The *Columbia* Disaster: Choice Points, Deficiencies, Dangerous Thinking

Frederick F. Lighthall

Cite as: Lighthall, F. F. (2014). The *Columbia* disaster: Choice points, deficiencies, dangerous thinking. www.high-techdanger.net

This analytical essay on the *Columbia* disaster's human causes unfolds in three parts. Part One is historical, tracing the accumulation of events and collective thinking. It begins with background history, traces six key choice points in deliberations about safety and danger, and ends with an account of the nature of signals and how some key signals were missed. Part Two of the essay is analytical, examining first how engineering and managerial standards of judgment had become degraded. It then takes a three-part look at near-term human causes of the accident. This is followed by a summary of deficient thinking that characterized the shuttle program in this segment of its history. Part Three examines the *Columbia* disaster's cumulative data to see the extent to which the syndrome of human causes I found in the *Challenger* disaster were present also in this *Columbia* accident. The essay concludes with a consideration of the extent, kinds, and levels of knowledge about human functioning that seems to have been absent and causal in NASA's deliberations that led to the *Columbia* disaster.

Part One: History

Background

After the *Challenger* accident, Allan McDonald and his Thiokol redesign team designed a creative cure for the boosters' field and nozzle joints. [1] NASA made improvements in other parts of the shuttle system and shuttle flights resumed with the launch of *Discovery*, STS-26R, on September 29, 1988.

Discovery completed its four-day mission two years and eight months after the *Challenger* accident. This successful flight was followed by eighty-five others, including one nearly catastrophic flight just after the *Challenger* accident. Then, the eighty-seventh flight, fourteen years after *Discovery*'s return to flight, ended once more in disaster and loss of all crew members. Returning to earth the morning of February 1, 2003, after sixteen days in space, the Orbiter *Columbia* disintegrated shortly after reentering the Earth's atmosphere.

Seventeen days earlier, 81 seconds after *Columbia's* January 16 launch, a chunk of foam insulation had torn loose from *Columbia's* external tank. The light, Styrofoam-like chunk of insulation had been struck almost head-on by the Orbiter's wing at a net speed of more than 500 miles an hour, bursting a large hole in the leading edge of *Columbia's* wing. *Columbia* ascended safely into its orbit with this hole in its wing. Returning to earth after completing its sixteen-day mission, the friction between *Columbia's* wing and the earth's atmosphere generated 3,000 °F of heat that penetrated the hole in its wing, weakening the wing's inner structure and causing disintegration of the Orbiter and the deaths of its crew.

Both foam shedding from the external tank and divots in the protective tile surfaces of the Orbiter's wings had occurred from the very beginning of shuttle flights. The very first shuttle flight suffered foam loss of enough debris striking the Orbiter's protective tiles that three hundred of the tiles had to be replaced before the next flight. [2] Despite the fact that both foam loss and divots in the Orbiter's TPS tiles were violations of design specifications, the overwhelming majority of foam debris from the external tank and of divots in Orbiters' protective tile surfaces had presented no physical danger.

The immediate issue after that first flight with its surprising damage to its tiles was whether, contrary to original design, some amount and size of foam debris, some velocities, and some divots in the tiles could be safely allowed. The first four shuttle flights, designated as test flights, provided sufficient technical information for officials to consider the then current extent of foam loss and minimal divots in the Orbiter's surfaces as both inescapable and safe.

The history of both foam loss and divots in the tiles since those test flights had generally proven the correctness of that judgment over the 112 flights up to the *Columbia* disaster. Despite changes in foam material and procedures, foam debris had occurred on every flight. The number of divots on the Orbiters' tiles had averaged 143 per flight on the upper and lower surfaces of the wing, with an average of thirty-one divots more than an inch long.[3] Danger lay in part in those averages, but also in the few deviations from those averages scattered across the eighty-seven flights since the *Challenger* disaster.

Deliberations about Danger from Foam Debris: Three Phases

Three major human causes of NASA's flawed deliberations about the dangers of foam debris arose over time in three sets of causes, each in its own time period. The first and perhaps most powerful cause developed gradually from the very first shuttle flight and accumulated force up to and beyond the *Challenger* accident to the *Columbia* disaster itself. The set of human causes in the second period arose in the shuttle flights immediately after the *Challenger* accident. These first two sets of causes brought about an uncontested collective mind-set (or "confirmation bias," see Lord, Ross, & Lepper, 1979 and Nickerson, 1998) regarding foam debris and gouges in the shuttle's protective tiles as disconnected from any "safety of flight issue," as causing only "turn-around" damage – in short, as safe.

Categorizing these gouges as merely turn-around damage helped to anaesthetize the engineers' subsequent diagnostic thinking about danger and the limits of safety. The engineers' time-tested (but not analytically tested) confidence short-circuited any tendency they might have possessed to apply the physics of energy, speed, and mass to the foam debris problem. This unwarranted confidence in the safety of foam debris led the engineers and managers to miss the most relevant parameter of foam's danger, its velocity. Finally, when an unusually large chunk of foam debris struck *Columbia*'s wing, the third set of human causes came into play, adding force to the previous two. Principal among this last set of causes, as was the case in the *Challenger* disaster, was reversal of the safety-protecting presumption of danger and the burden of proving safety (Lighthall, 2015, Chapter 8).

Six Choice Points

Six choice points arose in deliberations about flight safety that reveal crucial weaknesses in participants' engineering and managerial thinking. First were choice points arising with flight 27-R (the second flight after *Challenger*); then choices three and a half years later regarding flight STS-50; and then, ten years and sixty-two flights after that, choice points presented by flight STS-112 and by the fatal *Columbia* flight itself. The span of time, the number of flights, and mission successes across which these relatively few choice points were scattered constituted a shaping context for deliberations, a crucial context we must attend to in assessing the human causes of the disaster.

The first choice point, a deliberate decision, came in diagnosing and correcting damage that showed up on the second flight after the *Challenger* accident, flight STS-27R, the shuttle *Atlantis* flown December 2, 1988. A piece of lightweight insulation material had come loose from the right booster nose cone 85 seconds after launch, dislodging a thermal tile from *Atlantis*'s undersurface, exposing the Orbiter to the extreme heat of reentry. Reentry burn-through and destruction of STS-27R was prevented only by a thick aluminum plate that happened to be at the point where the tile was knocked loose.[4] The shuttle program had again experienced a near miss.[5]

Post-flight analysis of STS-27R's Orbiter (*Atlantis*) was confined to identifying the cause as being the impact on its wing by a small chunk of ablative insulation that covered its right booster's nose cone, coupled with the fact that it had come loose 85 seconds after launch, when the shuttle had accelerated to an advanced speed. The corrective action taken was simply to replace the nose cone ablative with the type used prior to the *Challenger* accident, ablative that had had no history of insulation loss. The first flawed decision was to accept this diagnosis and correction as sufficient to certify the next flight as ready. This decision was based on no analysis of how the small piece of relatively low-density ablative from the nose cone could generate enough striking force to dislodge a highly secure and well-tested thermal tile.[6]

The chief deficiency evident in the deliberations regarding STS-27R's damage was in the low standards of evidence and argument in judging the claims that the next flight would be safe, claims made in the absence of any quantitative analysis. The particular *source* of the damaging material, the ablative insulation of the booster nose cone, was replaced by a more reliable source; but the *dynamics of physical energy* by which the ablative material could dislodge the tile were ignored. That blind spot would figure fatally later in the *Columbia* disaster.

The second choice point arose three and a half years and twenty-one flights after 27-R, with the flight of the *Columbia* (STS-50), launched on June 25, 1992. A chunk of foam from the external tank (ET) bipod ramp 26 by 10 inches[7] came loose as *Columbia* ascended toward orbit. This chunk was the apparent cause of a gouge in the Orbiter's protective tile surface 9 inches by 4.5 inches by .5 inch—"the largest area of tile damage in shuttle history."[8] NASA's first response to this event therefore protected safety.

Both Marshall's ET Office (overseeing preparation of each external tank) and Johnson Space Center's Integration Office classified the foam strike as an in-flight anomaly (IFA). Designating an anomaly as an IFA required that the anomaly be investigated and either shown to be safe or shown to have been eliminated before another launch could take place.[9] Two different NASA offices had called for a halt in flights until the cause of this extensive tile damage was identified and rendered safe. Here, safety was being protected.

The second choice point lay in the responses of those same two offices to their own assessments of the danger of STS-50's tile loss. Instead of conducting kinetic-energy analysis of the striking force of foam debris, participants in STS-50's final Flight Readiness Review (FRR) concluded that the "shallow" (.5 inch) tile damage indicated that no excessive "aerodynamic loads" had been generated and that the large piece of foam debris had come from "inadequate venting" of the foam layer.[10] The engineers and managers participating in the FRRs leading up to the launch of the next flight after STS-50 failed to respond with engineering analysis to their own declarations that STS-50's unusual damage constituted an in-flight anomaly.[11] Reality was again sending a signal that almost succeeded in triggering a halt in flights pending a safety review. But two offices declaring an in-flight anomaly were not enough.

The engineers and managers in this case did at least turn their attention to "dynamic loads," addressing the issue of damaging energy. Whatever their analysis might have been, it almost certainly did not examine the kinetic energy of STS-50's foam debris striking its tiles. Such an analysis would have alerted all concerned that the main source of striking energy was neither the size nor the density of the foam insulation, the two factors that caught their attention. The key variable instead, one they consistently ignored, was the *velocity* of foam debris – that is, the net difference in speed between the chunk of foam debris and the Orbiter, as the Orbiter struck the debris.[12]

The third choice point came ten years and sixty-two flights later. Engineers and managers were trying to decide whether damage done to the left booster of the *Atlantis* shuttle (flight STS-112), flown just three months before the fatal *Columbia* launch, constituted a threat to flight safety. Engineers were alarmed by damage to the insulation covering one of *Atlantis*'s boosters that had been struck by an unusually large chunk of foam insulation from its external tank. Most alarming was how near the strike had come to an electronics box crucial to safe flight.

Bob Page, a NASA engineer at Kennedy and head of the Intercenter Photo Working Group, argued on behalf of the group that the 4 x 5 x 12–inch piece of foam debris from STS-112's bipod ramp, like the three previous debris incidents from the ET's bipod ramp, should be officially classified as an in-flight anomaly (IFA).

