertainties.

Five General Patterns Present in Columbia

Based on the material and analyses in the CAIB report, there are five classic patterns (Hollnagel, 1993) also seen in other accidents and research results:

- Drift toward failure as defenses erode in the face of production pressure.
- An organization that takes past success as a reason for confidence instead of investing in anticipating the changing potential for failure.
- Fragmented distributed problem solving process that clouds the big picture.
- Failure to revise assessments as new evidence accumulates.
- Breakdowns at the boundaries of organizational units that impedes communication and coordination.

1. The basic classic pattern in this accident is—*Drift toward failure as defenses erode in the face of production pressure*.

My colleague, Erik Hollnagel in 2002, captured the heart of the Columbia accident when he commented on other accidents:

> 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.

Hindsight bias, by oversimplifying the situation people face before outcome is known, often hides tradeoffs between multiple goals (Woods et al., 1994). The analysis in the CAIB report provides the general context of a tighter squeeze on production goals creating strong incentives to downplay schedule disruptions. With shrinking time/resources available, safety margins were likewise shrinking in ways which the organization couldn't see.

Goal tradeoffs often proceed gradually as pressure leads to a narrowing of focus on some goals while obscuring the tradeoff with other goals. This process usually happens when acute goals like production/efficiency take precedence over chronic goals like safety.

The dilemma of production/safety conflicts is: If organizations never sacrifice production pressure to follow up warning signs, they are acting much too risky. On the other hand, if uncertain "warning" signs always lead to sacrifices on acute goals, can the organization operate within reasonable parameters or stakeholder demands? It is precisely at points of intensifying production pressure that extra safety investments need to be made in the form or proactive searching for side effects of the production pressure and in the form or re-assessing the risk space—safety investments are most important when least affordable.

This generic pattern points toward several constructive issues:

• How does a safety organization monitor for drift and its associated signs, in particular, a means to recognize when the side effects of production pressure may be increasing safety risks?
• What indicators should be used to monitor the organization's model of itself, how it is vulnerable to failure, and the potential effectiveness of the countermeasures it has adopted.
• How does production pressure create or exacerbate tradeoffs between some goals and chronic concerns like safety?
• How can an organization add investments to safety issues at the very time when the organization is most squeezed. For example, how does an organization note a reduction in margins and follow through by re-building margin to boundary conditions in new ways?

2. Another general pattern identified in Columbia is that *an organization takes past success as a reason for confidence instead of digging deeper to see underlying risks*.
During the drift toward failure leading to the Columbia accident a mis-assessment took hold that resisted revision (that is, the mis-assessment that foam strikes pose only a maintenance and not a risk to orbiter safety). It is not simply that the assessment was wrong, but the inability to re-evaluate the assessment and re-examine evidence about the vulnerability that is troubling.

The missed opportunities to revise and update the organization's model of the riskiness of foam events seem to be consistent with what has been found in other cases of failure of foresight. Richard Cook and I have described this discounting of evidence as "distancing through differencing" whereby those reviewing new evidence or incidents focus on differences, real and imagined, between the place, people, organization and circumstances where an incident happens and their own context. By focusing on the differences, people see no lessons for their own operation and practices (or only extremely narrow well bounded responses). This contrast with the what has been noted about more effective safety organizations which proactively seek out evidence to revise and update this model despite the fact that this risks exposing the organization's blemishes (Rochlin, 1999; Woods, 2005).

Ominously, the *distancing through differencing* that occurred throughout the build up to the final Columbia mission can be repeated in the future as organizations and groups look at the analysis and lessons from this accident and the CAIB report. Others in the future can easily look at the CAIB conclusions and deny their relevance to their situation by emphasizing differences (e.g., my technical topic is different, my managers are different, we are more dedicated and careful about safety, we have already addressed that specific deficiency). This is one reason avoiding hindsight bias is so important—when one starts with the question, how could they have missed what is now obvious—one is enabling future distancing through differencing rationalizations.