Neil Otte, deputy manager of NASA's ET Project Office, and Jerry Smelser, ET Manager, disagreed. They argued that the evidence regarding foam debris did not demonstrate a danger to flight safety and did not warrant being classified as an IFA. Rather, foam debris constituted, they said, only "a turn-around issue." That is, foam debris would cause nicks and gouges in tiles requiring only additional time to repair or replace the tiles, thus extending the time required for the damaged Orbiter to be returned to service. The managers prevailed. They decided not to classify *Atlantis*'s foam loss and damage as an in-flight anomaly. If the foam loss had been labeled an in-flight anomaly, no further flights would have been allowed until either the cause or the effects of foam loss were eliminated. Instead, they decided that STS-112's foam loss and damage was merely worthy of further study.[13]

The CAIB's account reveals two important weaknesses. First, Ron Dittemore, who chaired the Board meeting, after hearing Otte's and Smelser's objections, assigned the action item to Smelser's group "to determine the root cause of the foam loss and to propose corrective action," and to report their findings on December 5, 2002, more than *three weeks after* STS-113 was scheduled to launch, November 10. Subsequently Smelser's due date was postponed until after the launch and return of the STS-107, *Columbia*. So Dittemore "decided to fly not one but two missions before resolving the STS-112 foam loss." [14]

Dittemore's confidence was based on his layman's sense of the light weight foam debris, uninformed by the physics of kinetic energy. So confident was Dittemore that a chunk of foam 4 x 5 x 12 inches could not do serious damage to the Orbiter that he subverted safety policy and practice in three ways. First, he disregarded Page's urgings -- urgings from a front-line engineer highly experienced in assessing foam debris -- that this debris like the previous ones be classified as an In-Flight Anomaly, thus allowing flights to continue before investigating and removing the source of the debris.[15] Second, the action item Dittemore charged Smelser to carry out identified the *source* of the foam loss *but not its potential for doing damage*. The physical cause of the foam loss was simply ignored – again, probably because Dittemore was subjectively convinced that the light foam material could not really damage the heat tiles. Finally, he allowed even that investigation to be postponed until after he approved two more flights, including the *Columbia*.

Despite Page's sense of palpable danger, and despite his persistence in arguing for an IFA, neither he nor his group members thought to present any quantitative analysis—the coinage of engineering persuasion. They failed to analyze the debris' energy. Were they, too, oblivious of the physics?

The engineers' and managers' deliberations failed on two counts. The engineers who saw the danger of STS-112's foam loss did not engage in enough engineering thinking to compute the kinetic energy that foam debris could acquire; and their managers never thought to task their engineers with analyzing the kinetic energy possible.

Three occasions had arisen, then, for participants to respond to unusual anomalous events with relatively simple engineering analysis of the kinetic energy of bodies in motion. The three opportunities for awareness had arisen over a period of three and a half years. In that period of time no one at NASA or NASA's contractors had wondered about how lightweight, low-density foam could acquire the energy sufficient to cause the damage that foam debris was credited as having caused. What could have caused this failing of engineering imagination and analysis? Part of the answer may lie hidden in the fourth choice point.

The fourth choice point was a decision to circumvent established accountability procedures designed to ensure safety. This decision was reached in preparation for the next flight after STS-112. It related to the flight readiness of the shuttle *Endeavor* (STS-113), flown November 23, 2002. The FRR to certify *Endeavor*'s external tank as flight ready reviewed the history and significance of ET foam loss. The review convinced two officials to refuse to sign *Endeavor*'s certificate of flight readiness. The dissenters were Pete Rodriguez, a member of NASA's Structures and Dynamics Laboratory, and Angela Walker, a safety manager. They were not convinced that the cause of foam loss and damage to *Atlantis*'s booster had been effectively examined, much less explained and eliminated. Here was actual behavioral resistance, the first and only instance of it, to the steady certification of foam loss and damage as harmless to flight in the post-*Challenger* shuttle history.[16] The decision in question[17] was to persuade the two dissenters who were withholding their signatures from *Endeavor*'s readiness certificate to sign the certificate or, if the dissenters still refused, to find some other person who would sign instead. One dissenter relented and signed; the other resisted, and a substitute official was located who did provide a signature.[18]

The procedures requiring responsible parties to sign their approval of flight readiness were put in place to insure that responsible, deliberate judgment would take place in flight certification. These procedures were part of the infrastructure to support serious deliberation and judgment. The subversion of these safeguarding procedures in this instance constituted an attack on a central pillar of deliberation, personal accountability for informed judgment. Cajoling a person to sign is to substitute the mere outer appearance of deliberation for the actual deliberation behind the two resistors' initial doubt and resistance; a triumph of power over informed thought, and of production over safety. Thereby is revealed a weakness in leadership, signaling also possible weaknesses more broadly in organizational culture.

Quality of deliberation suffers, and safety is undermined, when leaders develop such certainty in their thinking, or such desperation to meet a production schedule, that they dismiss all contrary thinking as either wrongheaded or troublemaking insubordination. We have to ask: what might have brought a manager to circumvent and subvert these safety-protecting procedures?

The fifth significant choice point was a failure on the part of several managers to respond to a weakness in safety reasoning, a weakness they themselves had sensed. The rationale presented at *Endeavor*'s FRR for certifying readiness was recognized by several managers as weak ("lousy," "stinks"), as below NASA's normal standards of evidence.[19] The managers could have probed further, objected to the rationale as inadequate and could have called for an action item to correct the cause of *Atlantis*'s threatening foam loss and damage before *Endeavor* would be certified ready. Instead they acquiesced and approved *Endeavor's* flight readiness despite their misgivings about the weak engineering rationale. The approval lacked quantitative evidence or analysis—no safety margins, no quantification of foam debris' net velocities, no quantification of kinetic energy.

The weak engineering rationale for the readiness of STS-113 in the face of too-easy acceptance of STS-112's anomalies and overt engineering resistance—a flawed rationale allowed by managers to stand—became the basis for launching *Endeavor*. *Endeavor*'s launch and mission turned out to be successful. *Endeavor*'s weak rationale, however, became the basis for certifying *Columbia* as ready for its flight.

The sixth and final disastrous decision was made five days *after Columbia was launched,* five days in which some engineers had calculated from the probable size, density, velocity, and angle of impact that the foam chunk from *Columbia*'s ET could have enough force to penetrate the Orbiter's thermal protection tiles.[20] This sixth decision had two parts. The first was the managerial judgment, based on the engineers' interpretation of an inappropriate computer model (the "CRATER" analysis), leading them to believe that the unusually large chunk of foam from the *Columbia*'s ET, shown by films to contact the Orbiter's left wing, would cause only "turn-around" damage, not a "burn-through."[21] The second part of the decision was the judgment by Linda Ham, chair of the Mission Management Team, that even if the wing was damaged, nothing could be done about it.[22]

Once again, in the face of films showing a chunk of foam debris dramatically larger than usual striking the Orbiter's wing, not even engineers who sensed the danger of that strike thought to calculate its dangerous *energy*. Rodney Rocha, division chief for structural engineering at Johnson Space Center and chair of JSC's debris assessment team, who became convinced that the films of *Columbia*'s foam strike signaled possible danger to Orbiter and crew, spearheaded the effort to obtain more definitive photographic imagery by military satellite. Despite persistent pleading, including some intense confrontations with others, Rocha never thought to calculate, or to ask others to calculate, the striking energy of *Columbia*'s unusually large chunk of debris.[23] Had he done so, he would have possessed a finding that would have halted all other thoughts.

By the third day of *Columbia's* mission the video and photo images of the foam debris from *Columbia's* bipod had been analyzed by Boeing engineers, They came up with two estimates of debris size: 20 x 20 x 2 inches, and 20 x 16 x 6 inches. The net speed of the debris was estimated to be *750 feet per second* or 511 miles an hour, and estimated to strike the leading edge of *Columbia's* wing at an angle of less than 20 degrees. [24] These early estimates were close to more refined ones made after the accident, which calculated the velocity at 546 miles per hour (800 feet per second) and debris size at 24 x 15 x 3.3 inches. The fact that the *Columbia*

Orbiter's wing struck that dislodged chunk of foam debris at the net velocity of 546 miles per hour, which would generate a striking energy of more than 15,500 foot pounds—fourteen times the muzzle force of a high-grain .44 Magnum pistol bullet—was so shockingly contrary to the prevailing view of the lightweight foam debris as harmless that any such computation would initially have been disbelieved.[25]

With such calculations Rocha could then have challenged all naysayers to do their own calculations and to look up their own equivalents of damaging force. Whoever might do those calculations would undergo a radical cognitive shift regarding the safety of foam debris. Then attention could turn to the plight of the seven astronauts, with thoughts leading to a rescue attempt. But that engineering analysis, and the *engineering imagination to think of its possibility*, was absent.

Part Two: Analysis

Causes: Missed Signals

What could account for this complete absence of a quantitative analysis of the physical potential of foam debris, a failure of both NASA and Rockwell engineers and managers over a span of years, to think of applying a relatively simple physical equation?

Consider the actual incidence of foam debris and Orbiter damage. Among the six choice points, only three were directly related to what could be regarded as *signals*: the three that involved unusual anomalies of actual damage to the Orbiter. Three signals of danger among many other instances of minor, unavoidable, and harmless damage. If we consider each of the three instances of damage—in flights 27-R, 50, and 112—in isolation from all other damage, we easily see them as clear warning signals that the physical realities of shuttle flight were sending to the engineers and managers.

But the research on what is known as *signal detection theory* shows that signals alone do not bring about detection.[26] Detection of signals depends, it turns out, on a *comparison*—between the frequency and strength of signals (their clarity, distinctiveness, dramatic qualities) and the frequency and strength of events similar to the signals. These similar but harmless events are considered “noise” surrounding the signals. A key determinant of whether signals of danger will or will not be detected is the signal-to-noise ratio.[27]

In the case of our three signals, they were all recognized and labeled by at least some engineers as unusually dangerous, as events that deviated from all the other instances of damage to the Orbiter. But at issue is the *believability* of apparent danger in any particular instance.

Consider: from the beginning of shuttle flights up to *Columbia*, there had been 112 flights. The Orbiters of those 112 flights *averaged* 143 divots per flight (thirty-one with more than one inch at their widest), most caused by foam debris. The accumulated experience of foam-caused damage, then, totaled 16,016 safe divots, of which three unusually large instances signaled danger. The signal-to-noise ratio, then, computes to .00019, or nineteen events in one hundred thousand.[28]

To put the matter another way, the percentage of all Orbiter divots that were non-threatening divots was 99.98. At the time 27-R was flown, furthermore, 3,718 divots would have occurred in the Orbiters, only *one* of which might have caused Orbiter damage significant enough to signal danger.[29] In that case the percentage of safe Orbiter divots to the total on the eve of flight 27-R would have been 99.97 percent. Not a signaling history that prompts vigilance for danger.

We must remember, however, that while we, examining the accident, do this kind of counting as we try to understand how the engineers and managers thought, they were not counting. Their experience had been flooded with many innocuous hits on each flight, scores of harmless divots that severely blunted awareness of foam debris as dangerous. Their predecessors had dealt with a single instance of an unusual loss of foam (STS-7) and they had been forced to view Orbiter damage from ET shedding as regularly requiring repairs ("turn-around" damage), nothing more.[30] Any initial search for signs of danger was soon diverted elsewhere, particularly to Thiokol's booster joints.

With 27-R's Orbiter damage, attention turned to eliminating its source altogether.[31] The engineers found the smoking gun, the nose cone ablative, and solved the problem, replacing the offending ablative with the type of ablative that had served well in the first twenty-four shuttle flights. The ablative strike to STS-27R's Orbiter was seen as a new problem, now solved—an event unrelated to ET debris.

One answer, then, to the question of why the engineers and managers failed so completely to quantify the striking energy of foam debris is that the overwhelming number of instances of Orbiter damage from foam debris were harmless, so consistently harmless that foam debris was automatically regarded as causing turn-around damage only: foam debris = turn-around damage = safe. Foam debris came soon to be understood *categorically*. You didn't have to quantify it.

Degraded Standards: Leadership and Competence

Another answer to the question of why no one turned an engineering eye to the energy or striking force of foam debris lay in an apparent weakness in the Flight Readiness Review system. Managers in charge of FRRs for the Orbiter and the ET failed to demand deeper analysis—a failure to insist on high standards of evidence, to probe arguments offered, and to reject weak evidence and arguments. It seems clear from the assessments of 27-R's damage and its source (the nose cone ablative) that no one in the FRR who reviewed its glaring damage demanded to know in quantitative terms how the ablative could have dislodged 27-R's tile, a demand that William Lucas almost certainly would have made, were he to have returned from the *Challenger* era.[32]

But Lucas had retired. Evidently, those who replaced him either did not possess his depth of technical knowledge and were therefore unable to probe into technical questions or simply tended to accept their subordinates' explanations at face value. Whatever the cause, it is clear that Lucas-like standards had become degraded after his retirement. Certainly, one contributing factor to lowered engineering standards in FRR assessments seems to have been a lowered level of analytical competence among key engineering managers.

For example, Cabbage and Harwood (2004) quote Jerry Smelser, manager of Marshall's External Tank office, revealing his limited analytical response to the problem of foam debris:

> "A piece of foam this size does not weigh much, although I don't have the background or the engineering expertise or all the tools to predict what is going to happen with the foam once it comes off. But from a practical engineer's standpoint, it did not appear to me that anything that light could do damage to an Orbiter. That was a practical farm boy engineer's judgment, but that also was substantiated by the people who did the analysis." 33

When engineering managers came before Marshall's Jerry Smelser to argue the readiness of the ET, he would not ask them to report levels of striking energy of the previous flight's foam debris. He would not do so because he was unable to do so. First, he did not have the analytical competence; second, he already possessed a "practical farm boy engineer's judgment" that nothing "that light could do damage to an Orbiter." Smelser, like others, had an image of the

ET’s pieces of foam debris, an image that presented foam debris and its striking power as innocuous: a piece of foam debris “was similar to a Styrofoam lid of a cooler hitting a windshield of a car or truck. . . . It was distracting, but not dangerous.’”[34]

So Smelser relied on others to do the technical analysis. But the knowledge-gap did not begin or end with Smelser. Smelser’s own supervisors clearly did not have sufficient knowledge to catch the gaps in his knowledge.[35]

Linda Ham, chair of the Mission Management Team and a manager with key responsibilities for decision making and oversight, relied on others for technical analysis: : “I personally [do not], nor does the MMT, do the analysis. We must rely on our contractor work force... we don’t have the tools to do that. We don’t have the knowledge to do that or the background or expertise to do that kind of thing” [36] Ham recognized Smelser’s rationale for safe launch of STS-112 as “lousy.”[37] Yet she lacked the technical engineering competence to confront Smelser and others with specific technical questions that would reinforce technical standards.

The post-*Challenger* (post-Lucas) engineering managers, with lowered technical knowledge and skills, had been allowed to become (hired to become?) more managerial, less engineering, and certainly had become less technically demanding of FRR presentations—and therefore more prone to certify as ready conditions that had not reached safe levels of engineering proof. A comparison, then, of the *Challenger* FRRs at Marshall with the *Columbia* FRRs at Marshall suggests the crucial importance of managers -- even high-level managers like Lucas, whose engineering knowledge goes deep enough to question and probe technical analyses—to demand penetrating, quantitative analysis whenever it is absent.

An image like Smelser’s—picturing foam debris in a highway scenario where a Styrofoam cooler lid comes off one vehicle and strikes another—can powerfully restrict thinking if the image itself is never questioned. Far from being questioned, the image spread—all the way up to Sean O'Keefe, NASA's top administrator. He likened the shuttle's foam debris to "a Styrofoam cooler blowing off a pickup truck ahead of you on a highway." [38]

Somewhat lower in NASA's management than O'Keefe, but well above Smelser's level, Ron Dittemore held a similar view of a very large piece foam debris from *Columbia*:

> "It's fragile, it's easy to break, and it's easy to break up into particles. . . . And so it's difficult for us to believe, as engineers, as management and as a team, that this particular piece of foam debris shedding from the tank represented a safety of flight issue. . . . Right now, it does not make sense to us that a piece of debris would be the cause for the loss of *Columbia* and its crew." [39]

Engineers and managers alike, with a few notable exceptions, had become blinded by the immediate, palpable properties of the foam insulation, the fluffy lightness of the material itself. The experience of holding a piece of that insulating material, with its low-density lightness, was so powerful it preempted all other physical considerations.[40] Their layman's view of the material blocked out considerations of velocity, the variable (with its quantity *squared*) that so dominates the equation for kinetic energy.

With respect to calculating the dangerous energy of foam debris, therefore, we see widespread failure of NASA's engineering mind. Effective engineering is carried forward by the mind that seeks always to quantify stresses and forces as the only way to understand them. Yes, the appealing highway scenario of a Styrofoam lid flying off and hitting a following car was suitable to reflect the harmlessness of 99.98 percent of the foam-caused divots. And, yes, key managers lacked sufficient technical knowledge to raise technical questions. But why not even a single engineer found time to muse for twenty minutes about the actual force of a chunk of foam traveling at *shuttle speeds in contrast to highway speeds* remains a mystery – until we understand that these engineers and managers, without exception, either lacked the knowledge of the physics of motion or failed to see the relevance of that knowledge to foam debris at shuttle speeds.

Many social scientists will have a ready explanation: *groupthink* (Janis, 1972). Groupthink refers to the situation where all members of a group form a pattern of thinking that excludes obvious realities that are relevant to, even crucial to, the group's effective decision making. The *Columbia* participants did reveal groupthink both with respect to the energy potential of flying foam debris and with respect to the metaphor of the Styrofoam lid flying off the pickup truck and striking the car behind. (The fact that a manager like Ham could use the term "kinetic energy" does not invalidate the fact that all participants simply failed to think of quantifying the strike force or energy of pieces of foam debris at shuttle speeds.)[41]

That NASA and contractor engineers and managers exhibited groupthink in this respect for more than fourteen years explains nothing, however. Rather, it begs for its own explanation. How was it that no engineering imagination ever strayed to the image of Styrofoam traveling at 500 miles per hour, then calculated that speed to be 733 feet per second, and then imagined what the striking force would be of a piece of foam the size of a loaf of bread traveling straight toward them across the length of almost two and a half football fields in a *single second*? How was it that no engineer in all those years was struck by the thought that a chunk of flying Styrofoam at highway speeds didn't match flying foam at shuttle speeds?[42]

This failure of engineering imagination and analysis led directly to the violation of one of the crucial requirements of effective deliberation about safety and danger, the first requirement of *relevance.* The relevance of the *relative velocity* of foam was completely missed. Even when velocity was mentioned in e-mails leading up to the final decision that the *Columbia* was safe, the writer mentioning it did not think to try computing what a piece of foam debris' net velocity might yield in terms of momentum or energy.[43]

In the *Challenger* deliberations the relevance of temperature was denied by accepting a NASA manager's false argument that there was no correlation between O-ring temperature and O-ring sealing performance (Lighthall, 2015, 75-78). The *Columbia* deliberations denied the relevance of foam debris by accepting the layman's naïve view that the potential for foam debris to do damage depended on its light mass, not on the underlying physical dynamics of mass, velocity, and energy. [44] Immediate sense perception was able to preempt mindful technical analysis.[45]

Near-Term Causes

The deficiencies already noted spanned the entire post-*Challenger* era up to the *Columbia* accident. Two conditions affected deliberations specifically regarding the *Columbia* flight and the flights immediately preceding it. First was pressure to meet a tight flight schedule; second was, once again, reversal of the safety-protecting presumption and burden of proof.

Launch Pressure. Unusual pressure to meet launch schedules resulted from a mid-2001 decision of NASA's top management to set a hard deadline to complete a Core segment of the International Space Station (ISS) called "Node 2." As shuttle flights proceeded in the year and a half before *Columbia*, preparations for each flight had to cope with a variety of problems that ate time and that placed NASA increasingly behind schedule. In response, work shifts were added, inspection requirements were reduced, operations normally done in sequence were performed simultaneously, and work time became increasingly compressed.

While the managers at the top of NASA's hierarchy were unaware of (or denied) an increase in launch pressure, the work force became acutely aware of managers' concerns—

communicated through management's addition of work shifts and time-saving tactics, and witnessed by a new screen saver NASA headquarters sent to all employees. The screen saver, appearing on all computer monitors, gave a line-by-line listing of the days, hours, minutes, and seconds before the launch scheduled to complete Node 2, on February 19, 2004. The heading above the listed days and times was "Countdown to Space Station Program U.S. Core Complete February 19, 2004." Every NASA manager—and every engineer who could see a monitor—could be reminded graphically of how the seconds before that key launch were flying by.