The distancing through differencing process that contributes to this breakdown also indicates ways to change the organization to promote learning. One general principle which could be put into action is—do not discard other events because they appear on the surface to be dissimilar. At some level of analysis, all events are unique; while at other levels of analysis, they reveal common patterns. Every event, no matter how dissimilar on the surface, contains information about underlying general patterns that help create foresight about potential risks before failure or harm occurs. To focus on common patterns not surface differences requires shifting the analysis of cases from surface characteristics to deeper patterns and more abstract dimensions. Each kind of contributor to an event then can guide the search for similarities.

To step back more broadly, organizations need a mechanism to generate new evaluations that question the organization's own model of the risks it faces and the countermeasures deployed. Such review and re-assessment can help the organization find places where it has underestimated the potential for trouble and revise its approach to create safety. A quasi-independent group is needed to do this—independent enough to question the normal organizational decision making but involved enough to have a finger on the pulse of the organization (keeping statistics from afar is not enough to accomplish this).

3. Another general pattern identified in Columbia is a *fragmented problem solving process that clouds the big picture*.

During Columbia there was a fragmented view of what was known about the strike and its potential implications. People were making decisions about what did or did not pose a risk on very shaky or absent technical data and analysis, and critically, *they couldn't see their decisions rested on shaky grounds* (e.g., the memos on p. 141, 142 of he CAIB report illustrate the shallow, off hand assessments posing for and substituting for careful analysis).

There was no place or person who had a complete and coherent view of the analysis of the foam strike event including the gaps and uncertainties in the data or analysis to that point. It is striking that people used what looked like technical analyses to justify previously reached conclusions, instead of using technical analyses *to test tentative hypotheses* (e.g., CAIB report, p. 126 1st column).

The breakdown or absence of cross-checks is also striking. Cross checks on the rationale for decisions is a critical part of good organizational decision making. Yet no cross checks were in place to detect, question or challenge the specific flaws in the rationale, and *no one noted that cross-checks were missing*.

The breakdown in basic engineering judgment stands out as well. The initial evidence available already placed the situation outside the boundary conditions of engineering data and analysis. The only available analysis tool was not designed to predict under these conditions, the strike event was hundreds of times the scale of what the model is designed to handle, and the uncertainty bounds were very large with limited ability to reduce the uncertainty (email on p. 151-152 CAIB). Being outside the analyzed boundaries should not be confused with not being confident enough to provide definitive answers. In this situation basic engineering judgment calls for large efforts to extend analyses, find new sources of expertise, and cross-check results as mission control both practices and does.

Seasoned pilots and ship commanders well understand the need for this ability to capture the big picture and not to get lost in a series of details. The issue is how to train for this judgment. For example, the Flight Director and his or her team practice identifying and handling anomalies through simulated situations. Note that shrinking budgets lead to pressure to reduce training investments (the amount of practice, the quality of the simulated situations, and the number or breadth of people who go through the simulations sessions can all decline).

I particularly want to emphasize this point about making technical judgments technically. The decision makers did not seem able to notice when they needed more expertise, data, and analysis in order to have a proper evaluation of an issue. NASA's evaluation prior to STS 107 that foam debris strikes do not pose risks of damage to the orbiter demands a technical base. Instead their "resolution" was based on very shaky or absent technical grounds, often with shallow, off hand assessments posing for and substituting for careful analysis (e.g., the memos on p. 141, 142).

The fragmentation of problem solving also illustrates Karl Weick's points about how effective organizations exhibit a "deference to expertise", "reluctance to simplify interpretations", and "preoccupation with potential for failure" none of which were in operation in NASA's organizational decision making leading up to and during Columbia (Weick, Sutcliffe, and Obstfeld, 1999).

The lessons of Columbia should lead organizations of the future to have a safety organization that ensures that adequate technical grounds are established and used in organizational decision making. To accomplish this, in part, the safety organization will need to define the kinds of anomalies to be practiced as well as who should participates in those simulation training sessions. The value of such training depends critically on designing a diverse set of anomalous scenarios with detailed attention to how they unfold. By monitoring performance in these simulated training cases, the safety personnel will be better able to assess the quality of decision making across levels in the organization.

4. The fourth pattern in Columbia is a *Failure to revise assessments as new evidence accumulates*.

I first studied this pattern in nuclear power emergencies 20 plus years ago (Woods et al., 1987). What was interesting in the data then was how difficult it is to revise a mis-assessment or to revise a once plausible assessment as new evidence comes in. This finding has been reinforced in subsequent studies in different settings (Johnson et al., 1991; Feltovich et al., 1997).

Research consistently shows that revising assessments successfully requires a new way of looking at previous facts. We provide this **"fresh"** view:
(a) by bringing in people new to the situation
(b) through interactions across diverse groups with diverse knowledge and tools,
(c) through new visualizations which capture the big picture and re-organize data into different perspectives.

One constructive action is to develop the collaborative inter-changes that generate fresh points of view or that produce challenges to basic assumptions. This cross checking process is an important part of how NASA mission control and other organizations successfully respond to anomalies (for a case where these processes break down see Patterson et al., 2004). One can also capture and display indicators of safety margin to help people see when circumstances or organizational decisions are pushing the system closer to the edge of the safety envelope. (this idea is something that Jens Rasmussen one of the pioneers of the new results on error and organizations has been pushing for two decades, e.g., Rasmussen, 1990; Rasmussen et al., 1994).

The crux is to notice the information that changes past models of risk and calls into question the effectiveness of previous risk reduction actions, without having to wait for completely clear cut evidence. If revision only occurs when evidence is overwhelming, there is a grave risk of an organization acting too risky and finding out only from near misses, serious incidents, or even actual harm. Instead, the practice of revising assessments of risks needs to be an ongoing process. In this process of continuing re-evaluation, the working assumption is that risks are changing or evidence of risks has been missed.

What is particularly disappointing about NASA's organizational decision making is that the correct diagnosis of production/safety tradeoffs and useful recommendations for organizational change were noted in 2000. The Mars Climate Orbiter report of March 13, 2000 clearly depicts how the pressure for production and to be 'better' on several dimensions led to management accepting riskier and riskier decisions. This report recommended many organizational changes similar to the CAIB. A slow and weak response to the previous independent board report was a missed opportunity to improve organizational decision making in NASA.

The lessons of Columbia should lead organizations of the future to develop a safety organization that provides "fresh" views on risks to help discover the parent organization's own blind spots and question its conventional assumptions about safety risks.

5. Finally, the Columbia accident brings to the fore another pattern: *Breakdowns at the boundaries of organizational units.*

The CAIB analysis notes how a kind of catch 22 was operating in which the people charged to analyze the anomaly were unable to generate any definitive traction and in which the management was trapped in a stance shaped by production pressure that views such events as turn around issues. This effect of an '*anomaly in limbo*' seems to emerge at the boundaries of different organizations that do not have mechanisms for constructive interplay. It is here that we see the operation of the generalization that in risky judgments we have to defer to those with technical expertise and the necessity to set up a problem solving process that engages those practiced at recognizing anomalies in the event.

This pattern points to the need for mechanisms that create effective overlap across different organizational units and to avoid simply staying inside the chain of command mentality (though such overlap can be seen as inefficient when the organization is under severe cost pressure).

This issue is of particular concern to many organizations as communication technology has linked together disparate groups as a distributed team. This capability for connectivity is leading many to work on how to support effective coordination across these distributed groups, e.g., in military command and control (Klein et al., in press).

The lessons of Columbia should lead organizations of the future to develop a safety organization with the technical expertise and authority to enhance coordination across the normal chain of command.

Managing Resilience in Organizations

The insights derived from the above five patterns and other research results on safety in complex systems point to the need to monitor and manage risk continuously throughout the life cycle of a system, and in particular to find ways of maintain a balance between safety and the often considerable pressures to meet production and efficiency goals (Reason, 1997; Weick, et al., 1999; Adamski and Westrum, 2003). These results indicate that safety management in complex systems should focus on resilience—the ability to adapt or absorb disturbance, disruption and change. A systems' resilience captures the result that failures are breakdowns in the normal adaptive processes necessary to cope with the complexity of the real world (Rasmussen, 1990; Rasmussen et al., 1994; Woods and Cook, 2003; Sutcliffe and Vogel, 2003).