Managers had even more graphic and detailed reminders. Some presented and others listened to monthly briefings at NASA headquarters about how well or poorly the Node 2 schedule was being met. Briefing slides presented timelines marked by triangles to show schedule interruptions (green if ahead of schedule; red if behind—red dominating), with squares to show corrections for each interruption that would make up for lost time. The vertical axis of these slides marked reserve margins in months, with most of the triangles and squares in the minus area below the break-even line that represented "on schedule." Attention at headquarters was on three questions: What caused each slow-down? What can be done to prevent such interruptions? And how can the program make up for the time lost? Time and schedule. Safety had no comparable briefings, graphs, or screen saver.

The net result, socially and psychologically was an impact on lower managers and engineers that induced unwritten rules: Get it done, do it quickly; stay on task; don't lose time with irrelevant details; shorten or omit unnecessary steps. Mental worlds became infused with little, constant, back-of-the-mind self-reminders of time and urgency. Some unattributed quotes from NASA workers capture that sense of urgency:

> *"A one-week hit on a particular launch can start a steam roll effect including all [the]*

> *constraints and by the time you get out of here, that one-week slip has turned into a couple of months."*
>
> *"The thing that was beginning to concern me . . . is I wasn't convinced that people were being given enough time to work the problems correctly."*
>
> *"I have to think that subconsciously that even though you don't want it to affect decision making, it probably does."*[46]

NASA's pressures on all to meet flight schedules in the two years leading up to the *Columbia* accident, therefore, took forms more pointed and reached intensities far greater than the considerable launch pressures that affected the decision making in the *Challenger* case (Lighthall, 2015, Chapter 3).

It is clear that both the standards for proving the shuttle's safe readiness and managers' capacity for probing evidence had become weak. This was apparent in the discussions about STS-112's debris and damage to the insulation of its booster and discussions in the FRRs for the *Endeavor* flight (STS-113), where the rationale for readiness was recognized as "lousy." Acceptance of that weak rationale and resistance to classifying STS-112's chunk of foam debris as an IFA reflected lower standards already part of the engineering subculture, but certainly exacerbated by organization-wide pressures to meet the flight schedule. Those pressures and the lowered standards were reflected in the efforts to press two officials to sign *Endeavor*'s certificate of flight readiness—standards so lowered as to bring Rodriguez and Walker to the point of refuting official claims of having proved flight readiness, and pressures so heavy as to subvert established accountability procedures by cajoling Rodriguez to sign and by finding a substitute to sign for Walker.

Here we had two members of an FRR, engineers Rodriguez and Walker, offering critiques of the readiness rationale, performing precisely the questioning function that safety required and that FRRs were designed to carry out, two questioners being silenced or bypassed so that a launch could proceed. Such was NASA's capacity to probe and question, weak already but further subverted under pressure to meet the flight schedule.

Presumption and Burden. The second condition affecting *Columbia*'s deliberations was that the safety-protecting presumption and burden of proof, recognized even in post-*Challenger* FRRs, became once again reversed. As soon as *Columbia* was launched, all participants began to focus on seeking evidence of danger while tacitly presuming the shuttle would be safe if danger could not be verified.

From the moment on the day after launch that engineers recognized dramatic danger in films of *Columbia*'s foam debris, all efforts were geared to finding out whether damage had taken place or how severe the damage might be. Many managers, Ham and Dittemore among them, simply assumed that unless *foam density or mass* could be shown to be dangerous, it could cause only turn-around damage. Others saw the danger of the unusually large size of *Columbia's* chunk of foam debris, but sought to gather clearer images of the damage it might have inflicted, the tacit assumption being that unless and until damage could be shown, the *Columbia* Orbiter and crew were safe. It is painfully ironic that, even under this reversed burden of proving danger, engineers or managers who might have applied the algorithm for computing the energy of a moving body could have demonstrated the gut-wrenching level of peril in which *Columbia* and its crew were caught, thus *proving flight danger*.

In the first week after *Columbia* was launched, no engineer or manager faced as real the possibility that fatal damage had occurred. A participant's summary of the day-six discussion of the debris assessment team at Johnson lists the questions it addressed: "Where is it [the foam debris] hitting? What is that cloud [of debris]? How big is that piece? Is it really coming from the bipod area or not? Is it spinning? Can we see damage?" All of these questions addressed the specifics of source and likelihood of damage, but none focused on the velocity of the debris. All of the questions assumed, even on day six, that the Orbiter and crew were in no danger because no damage had yet been proven.[47]

In contrast, the safe path of deliberations would have been to assume, immediately after viewing the shocking film images on day two, that fatal damage had been done, that crew and vehicle were in serious danger until evidence proved otherwise. That rule is comparable to assuming all firearms are loaded and ready to fire until someone demonstrates that they are unloaded. Having assumed danger on day two of *Columbia*'s mission (while simultaneously gathering evidence of safety), steps could then have been taken to plan for the crew's rescue. [48]

One engineering error in the first days after *Columbia*'s launch drew a specific conclusion in the absence of specific calculations. Mike Stoner, a thermal protection technical manager at United Space Alliance, wrote a memo to colleagues assuring them of no danger to the Orbiter's wing leading edge (WLE), protected by the RCC panels. After conferring with Calvin Schomburg, NASA's tile specialist, he felt confident in asserting that "at T+81 seconds [81 seconds after *Columbia*'s launch], the piece wouldn't have had enough energy to create large damage to the RCC WLE system." [49] He was precise in the relevant physics of safety, but precisely in error.

Stoner was offering proof of safety to his colleagues clearly without having made the calculations of the algorithm for kinetic energy, even while citing an indicator of the shuttle's velocity, time elapsed since launch. His memo is notable also both for its explicit focus on the kinetic energy of a piece of foam debris and for its acceptance of the proper burden of proof. But while his memo reflects the safety-protecting burden and presumption, his argument substitutes assurance and erroneous conclusion for accurate calculation.

The only explicit reference to burden of proof in all of the deliberations was clear but indirect. In an exchange on day eight of *Columbia*'s mission between Rodney Rocha, chief of the structural engineering division for the shuttle at JSC and Calvin Schomburg, Rocha told Schomburg that they should assume *Columbia* was unsafe until they could prove it was safe. Schomburg replied that it was unreasonable to believe *Columbia* was in danger, that the history of foam debris showed that the *Columbia* damage would cause nothing more than tile replacement — the familiar "turn-around" view of debris. [50] Clearly, Schomburg had not the foggiest idea of burden or presumption (or of "confirmation bias"), missing Rocha's meaning completely. It is also clear that Schomburg never framed the problem of debris strikes as a problem of kinetic energy in terms of its formulation, E (energy) = [M (mass) x V^2 (velocity squared)]/2. In Schomburg's view, foam debris simply, categorically, and absolutely posed no risk to flight safety.

When Rocha went on to discuss how damage could be mitigated at reentry by Orbiter maneuvering, Schomburg reassured him that the reentry had already been designed to minimize damage. The exchange became more heated.[51] But no further reference was made, in this exchange or any other, to presumption or burden of proof.

The Deliberative Situation. Were *Columbia's* deliberations under the same kind of time squeeze that was true for the *Challenger* deliberations? Were *Columbia*'s deliberative resources diminished by the number of days (about thirty)[52] that the crew could stay in space before their physical survival required a return to earth? In principle, if either the right presumption had been recaptured in managerial consciousness or the results of energy calculations had penetrated management, either one could have taken place in time for a rescue to have been attempted within the launch window of an expedited next flight. [53] Unlike the *Challenger* deliberations, the deliberative situation triggered by the discovery of *Columbia*'s foam strike provided sufficient time for effective deliberation, had the collective intellectual resources been available.

Deficient Thinking: Summary

My reading of the CAIB's six volumes and of Cabbage and Harwood's (2004) indispensable narrative has suggested eight causes of deficient thinking and deliberation, causes of the accident that spanned three segments of shuttle history.

First, the dominance of relatively harmless instances of damage over rare instances of dangerous damage—the unfavorable noise-to-signal ratio—spanned flights from early in shuttle history right through to the *Columbia* accident. This preponderance of small divots and the lightweight quality of the foam debris created the widespread perception of foam debris as non-threatening, as a turn-around issue only, a perception that became part of the engineering and managerial culture. It was a culture over which a few engineers with clear eyes to danger could not prevail. Foam debris was considered simply irrelevant to flight safety.

A set of four causes arose in the second span of shuttle flights, beginning just after the *Challenger* accident and carrying through to the *Columbia* disaster. Lowered standards of evidence and argument in FRRs, weak or absent probing of evidence and argument in FRRs, and strikingly weak analytical competence on the part of managers at several levels of NASA all undermined the effectiveness of the readiness review system as a quality assurance and safety-protecting system. These causal conditions seem to have been caused by a fourth condition arising immediately in this second span of flights.

A change in safety-protecting organizational structures had taken place. William Lucas, who directed the Marshall Center for twelve years prior to the *Challenger* accident, had early instituted an additional FRR that he himself would chair, the "Marshall Center Board."[54] This additional board was intended to ensure that all shuttle elements being certified at Marshall would be up to Lucas's own standards of flight readiness. Further, because Lucas's standards were so demanding, Stanley Reinartz, who reported to Lucas, often held yet another FRR at the Shuttle Projects Office to make sure that presentations being prepared for Lucas's board would pass muster.

With Lucas's retirement and replacement, both of these FRRs were eliminated,[55] and the role of the director the Marshall Flight Space Center (MSFC) shifted away from penetrating technical oversight.[56] The loss of these two FRRs, and of Lucas's demanding influence on all readiness reviews below his level, allowed or promoted weaknesses in standards, probing, and technical competence of managers. All five of these conditions weakened the post-*Challenger* system of reviews designed to protect quality and safety.

Add to these causes of weaknesses in deliberation the increasing pressures to meet flight schedules, especially intense in the flights leading up to the *Columbia* accident—pressures to complete Node 2 of the International Space Station—and you have a formula for the domination of managerial commitment to production over (a weakened) engineering bias toward quantitative analysis of engineering performance.