A system's resilience includes 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*—the system's ability to restructure itself in response to external changes or pressures, *margin*—how close the system is currently operating relative to one or another kind of performance boundary, *tolerance*—does the system gracefully degrade as stress/pressure increase or collapse quickly when pressure exceeds adaptive capacity. Cross-scale interactions are another important factor as the resilience of a system defined at one scale depends on influences from scales above and below: downward in terms of how organizational context creates pressures/goal conflicts/dilemmas and upward in terms of how adaptations by local actors in the form of workarounds or innovative tactics reverberate and influence more strategic issues.

Managing resilience, or Resilience Engineering, then focuses on what sustains or erodes the adaptive capacities of human-technical system in a changing environment (Hollnagel et al., 2005). The focus is monitoring organizational decision making to assess the risk that the organization is operating nearer to safety boundaries than it realizes (or more generally, that the organization's adaptive capacity is degrading or lower than the adaptive demands of its environment).

To put it in terms of the basic failure pattern evident in the Columbia accident—managing an organization's resilience is concerned with assessing the risk that holes in organizational decision making will produce unrecognized drift toward failure boundaries, or monitoring for risks in how the organization monitors its risks. Resilience Engineering seeks to develop engineering and management practices to measure of sources of resilience, provide decision support for balancing production/safety tradeoffs, and create feedback loops that enhances the organization's ability to monitor/revise risk models and to target safety investments (e.g., Cook et al., 2000; Carthey et al., 2001; Woods and Shattuck, 2000; Hollnagel, 2004). For example, Resilience Engineering would monitor evidence that effective cross checks are well-integrated when risky decisions are made or would serve as a check on how well the organization prepares to handle anomalies by checking on how it practices handling of simulated anomalies (what kind of anomalies, who is involved in making decisions).

The focus on system resilience emphasizes the need for proactive measures in safety management—tools to support *agile, targeted, and timely* investments to *defuse emerging vulnerabilities* and sources of risk before harm occurs.

Sacrifice Decisions

To achieve resilience organizations need support for decisions about production/safety tradeoffs. Resilience engineering should help organizations decide when to relax production pressure to reduce risk, or, in other words, develop tools to support sacrifice decisions across production/safety tradeoffs.

When operating under production and efficiency pressures, evidence of increased risk on safety may be missed or discounted. As a result, organizations act in ways that are riskier than they realize or want, until an accident or failure occurs. This is one of the factors that creates the drift toward failure signature in complex system breakdowns.

To make risk a proactive part of management decision-making requires ways to know when to relax the pressure on throughput and efficiency goals, i.e., making a *sacrifice* decision—how to help organizations decide when to relax production pressure to reduce risk (Woods, 2000b). I refer to these tradeoff decisions as sacrifice judgments because acute production or efficiency related goals are temporarily sacrificed, or the pressure to achieve these goals relaxed, in order to reduce risks of approaching too near safety boundary conditions. Sacrifice judgments occur in many settings: when to convert from laparoscopic surgery to an open procedure (e.g., Cook et al., 1998), when to break off an approach to an airport during weather that increases the risks of wind shear, and when to have a local slowdown in production operations to avoid risks as complications build up.

New research is needed to understand this judgment process in organizations. Indications from previous research on such decisions (e.g., production/safety tradeoff decisions in lapraroscopic surgery) is that the decision to value production over safety is implicit and unrecognized. The result is that individuals and organizations act much riskier than they would ever desire. A sacrifice judgment is especially difficult because the hindsight view will indicate that the sacrifice or relaxation may have been unnecessary since “nothing happened.” This means that it is important to assess how peers and superiors react to such decisions.

The goal is to develop explicit guidance on how to help people make the relaxation/sacrifice judgment under uncertainty, to maintain a desired level of risk acceptance/risk averseness, and to recognize changing levels of risk acceptance/risk averseness. For example, what indicators reveal a safety/production tradeoff sliding out of balance as pressure rises to achieve acute production and efficiency goals. Ironically, it is at these very times of higher organizational tempo and focus on acute goals that require extra investments in sources of resilience to keep production/safety tradeoffs in balance—valuing thoroughness despite the potential for sacrifices on efficiency required to meet stakeholder demands.

An Independent, Involved, Informed, and Informative Safety Organization

While NASA failed to make the production/safety tradeoff reasonably in the context of foam strikes, the question for the future is how to help organizations make these tradeoffs better? It is not enough to have a safety organization; safety has to be part of making every day management decisions by actively re-considering and revising models of risks and assessments of the effectiveness of countermeasures. As Feynman also noted in his minority report on Challenger, a high risk, high performance organization must put the technical reality above all else including production pressure.

One traditional dilemma for safety organizations is the problem of "cold water and an empty gun." Safety organizations raise questions which stop progress on production goals—the "cold water." Yet when line organizations ask for help on how to address the safety concerns, while being responsive to production issues, the safety organization has little to contribute—the "empty gun." As a result, the safety organization fails to better balance the safety/production tradeoff in the long run. In the short run following a failure, the safety organization is emboldened to raise safety issues, but in the longer run the memory of the previous failure fades, production pressures dominate, and the drift processes operate unchecked (as has happened in NASA before Challenger, before Columbia, and could happen again with respect to space station).

From the point of view of managing resilience, a safety organization should monitor and balance the tradeoff of production pressure and risk. To do this the leadership team needs to implement a program for managing organizational risk—detecting emerging 'holes' in organizational decision making. As a result, a safety organization needs the resources and authority to achieve the "I's" of an effective safety organization -- independence, involvement, informed and informative:

- provide an *independent* voice that challenges conventional assumptions within senior management,
- constructive *involvement* in targeted but everyday organizational decision making (for example, ownership of technical standards, waiver granting, readiness reviews, and anomaly definition).
- actively generate *information* about how the organization is actually operating, especially to be able to gather accurate information about weaknesses in the organization (informed and informative).

Safety organizations must achieve independence enough to question the normal organizational decision making. At best the relationship between the safety organization and line senior management will be one of *constructive tension*. Inevitably, there will be periods where senior management tries to dominate the safety organization. The design of the organizational dynamics needs to provide the safety organization the tools to resist these predictable episodes by providing funding directly and independent from headquarters. Similarly, to achieve independence, the safety leadership team needs to be chosen and accountable outside of the normal chain of command.

Safety organizations must be involved in enough everyday organizational activities to have a finger on the pulse of the organization and to be seen as a constructive part of how the organization balances safety and production goals. This means the new safety organization needs to control a set of resources and have the authority to decide how to invest these resources to help line organizations provide high safety while accommodating production goals. For example, the safety organization could decide to invest and develop new anomaly response training programs when it detects holes in organizational decision making processes.

In general, safety organizations risk becoming information limited as they can be shunted aside from real organizational decisions, kept at a distance from the actual work processes, and kept busy tabulating irrelevant counts when their activities are seen as a threat by line management (for example, the 'cold water' problem). Independent, involved, informed and informative—these properties of an effective safety organization are closely connected, mutually reinforcing and difficult to achieve in practice.

Conclusion

Researchers on organizations and safety are not simply commentators on the sidelines, but participants in the learning and change process with the responsibility to expand windows of opportunity, created at such cost, and to transfer what is learned to other organizations. General patterns have emerged from the study of particular accidents like Columbia and other forms of research on safety and complex systems. These results define targets for safety management to avoid more repeats of past organizational accidents.

Organizations in the future will balance the goals of both high productivity and ultra-high safety given the uncertainty of changing risks and certainty of continued pressure for efficient and high performance. To carry out this dynamic balancing act, a new safety organization will emerge designed and empowered to be independent, involved, informed and informative. The safety organization will use the tools of Resilience Engineering to monitor for "holes" in organizational decision making and to detect when the organization is moving closer to failure boundaries than it is aware. Together these processes will *create foresight* about the changing patterns of risk before failure and harm occurs.