The *Columbia* flight itself saw two additional conditions that combined with the previous six to weaken deliberations about safety and danger. The first was, once again, a reversal of the safety-protecting presumption and burden of proof; the second was the passive leadership stance that "nothing can be done about it."

After the *Challenger* accident, Marshall leadership became more passive, more accepting, less critical and less technically competent, leading to a weaker FRR process—lower standards of evidence and argument, weaker probing—showing a clear lack of understanding about how crucial to safety high standards are, and how crucial to high standards is a regime of informed, critical challenge. This weaker leadership then tilted in the managerial direction of production toward setting and meeting a tight flight schedule. The production culture became oblivious to the concerns of the few (Page, Rocha, Rodriguez, Walker) who doubted safety.

Finally, when engineers and managers became aware of possible damage to *Columbia* from the unusually large chunk of foam, all participants fell into the dangerous presumption that *Columbia* and its crew were safe until evidence of severe damage were found, showing once again widespread ignorance of the connection between safety, on the one hand, and the roles of cognitive framing, presumption, and burden of proof in assessing possible danger, on the other. Lacking that understanding, -- and lacking the clear evidence of danger that calculation of the kinetic energy of that piece of debris would show -- deliberations were foredoomed to concluding that the *Columbia* and its crew were safe.

The basic causes of the *Columbia* accident seem to boil down to three: rare signals of danger among frequent signals of safety; weakened post-*Challenger* competence and leadership in the FRRs; and reversed presumption and burden of proof.

Part Three: Columbia's Shared Vulnerabilities

I discovered (Lighthall, 2015) what I called a "syndrome" of four hidden vulnerabilities that characterized the fatal individual and collective thinking in the *Challenger* disaster. I argued that those four vulnerabilities would apply generally to high-technology situations where some new, possibly disastrous operating or environmental condition threatening the whole enterprise was detected a) during ongoing operations, b) by technical experts who c) communicated their evidence of the danger to managers responsible for deciding action implications. The four vulnerabilities threatening deliberation and accurate situation assessments were:

-- restricted time for participants to assess the likelihood and seriousness of the sudden threatening evidence and mixed, conflicting evidence;

-- the relative remoteness of managers from direct observation of technical conditions as contrasted with the immediate and ongoing contact of engineers with quickly shifting technical dynamics, causing a crucial lag in managers' technical knowledge when they must assess those dynamics promptly and accurately;

-- the mode of data-based deliberation and resolution of conflicting views about the new threat that entails a conflict-resolving *argument* a) whose elements, protections, and pitfalls are likely to be completely unknown to both engineers and managers and b) a form of argument that can itself determine the outcome of the issue when the evidence is mixed;

and

-- the subtle but gripping power of an organizational culture geared overwhelmingly to production on schedule, a culture defined by specific organizational routines tied directly to schedule-driven actions geared to produce a well-defined product or service on time in a high stakes setting of chronic danger.

Consider each of these as applying to the *Columbia's* deliberations. Time was limited to assess the *Columbia's* situation and to avert catastrophe if it existed. While correct assessment of the shuttle's and crew's disastrous situation could in principle have been achieved in two or three days, given the data available, planning and executing a rescue mission could only have been rushed. Given the framing of the assessment task as determining the presence of danger, i.e., to determine whether the Orbiter's wing was actually struck by the (observably large) chunk of foam, time for *that* task was limited. So *Columbia* was vulnerable on the first count.

The efforts of engineers like Page, Rocha, Rodriguez, and Walker to communicate their cautioning observations and reasoning to their managers met with resistance based on managers' confidence in the harmlessness of the light fluffy insulating material. So clearly evident here again, as in the *Challenger* deliberations, was the role-structured lag in technical knowledge between managers distant from direct observations of emergent dynamics and the front-line technical observers who saw and understood the new anomaly. *Columbia* was vulnerable on the second count.

The third element of the syndrome of vulnerabilities, use of reasoning that tacitly presumed safety and sought to prove danger was explicitly present, as already explained. Columbia was vulnerable on the third count.

What about the fourth vulnerability, a production-dominated organizational culture? The Columbia deliberations present a case study of an ethos of extreme production-dominating consciousness, a consciousness reinforced on every computer monitor in the organization by the running seconds, hours, and days to complete the space station's Node 2. *Columbia* was vulnerable on count four.

So we do see in the *Columbia* deliberations the same syndrome of dangerous vulnerabilities found in the *Challenger* disaster. Yet the *Columbia* deliberations were distinctly different from those in the *Challenger* case. The *Columbia's* deliberations were marked by a distinct new metaphor, a metaphor that falsely shaped managers' thinking up and down the line of management: the light Styrofoam lid coming off the cooler in the back of the pick-up truck and striking the windshield of the car following the truck. That neat, concise metaphoric mental model, completely misrepresenting the lid's velocity as modeling the foam shedding in the case of the space shuttle, was an entirely new element that arose in the *Columbia* disaster. Its prominence has much to suggest, all by itself, about NASA's organizational culture at the time.

Deliberative Deficiencies: Three Causes

Three human causes of these deliberative deficiencies present themselves—level of knowledge and skill, leadership, and organizational orientation.

1. Knowledge and skill. The level of engineering knowledge, analysis, and reflection evident in these *Columbia* deliberations was deficient in four important respects.

First, the novel and high-stakes nature of space exploration demands the ability to make acute differentiations between safe and dangerous degrees of a threatening condition (e.g., foam debris, O-ring erosion). This means that engineers need concepts and tools of detecting when new intensities of a usually safe condition arise, that is, skill in detecting a signal of possible danger – a signal that a familiar kind of phenomenon, one that is usually safe, has sharply shifted its frequency or intensity to a dangerous level.

Sensitivity to such changes in operating conditions would be increased in these safety-critical enterprises by wide organizational understanding of the idea of signals amidst noise, where tolerable sizes of foam debris, for example, are seen as a *noisy foreground* that hides unknowns behind its screen. But in these high-tech explorations detecting the emergence of a rare deviation from that normal foreground requires a firm, quantified record of that normal foreground, a record whose quantified properties (e.g., range of sizes of foam debris) are clear enough to reveal the sharpness of that deviation from normal.

In these enterprises, furthermore, detection of such rare deviations is but the first of four necessary steps in an effective *organizational* response to that new shift in conditions. The second step is to communicate the fact and severity of the new condition. Quantification of the degree of deviation from its normal (and quantified) foreground makes it possible to communicate the new deviation more clearly to managers, managers tasked with deciding the third step, the organizational interpretation and response to be made. The fourth step is effective organizational execution of the managerial action plan. A pre-condition of success in executing that plan in these enterprises, however, is the *clear communicability* of the rare deviation. Its rarity makes it unbelievable. It announces that the normal, apparently safe operating conditions, apparently confirmed as safe over time (see Lord, Ross, & Lepper, 1979 and Nickerson, 1998) have suddenly changed. That suddenly arising deviant implication will usually be unbelievable to other organizational members unless clear evidence of the extent of the shifted conditions can be presented in unambiguous, quantitative form.

The *Columbia* engineers and managers had *not* accumulated data on the size of their foam debris. Page and his colleagues had become experts in photo analysis and were able to detect dramatic deviations from their memories of usual sizes. However, in order to communicate convincingly any dramatic and other deviations from normal they would have had to keep measurements of all anomalies even when 99.5 % of them posed no actual flight threat. Keeping measurements of *all* anomalies (like depth of O-ring impingement erosion and O-ring temperature in the *Challenger* disaster) is a crucial step in detecting the rare serious ones.

The second gap in knowledge and skill was the engineers' and managers' ignorance of the dynamics of argument. It is unlikely in the extreme that those who specialize in engineering and engineering management receive training, formally or informally, in the nature and dynamics of argumentation. It is only remotely possible, also, that their exposure to television or literary depictions of courtroom dramas would lead them to draw any parallels to engineering arguments in a high-tech setting. Culturally, the informed world of evidence-based arguments and arguing is distant from the world of engineering arguments.

It seems to me that if any profession is culturally defined as being disciplined about argument it is not engineering, or medicine, or even science but rather the law. Further, our culture emphasizes a distinct contrast between engineering and the law, where engineering truth is found in exact quantification while legal truth is found by collecting subjective judgments about facts. The jury method of finding truth contrasts markedly with the idea of quantitative measurement. Yet the fundamentals of legal argument to examine evidence on a question of fact must become part of normal discourse in these safety-critical enterprises if decision making about a newly revealed operating threat is to be effective.

Whatever caused the *Challenger's* space engineers' and managers' ignorance of the dynamics and pitfalls of argument, knowledge among their future counterparts must expand to correct that deficiency. To protect safety in decision making, they and their successors must internalize five fundamental elements of effective engineering argument. They must come to understand first, the power over outcomes exerted by arguments or assumptions that make relevant facts seem irrelevant, as happened with the dismissal of O-ring temperature (*Challenger*) and of the damaging power of light-weight pieces of debris (*Columbia*).

They must understand the four other fundamental properties of formal argument: the power to determine a final decision exerted by the side of the argument that does not have the burden to prove its case (and the weak arguing position of the one so burdened); the framing of the case that must be proven (e.g., danger or safety); the standards of evidence and reasoning allowed or insisted upon; and verification, challenge, and rebuttal that is brought to bear on statements, assumptions, and conclusions. Wherever safety is at risk, and in these high-tech enterprises safety is *always* at risk by virtue of the unknowns, achieving safe outcomes when engineering disagreements arise will depend on participants' understanding and real-time management of these five pillars of effective and safe assessment of reality.

A third weakness of knowledge showed up in the *Columbia* engineers' and managers' simplification of physical complexities through use of analogy. The misleading power of a simple simile between Styrofoam debris traveling from a truck to a following car at highway speeds and Styrofoam-like debris traveling between the shuttle's ET and its Orbiter's wing at shuttle speeds replaced analytical *engineering* thinking among both engineers and managers.

What kinds of training and reminding will equip them and their successors habitually to doubt the comforting and simplifying effects of such similes? What training will lead them habitually to react skeptically, to replace “x is *like* y” with at least “x may be *kind of, sort of* like y,” or with “in *what ways* is x like y, and in *what ways is x not* like y?” To be cautious about one’s own thinking, to reflect skeptically about accepted explanations of technical matters requires the habit of asking oneself and one’s colleague, “What if we assume the opposite; can there possibly be *any* truth to *that*?”