Acknowledgments

The analyses presented are based on the author's role as a consultant to the Columbia Board including participating in the Board's Safety Symposium on Organizational Factors, Houston TX, April 27-28, 2003 and on the author's discussions with US Senate and House science committees regarding the effectiveness of NASA's planned safety reforms, including invited testimony, hearing on "The Future of NASA", Committee on Commerce, Science and Transportation, John McCain, Chair, Washington DC, October 29, 2003.

This work also was supported by a cooperative agreement with NASA Ames Research Center (NNA04CK45A) to study how to enhance organizational resilience for managing risks.

References

Adamski, A. J. and Westrum, R. (2003). Requisite Imagination: The Fine Art of Anticipating What Might Go Wrong. In E. Hollnagel (ed.), *Handbook of Cognitive Task Design*. Erlbaum, 2003.

Brown, J. P. (2005). Ethical Dilemmas in Healthcare. In M. Patankar, J. P. Brown & M. D. Treadwell (eds), Ethics in Safety. Cases From Aviation, Healthcare, And Occupational And Environmental Health. Ashgate, Burlington VT, in press.

Carthy, J., de Leval, M. R. and Reason, J. T. (2001). Institutional Resilience in Healthcare Systems. *Quality in Health Care*, 10: 29-32.

Chow, R., Christoffersen, K. and Woods, D.D. A Model of Communication in Support of Distributed Anomaly Response and Replanning. In Proceedings of the IEA 2000/HFES 2000 Congress, Human Factors and Ergonomics Society, July, 2000.

Cook, R. I., Woods, D. D. and Miller, C. (1998). *A Tale of Two Stories: Contrasting Views on Patient Safety*. Chicago, IL: National Patient Safety Foundation.

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.

Dekker, S. W. A. (2002). *The field guide to human error investigations*. Bedford, UK: Cranfield University Press/Aldershot, UK: Ashgate.

Dekker, S. W. A. (2004). Ten Questions about Human Error: A new view of human factors and system safety. Lawrence Erlbaum, Hillsdale NJ, in press.

Feltovich, P., Spiro, R., & Coulson, R. (1997). Issues of Expert Flexibility in Contexts Characterized by Complexity and Change. In P. J. Feltovich, K. M. Ford & R. R. Hoffman (Eds.), *Expertise in context : human and machine*. Cambridge, MA: MIT Press, 1997.

Gehman, H. W. (2003) *Columbia Accident Investigation Board Report*, August 2003.

Hollnagel, E. (2004). *Barrier Analysis and Accident Prevention*. Taylor & Francis, London.

Hollnagel, E., Woods, D. D. and Leveson, N., editors, (2005). *Resilience Engineering: Concepts and Precepts*. Brookfield, VT: Ashgate Publishing Company.

Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, MA: MIT Press.

Johnson, P. E., Jamal, K. and Berryman, R. G. (1991). Effects of framing on auditor decisions. *Organizational Behavior and Human Decision Processes*, 50, 75-105.

Klein, G., Feltovich, P., Bradshaw, J. M. and Woods, D. D. (in press). Coordination in Joint Activity: Criteria, Requirements, and Choreography. In W. Rouse and K. Boff (Ed.). Organizational Dynamics in Cognitive Work, Wiley.

Lanir, Z. (1986). *Fundamental Surprise: The National Intelligence Crisis* Eugene, OR: Decision Research. (originally Tel Aviv: HaKibbutz HaMeuchad, 1983, Hebrew).

Low, B., Ostrom, E., Simon, C. and Wilson, J. (2003). Redundancy and Diversity: Do they influence optimal management. In F. Berkes, J. Colding and C. Folke (eds.), *Navigating Social-Ecological Systems: Building Resilience for Complexity and Change*, Cambridge University Press, NY, 83-114.

Patterson, E.S., Watts-Perotti, J.C. and Woods, D. D. (1999). Voice Loops as Coordination Aids in Space Shuttle Mission Control. *Computer Supported Cooperative Work*, 8, 353—371.

Patterson, E.S. and Woods, D.D. (2001). Shift changes, updates, and theon-call model in space shuttle mission control. *Computer Supported Cooperative Work: The Journal of Collaborative Computing*, 10(3-4), 317-346.

Patterson, E. S., Cook, R. I., Woods, D.D. and Render, M.L. (2004). Examining the Complexity Behind a Medication Error: Generic Patterns in Communication. *IEEE SMC Part A*, 34(6), 749-756.

Patterson, E.S., Cook, R. I. and Woods, D.D. (in press). Gaps and Resilience. In M. S. Bogner (ed.) *Human Error in Medicine*, second edition. Erlbaum.

Rasmussen, J. (1990). Role of Error in Organizing Behavior. *Ergonomics, 33*, 1185-1190.

Rasmussen, J., Pejtersen, A. M. and Goodstein, L. P. (1994). At the periphery of effective coupling: human error. In *Cognitive Systems Engineering*. New York: John Wiley & Sons, pp. 135-159.

Reason, J. (1997). *Managing the Risks of Organizational Accidents*. Brookfield, VT: Ashgate Publishing Company.

Rochlin, G. I. (1999). Safe operation as a social construct. *Ergonomics, 42* (11), 1549-1560.

Stephenson, A. G. et al. (2000). *Report on Project management in NASA by the Mars Climate Orbiter Mishap Investigation Board*. NASA, March 13, 2000.

Sutcliffe, K. and Vogus, T. (2003). Organizing for resilience. In K.S. Cameron, I.E. Dutton, & R.E. Quinn (Eds.), *Positive Organizational Scholarship*. San Francisco: Berrett-Koehler, p. 94-110.

Weick, K. E., Sutcliffe, K. M., & Obstfeld, D. (1999). Organizing for high reliability: Processes of collective mindfulness. *Research in Organizational Behavior, 21*, 13-81.

Watts, J.C., Woods, D.D. and Patterson. E.S. (1996). Functionally Distributed Coordination during Anomaly Response in Space Shuttle Mission Control. Proceedings of Human Interaction with Complex Systems, IEEE Computer Society Press, Los Alamitos, CA.

Woods, D. D. (2000 a). Behind Human Error: Human Factors Research to Improve Patient Safety. *National Summit on Medical Errors and Patient Safety Research*, Quality Interagency Coordination Task Force and Agency for Healthcare Research and Quality, September 11, 2000. http://www.apa.org/ppo/issues/shumfactors2.html

Woods, D. D. (2000 b). Designing for Resilience in the Face of Change and Surprise: Creating Safety Under Pressure. Plenary Talk, Design for Safety Workshop, NASA Ames Research Center, October 10, 2000.

Woods, D. D. (2002). *Steering the Reverberations of Technology Change on Fields of Practice: Laws that Govern Cognitive Work*. Plenary address at the 24th Annual Meeting of the Cognitive Science Society. http://csel.eng.ohio-state.edu/laws

Woods, D. D. (2005). Conflicts between Learning and Accountability in Patient Safety. *DePaul Law Review*, in press.

Woods, D. D., Johannesen, L. J., Cook, R. I., & Sarter, N. B. (1994). *Behind Human Error: Cognitive Systems, Computers, and Hindsight* (State-of-the-art report). Wright-Patterson Air Force Base, OH: Crew System Ergonomics Information Analysis Center.

Woods, D.D. and Cook, R.I. (2002). Nine Steps to Move Forward from Error. *Cognition, Technology, and Work*, 4(2): 137-144.

Woods, D.D. and Cook, R.I. (2003). Mistaking Error. In M. J. Hatlie and B. J. Youngberg (Eds.) *Patient Safety Handbook*, Jones and Bartlett.

Woods, D. D. and Shattuck. L. G. (2000). Distant supervision—local action given the potential for surprise *Cognition, Technology and Work*, 2, 86—96.