The fourth weakness of engineering analysis was the failure to frame the foam debris problem more abstractly, to shift attention away from the debris’ lightweight quality toward the more abstract question, “What do we know about the striking power of bodies in motion?” But the imagination to shift from immediate sense data to a more abstract generalization about those same data requires an abstract image like the formula for computing the energy produced by a moving body: $E=MV^2/2$.

That kind of abstract image is necessary but not sufficient, of course, but NASA engineers and managers had at least 17 years to connect foam debris to $E=MV^2/2$, plenty of time for even weak imaginations to stumble on the abstract question. The chief cause of failure in this case, it seems, was a sheer lack of knowledge of physics (or some shared professional prejudice against pure physics?).

We must ask what kinds of interventions or change will have to take place to bring NASA engineering to the deeper knowledge and the higher technical standards required to protect spacecraft and astronaut safety? What interventions will be necessary for *managers* to be sufficiently skilled in technical engineering to pose the necessary challenges to their engineers’ analyses and for the *engineers* to produce coherent flight rationales based on high standards of quantitative evidence?

Each of these weaknesses in knowledge calls out for education and training. What kinds of education and training are required to ensure *updating of new knowledge*? What kind of education at NASA will *extract* the lessons of near misses and then lead to *mastery* of those lessons?

2. Leadership. The second major kind of weakness was in *Columbia's* leadership. The decision to set a deadline for completion of Node 2 (February 19, 2004) in order to demonstrate to Congressional funding leaders NASA's efficiency in meeting flight schedules revealed deep ignorance of the effects of imposing a rapid flight rate on care, on thoughtfulness, and on accuracy, a profound numbness to the many hidden dangers in this infant stage of space exploration (see Fitts,1966; Hollnagel, 2009; and Wickelgren, 1977 for the effects of imposing an unaccustomed time requirement on task completion). Weak leadership also at the operational levels of flight readiness reviews led to lowered standards due, in turn, to managers' own low levels of technical competence. And when a rationale to demonstrate flight readiness was recognized by a leader as "lousy," neither the leader nor those who agreed with "lousy" took steps to correct either that particular FRR or, more importantly, the FRR process leading to that specific rationale. Both conceptual understanding and analytical knowledge and skill had slid to dangerously low levels.

What kinds of intervention, at what organizational levels, will bring to NASA's institutional and operational managers an awareness of the effects of their policies on the performance of workers far below them in the organization? What kinds of intervention and change will provide strong, reliable support for managers and engineers who need time to gather necessary data, time to do proper analysis, and time to assimilate and communicate the meanings of analyses in order to protect safety against unmitigated production? What kinds of intervention and change will it take for institutional managers to understand that operational managers require technical knowledge of engineering if they are to be able to supervise engineers effectively and to hold the space program to high technical standards?

3. A culture of safe production. The third underlying cause of deliberative failure was NASA's unbalanced organizational culture, oriented toward meeting a flight schedule based on an arbitrary completion date. Little or no effort was made toward building a consciousness in managers of the fact that dangers still lay hidden. The high standards of engineering and flight readiness that William Lucas had established and that had permeated Lucas's FRRs had disappeared, and the technical competence of his successors was inferior. The independent safety system that NASA had supposedly instituted after the *Challenger* disaster was barely evident and clearly ineffectual in the *Columbia* deliberations – with the exception of two, Rodriguez and Walker, who raised their voices but whose resistance to approving a weak flight rationale, resistance in the service of safety, was consciously circumvented by managers.

Repeated boards of experts have been convened over the years, with repeated reports issued. But at no time has the NASA leadership called for an *in-house* capacity for studying and correcting its own deliberative and communication weaknesses. For an organization to learn about weaknesses in its own processes its members must be supported, with the help of skilled professionals, to turn analytical abilities upon the actual practices and decisions that have guided members' work. For a space program to avoid the kinds of deficient deliberation evident in the *Columbia* and *Challenger* cases, it must not only develop a robust training program to correct those deficiencies and must not only hire the leadership sufficient to institute corrections, but must also do the more difficult task of recreating itself as a learning organization. That would be a systemic change in NASA's culture.

The shuttle program now gives way to the next phase of space exploration. The debate has been whether we should explore the moon as a launch base for wider space exploration or to go directly to Mars. Present inclinations favor the latter. But no matter how that issue turns out, our next efforts in space will put us back virtually at the beginning stages of the new technology required. Yes, we will have wiser vehicle designs, but they will be addressing far more challenging realities. Not only will the next generation of vehicles be, and continue to be experimental; the organization of engineering and managerial talent to build them will also be experimental – still to be tested against the realities that travel to the moon and Mars will throw up at us.

Just as the shuttle and its successor vehicles must still be regarded as experimental vehicles, so the *space program* must be considered an experiment in how to explore space. The *Challenger* and *Columbia* are still speaking, sending signals that I have been at some pains to spell out (see also Lighthall, 2015). Now NASA's space program must learn how to listen to its own signals, most particularly regarding how it deliberates effectively and ineffectively to reach its decisions, and how it might come to do so better. (A sample of the kinds of expertise I think are needed in NASA's space program is provided in the following endnote. See also Lighthall, 2015, Chapter 9). [57]

The demand for more effective *mind work,* much more educated and incisive deliberation, now has higher stakes. The planet Earth itself is more demanding, putting strains on our resources to cope – with global warming, cultural and religious conflict, domestic productivity and jobs, population growth, education, health care. Space exploration now must compete with other priorities that scream louder than ever. An electorate in an economy that is anything other than booming will have to make tougher choices. Space exploration may not have the same urgency among the electorate generally that it did when Sputnik seemed an imminent threat to our national security.[58] Space disasters like the *Challenger* and *Columbia*, caused by severe deliberative deficiencies, may not lead so readily to refunding an ambitious space program with human crews. When priorities are many and choices tough, if our space exploration is to be sustainable, the decision makers in the space program must bring a much sharper and deeper deliberative capacity to the task of detecting and coping with the dangerous realities of space and of their new vehicles.

End Notes

---

1. See McDonald, A. J. 1989. Return to flight with the redesigned solid rocket motor. AIAA Paper No. 89-2404, 25th Joint Propulsion Conference (AIAA/ASME/SAE/ASEE), Monterey, CA, 10-12 July, 1989.
2. This number represented 1.25 percent of the Orbiter's 24,000 tiles, but the loss of a single tile, depending on where it was located on the wing, could bring disaster.
3. See the discussion of the history of divots in Columbia Accident Investigation Board (2003) *Report* (pp. 121–31). Washington, DC: US Government Printing Office (hereafter, CAIB *Report*).
4. A photograph of the missing "chine" tile and schematic drawing of its location on STS-27R's undersurface can be found at http://spaceflight.nasa.gov/shuttle/rtf/tps/tiledamage.pdf. A notation accompanying this photograph asserts that "insulation and ice were the cause of the damage." The claim of ice debris as the cause is contradicted by CAIB's *Report* (see vol. 1, 127).
5. See Lighthall, (2014) on this website.
6. The CAIB pointed to an earlier flight (STS-7) preceding the *Challenger* accident as providing a danger signal in which a piece of foam much larger than average had come loose. The board identified large chunks of foam that tore loose from a forward strut of external tanks (the "bipod" legs connecting the external tank to the Orbiter) as particularly dangerous, and saw the first instance of such a chunk, from STS-7, as the first clear danger signal warning of the catastrophe that eventually occurred in the *Columbia* accident. The board saw weakness in NASA's failure to determine the *cause* of STS-7's bipod foam loss.

   This first instance of an unusually large chunk of foam loss, however, caused a level of damage well within the range already determined safe, the kind that would necessitate repair but not the kind that would threaten flight safety. When a new condition, like a large chunk of foam debris, causes no particular damage, why would you halt flights to determine and eliminate the cause of that foam loss? It is only in hindsight, from the vantage point of the accident 105 flights later, hindsight from an accident in which a large chunk of foam, *struck by the Orbiter at a relative speed of more than 500 miles per hour*, did contribute to cause an accident (*Columbia*), that you might seize upon "large chunk" as a definitive signal of danger. But "large chunk" was not definitive, since large chunks of foam at slow velocities and small angles of impact would be harmless.

It was only with the flight of STS-27R (*Atlantis*), where a relatively small piece of light- to medium-weight debris was accompanied by clearly dangerous damage, that a flight provided a clear signal of the disastrous danger of foam debris. Paradoxically, while the CAIB focused on STS-7 as a signal that should have triggered investigation of the cause of foam loss, it relegated the far more dangerous damage of flight 27-R to the category of "other foam/debris events"—on the basis, apparently, that the foam that did the damage to 27-R's wing was not a "bipod foam event" but involved ablative foam from a booster nose cone.

STS-27R foreshadowed the *Columbia* accident in the severity and danger of its damage and in the fact that the foam loss from 27-R's booster nose cone came at a point in its ascent, 85 seconds—when shuttle speed had attained a high velocity, generating a high velocity of debris impact—very similar to that of the *Columbia*, which shed its large chunk of foam 81 seconds after launch.

7. The CAIB *Report* gives no third dimension of this piece of foam (vol. 1, 124).

8. Ibid, 124.

9. See CAIB *Report*, vol. 1, 123.

10. The theory behind NASA's poking holes in the foam to vent it was an engineering guess about the cause of foam separation. The venting idea had apparently never been investigated experimentally or with any probing analysis.

11. The language of the CAIB *Report* suggests that a quantitative analysis had been carried out on aerodynamic loads (see CAIB *Report*, vol. 1, 124). Computing the striking energy of an object requires that one quantify all four of the key variables—mass, density, velocity, and angle of impact—and then compare the resulting quantity of impact energy to the tiles' strength to resist that level of impact. The reported language of the Integration Office's certification of readiness implies that the engineers had calculated a quantitative safety margin of safe impact energy and

had determined—by computing the four-variable algorithm—that STS-50's "26 x 10" inch piece of foam had generated a level of impact energy within that safety margin.

Yet no such safety margin had ever been established, either by experiment or by analysis. Extensive review of the CAIB's six-volume *Report* reveals no hint of pre-accident safety margins, for either tiles or RCC panels, no margins for foam momentum, for angle of incidence, or for total impact energy.

If that kind of analysis had been the response to STS-50's damage, it would surely have been employed to assess subsequent foam-loss events, particularly the very large one that led to the *Columbia* disaster. The only analysis carried out in deliberations about the fatal *Columbia* flight was an algorithm of dubious applicability pressed into use from a very different context, the so-called "CRATER" computer model, noted below.

12. Ham and Dittemore both revealed their narrow conception of the debris problem in e-mails sent after *Columbia*'s launch. Ham: "Can we say that for any ET foam lost, no 'safety of flight' damage can occur to the Orbiter because of the density?" (CAIB *Report*, vol. 1, 154; Cabbage and Harwood (2004, 115). Dittemore: "Another thought, we need to make sure that the density of the ET foam cannot damage the tile to where it is an impact to the Orbiter" (CAIB *Report*, vol. 1, 155).

The kinetic energy of any body in motion is equal to the body's mass multiplied by the *square of its net velocity* divided by two. The formula for kinetic energy looks like this:

Ek in foot pounds of energy = $(M_c \text{ x } V^2)/2$

Where

Ek = Kinetic Energy—the striking energy of a body in motion

M_c = Mass converted to foot units of gravitational pull, that is

Mass =Size (in cubic feet) X density (pounds per cubic foot)

Mass_c = Mass/32.2 ft. per second

V = Net Velocity (feet per second), that is, the difference between the velocity of the object being struck and the striking object—for example, the difference between the speed of the shuttle and the speed of a chunk of foam that the shuttle strikes.

I thank Allan McDonald for setting me straight on the correct details of this formula.

13. See the discussion by Cabbage and Harwood (2004, 61–63).

---

14. CAIB Report, Vol. I, 125.

15. The disconnect between a front-line engineer's understanding of crucial dynamics, especially a shift in those dynamics, and the levels of understanding acquired by the engineer's manager, on the other hand, is not a new story (Lighthall, 2015).

16. Cabbage and Harwood (2004, 66).

17. A decision made apparently by the deputy manager of NASA's ET Project Office, Neil Otte.

18. See Cabbage and Harwood (2004, 61–66). Rodriguez relented and signed; Walker refused to do so. The CAIB makes no mention of this episode of dangerous and unprofessional pressure and of professional dissension in its report.

19. The important decision-making narrative provided by Cabbage and Harwood (2004) indicates there were at least four managers with sufficient authority to challenge the flawed FRR reasoning that gave assurance of *Endeavor*'s safety, managers who also were aware that the reasoning supporting *Endeavor*'s launch readiness was below standards: Ron Dittemore, manager of the Space Shuttle Program; Don McCormack, lead manager of NASA's Mission Evaluation Room for the *Columbia* flight; Paul Shack, shuttle manager for NASA; and Linda Ham, NASA's shuttle integration manager and program integration manager for the *Columbia* flight, and chair of *Columbia*'s Mission Management Team (see Cabbage and Harwood 2004, 69–70 and 106–8).

    The weak assessment of the danger posed by *Atlantis*'s foam loss and its damage to its booster, given initially by Jerry Smelser, NASA's external tank manager at Marshall, was later questioned by NASA's Bryan O'Connor, NASA's chief safety and mission assurance officer. O'Connor questioned Smelser whether the size of *Atlantis*'s chunk of foam loss was too small to cause damage, and to Smelser's reply that it was a large chunk but still not a "safety of flight issue," O'Connor wondered whether Smelser wasn't relying simply on the success of past experience. When Smelser replied that there had never been evidence of foam doing more than maintenance damage (a lapse of memory or of judgment regarding the foam-caused damage and near miss of flight 27-R), O'Connor sought information from NASA's *Integrated Hazard Report 37*. In it, he found support for officially regarding any foam hitting the Orbiter as possibly "catastrophic," and therefore improperly being considered merely a matter of maintenance.

    Smelser then agreed with O'Connor that more proper wording would be to view the foam as an "accepted risk to fly" rather than a "non–safety of flight issue." With that semantic reclassification, O'Connor's challenge ended, and no other manager questioned Smelser's rationale for *Endeavor*'s flight readiness (see Cabbage and Harwood's (2004) account of the

O'Connor-Smelser interchange, 68–71).

20. See Cabbage and Harwood's (2004, 101-2) account of Boeing engineers' calculations of the penetrating force of foam.

21. Classifying *Columbia*'s estimated damages as "turn-around" was accomplished in the absence of any quantitative analysis establishing a safety margin that would distinguish between "turn-around" and "burn-through" damage. While the various streams of deliberation by different groups of engineers and managers took place over several days, the view that *Columbia*'s Orbiter was safe to return to earth had crystallized by the ninth day of *Columbia*'s sixteen-day mission.

22. Cabbage and Harwood (2004, 106) quote Linda Ham, NASA's co-chair of *Columbia*'s MMT (apparently from a tape recording of an MMT meeting), as having earlier reached the view that "there is not much we can do about it". Ham later clarified the context of her remarks in a July 22, 2003, press roundtable (see National Aeronautics and Space Administration (2003). *Facts*. Press roundtable STS-107 mission management," July 22, 2003 (hereafter, NASA *Facts*, 2003) 8-9.

23. See Cabbage and Harwood's extended discussion of the efforts of Page and Rocha to pursue the issue of danger (2004, 94–102, 108–9, 113–23, 131–34).

24. CAIB *Report,* Vol. I, 60-61, 143.

25. Allan McDonald related to me in a phone conversation that when he first heard the news from NASA's early reports detailing the estimated speed of the Orbiter and the chunk of foam debris, he worked out the amount of kinetic energy involved in the foam contacting the Orbiter's wing. He could not believe the results of his own calculations, thinking that he must have misplaced a decimal somewhere. Only after repeated calculations did he come to accept his own results.

    When he compared his figure for kinetic energy to ballistics tables, he was astonished to find that NASA's estimates yielded the kinetic energy of a 30.06 rifle bullet. Still dubious, he re-calculated, but came up with the same result. He, too, had a hard time convincing himself that the light weight foam debris could generate such energy. But his working knowledge of physics gave him the advantage, not available to NASA engineers and managers, of seeing the immediate relevance of the formula for kinetic energy, a formula in which a key variable, velocity, was *squared,* revealing how velocity could easily dominate the other variables in determining the potency of kinetic energy.

26. See Swets (1964), Swets, Dawes, & Monahan, (2000), and Wickens (2002); the website

http://en.wikipedia.org/wiki/Detection_theory provides a concise summary of SDT.

27. Another key variable affecting whether a signal will be detected is the perceiver's (engineers' and managers') conscious or unconscious willingness to accept possible risk in order to meet flight schedules or willingness to interrupt flight schedules in order to protect safety. In statistical terms, this is the question whether the perceiver prefers making a Type I error, seeing a flight as dangerous when conditions are actually safe, to a Type II error, seeing a flight as safe when it is actually in danger.

28. The CAIB *Report* identified seven flights with notable Orbiter damage, which would bring the signal-to-noise ratio to .00044, or forty-four red marbles in a barrel of one hundred thousand pink marbles.

29. STS-7, flown June 18, 1983, four and a half years before 27-R, shed a large chunk of bipod foam from its ET, one classified as an IFA. The CAIB *Report* makes no mention of the size of Orbiter damage that may have resulted, but it was judged at the time to be only "turn-around" damage—a judgment, again, based apparently on no quantified safety margin or analysis of kinetic striking energy (see CAIB *Report*, vol. 1, 123).

30. It might be argued that engineers, especially space engineers, with their tendencies to quantify realities, might be equipped to break out of the grip of the noise surrounding signals of danger, might be sensitive to the quantitative differences between the Orbiter damage of STS-27R and STS-50 and all the other merely turn-around damage. Yet that same quantitative orientation would also give weight to the overwhelming number of harmless Orbiter divots as a quantitative reality. It would also lead engineers to pay attention to signals of danger with much less noise elsewhere, like eroding or leaking O-rings in the boosters and similarly demanding anomalies in the Orbiter, main engines, and ET. What is required in detecting the significance of rare anomalies is not merely quantitative sensitivity, but also sensitivity to the signal-to-noise problem itself, whose solution requires *measurement* of small anomalies, however small that might shift from small and innocuous to big and dangerous, where measurements would show that shift.

31. The first post-*Challenger* flight, 26-R, had also provided a signal of danger, described in detail by McDonald & Hanson (2009, 523-525).

32. See Lighthall (2015, chapter 3) for an extended analysis of Lucas's leadership style, his impact on subordinates, and the organizational culture he created at NASA's Marshall Space Flight Center.

33. Cabbage and Harwood (2004, 68). See similar comments by Linda Ham regarding limitations of her own analytical capabilities in NASA *Facts* (2003), 9-10, 17, 23, for example: "I personally

[do not], nor does the MMT, do the analysis. We must rely on our contractor work force... we don't have the tools to do that. We don't have the knowledge to do that or the background or expertise to do that kind of thing" (p. 17). Managers with limited technical knowledge, with limited ability to ask pertinent questions or contemplate alternative analyses or arguments, can hardly hold their subordinates to high engineering standards.

34. Quoted from Cabbage and Harwood's interview of Smelser (2004, 68).

35. Technical weakness is revealed also in the interchange where NASA's chief safety officer, Bryan O'Connor, challenged Smelser's assessment of STS-112's foam loss. O'Connor's challenge focused on a technical *classification* but O'Connor never thought to ask Smelser for an engineering analysis of the striking power of foam debris. See Cabbage and Harwood's account (2004, 68–71).

36. See NASA *Facts* (2003), 9-10, 17, 23.
37. CAIB *Report*, Vol. I, 148.
38. O'Keefe is quoted in Tompkins (2005, 130–31).

39. Cabbage and Harwood (2004, 185).

40. Linda Ham, a key participant in the decisions running up to the *Columbia* accident, focused on the properties of the foam material itself. In explaining the importance of earlier rationales for approving flights in the context of foam shedding, she explained that she was searching for an earlier "rationale *based on the material properties of that foam,* so that even on the bipod ramp, if that foam would come off, that it would do no damage to the Orbiter" (emphasis added).

    Ham was, she said, "hoping" that previous FRR analysis would show a safety margin against foam debris, that analysis showed that there "won't be enough kinetic energy to hurt the Orbiter anywhere." But referring to kinetic energy was far from doing the analysis, or from demanding that others do it. See NASA *Facts*, 2003, 21–22.

41. This groupthink failure extends, apparently, to the CAIB and its staff, no member of which thought to compute the kinetic energy for the *Columbia*'s large chunk of foam debris despite possessing all the data necessary to assess various scenarios.

    The board's focus was apparently on the *size* of *Columbia*'s chunk of debris, and of earlier instances of foam shedding, since it placed so much emphasis on bipod foam loss, which produced the largest pieces of debris. Nowhere does the board's report include a table reporting both size and estimated *velocity* of foam debris for various flights.

42. Often investigation after a disaster discovers an ignored report warning of the dangerous condition or practice that actually caused the disaster. I am waiting for someone to discover such a report in the *Columbia* case, one that includes some computations of energy or striking force and an explicit warning of a possible catastrophe.

43. See the e-mail from L. D. Austin to Linda Ham, CAIB *Report*, vol. 1, 155–56, where Austin summarizes the relevant variables: "the impact damage significance is always a function of debris size and density, impact velocity, and impact angle—these latter two being a function of the flight time at which the ET foam becomes debris."

44. McDonald, who early understood the importance of velocity in these foam strikes, noted that: "NASA folks missed the importance of both mass and relative velocity. They … never recognized that the ramp foam was more than an order of magnitude higher in mass [than all earlier foam debris]. I was totally surprised myself because the piece of cork that allegedly hit STS-26R had an assumed velocity of 1180ft/sec, which was more than 50% higher than the foam that hit the *Columbia.* However, the cork weighed a mere 2.1 oz. compared to 1.67 pounds for the ramp foam [that struck *Columbia*] – the mass was nearly 13 times higher. Therefore, the kinetic energy at impact was 5.5 times higher for *Columbia* even with a 50% lower relative velocity at impact" (Comment by A. J. McDonald in personal correspondence).

45. Some parallels can be found between the deliberations preceding the *Columbia* accident and those preceding *Challenger*. For example, deliberations of both suffered from a red herring. O-ring erosion was the red herring that captured the minds of many in the *Challenger* deliberations, just as the *Columbia* deliberations were captured by the lightweight quality of the foam material. (Even Nobel Prize laureate Richard Feynman's attention repeatedly returned to O-ring erosion, reflected in CAIB's quote of Feynman, vol. 1, 130 and in Feynman's memo, "Personal observations on reliability of shuttle," P C *Report*, vol. II, pp, F-1, 2.) Engineers before the *Challenger* accident had established empirical safety margins for erosion from proper engineering analysis of O-ring impingement erosion in various scenarios of safe and dangerous sealing of the joints (a fact missed by Feynman). In contrast, engineers and managers in deliberations from STS 26-R onward substituted unexamined assumptions for engineering analysis regarding the dangers of foam loss.

46. Unattributed quotes of NASA workers, see CAIB *Report*, vol. 1, 134.

47. See Cabbage and Harwood (2004, 108–123) for deliberations and actions on days six through eight of the seventeen-day space mission.

48. The collective search for evidence of danger, even on its path of dangerous thinking, did reveal errors of judgment and illogic, errors that might well have been corrected with time and further discussion. Both engineers and managers could engage in twisted thinking. In response to film analysts' need for images of damage, Trish Petete, Paul Shack, and Calvin Schomburg took the position that it was better to wait for the damage assessment team to report their findings before deciding if DOD satellite photos were necessary. It apparently never occurred to them that the very photos they were postponing might be crucial for assessing whether damage was severe enough to warrant obtaining the photos!

So Rocha and others seeking orbital photos were in the position of having to prove sufficient damage on the basis of insufficient films to justify further photos that could show whatever damage had occurred. Not only was the burden of proof on them to prove damage. They were also required to do so with evidence from films whose insufficiencies they had pointed out (see Cabbage and Harwood 2004, 113–14; CAIB *Report*, vol. 1, 156–57).

49. Cabbage and Harwood (2004, 97–98).

50. This category, "turnaround damage," and its contrasting twin, "safety-of-flight concern," provided engineers and managers an easy mechanism for substituting qualitative thinking for quantitative thinking, allowing simple judgment to replace quantitative analysis of the actual properties of foam debris. It allowed the whole range of quantitative differences (in size and velocity) among the many instances of foam debris to be reduced to these two categories. This simplification promoted a blindness to actual variation in degrees of danger that surely qualifies as "normalization of deviance" (Vaughan, 1996), a case of assimilating repeated occurrences of foam debris into the established schemas (in Piaget's terms) of a) turnaround issue only, and b) another piece of Styrofoam-like debris floating off a pickup truck on a highway – a distraction, but nothing dangerous.

51. See Cabbage and Harwood's (2004, 120-23) account of the exchange and of the deliberations of day eight.

52. The CAIB *Report* (173–74) estimated that the crew's food, water, and oxygen would be sufficient if a rescue mission reached the *Columbia* in orbit no later than flight day 30 (February 15).

The CAIB's own analysis reveals a shaky understanding of the relevance of burden of proof to protecting safety. At points it seems to comprehend, as when it states "The engineers found themselves in the unusual position of having to prove that the situation was unsafe – a reversal of the usual requirement to prove that a situation is *safe"* (p.169, italics in original). In the CAIB's summary of MMT decision making, however, the CAIB frames the focal question to be answered as one of determining if the situation was dangerous: "…managers failed to avail themselves of the wide range of expertise and opinion necessary to achieve the best answer to the debris strike question -- '*Was this a safety-of-flight concern?*'" (p. 170, italics in original). While the CAIB does include the reversal of the "usual" burden of proof as the twenty-second among its twenty-nine Findings, it makes no reference to decision rules, modes of argument, or burden of proof in its recommendations.

53. See CAIB *Report*, vol. 1, 173-74 for the CAIB's rescue and repair scenario.

54. Lucas's Marshall Center Board was sometimes referred to as the "Level 2-1/2" FRR, since it followed Marshall's Level III FRR and preceded the Level II FRR at the Johnson Center.

55. A diagrammatic view of the post-*Challenger* FRR system—renamed the "Milestone Review Process"—can be found in Figure 6.2.1-1 of the "External Tank Working Group Final Report," CAIB *Report*, vol. 4, 65.

56. Compare, for example, the scientific and engineering education of William Lucas as director of Marshall (master's degree in chemistry, PhD in metallurgy) with that of Arthur Stephensen as director of Marshall in the four years preceding the *Columbia* accident (bachelor's degree in electrical engineering).

57. In addition to the disciplines of physics and engineering, improvement of NASA's deliberations about safe and risky flight requires the perspectives, knowledge, and skills of many other disciplines, including psychology, social psychology, sociology, anthropology, and especially, education.

Active in-house and visiting consultants from these disciplines are needed to develop in NASA's engineers and managers the analytical insights and skills reflected in the works of Argyris and Schön (1994), John S. Carroll and colleagues (Carroll, Perin, & Marcus, 1991; Carroll, Randolph, & Hatakenaka, 2002), Fischhoff and colleagues (Fischhoff, et al., 1983), Gaskins (1995), Hale and colleagues (Hale, Wilpert, & Freitag, 1997), Hammond (1995, 2000), Helmreich and Merritt (1998), Hutchins (1995), Jarvis (1987, 2006), Kahneman & Tversky (1974), Klein and colleagues (Klein, 1993, 1997, 2003), Paté-Cornell (1990), Reason (1990), Roberts and colleagues (Roberts, K. H., 1993), Rogers (2006), Schein (1993), Swets and colleagues (2000), Weick and colleagues (Weick & Sutcliffe, 2001), Wilpert and colleagues (Wilpert & Qvale, 1993), Woods and colleagues (Zelik, Patterson, & Woods, 2007), and Zohar & Luria (2003, 2004) -- as a non-exhaustive sample of expertise needed in the space program's improvement of safety.

58. Starbuck and Stephenson (2005) address NASA's more vulnerable situation: "Taxpayer support will be critical to NASA's future efforts and NASA should begin now to explain better how the billions of dollars given NASA each year ultimately benefit ordinary Americans" (p. 332) and "NASA has generally paid too much attention to outer space and not enough attention to Earth" (p.333).

References

Argyris, C., & Schön, D. (1994). *Theory in practice: Increasing professional effectiveness.* San Francisco, CA: Jossey-Bass.

Cabbage, M., & Harwood, W. (2004). *Comm check… The final flight of shuttle* Columbia. New York: Free Press.

Carroll, J. S., Perin, C., & Marcus, A. A. (1991, December). Organizational learning at nuclear power plants. *Report of the panel meeting, nuclear power plant advisory panel, on organizational learning,* MIT Sloan School of Management, Cambridge, MA, October 17–18.

Carroll, J. S., Rudolph, J. W., & Hatakenaka, S. (2002). *Organizational learning from experience in high-hazard industries: Problem investigation as off-line reflective practice* (Working Paper 4359-02). Cambridge: MIT Sloan School of Management. Available at: http://ssrn.com/abstract_id=305718

Columbia Accident Investigation Board (2003). *Report,* Volumes I-VI, Washington, DC: US Government Printing Office.

Fischhoff, B., Slovik, P., Lichtenstein, S., Read, S., & Combs, B. (1978). How safe is safe enough? A psychometric study of attitudes toward technological risks and benefits. *Policy sciences, 9,* 127–152.

Gaskins, R. H. (1995). *Burdens of proof in modern discourse.* New Haven, CT: Yale University Press.

Hale, A. R., Wilpert, B., & Freitag, M. (Eds.) (1997). *After the event: From accident to organizational learning.* Howard House, Wagon Lane, Bringley, UK: Emerald Group Publishing.

Hammond, K. R. (2000). *Judgment under stress*. New York, NY: Oxford University Press.

Helmreich, R.L., & Merritt, A.C. (1998). *Culture at work in aviation and medicine: National, organizational, and professional influences*. Aldershot, U.K.: Ashgate.

Hollnagel, E. (2009). The ETTO Principle: Efficiency-thoroughness trade-off—Why things that go right sometimes go wrong. Burlington, VT: Ashgate.

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

Janis, I. L. (1972). *Victims of groupthink.* Boston, MA: Houghton Mifflin.

Jarvis, P. (1987). *Adult learning in the social context*. Beckenham, England: Croom Helm.

Jarvis, P. (2006). *Towards a comprehensive theory of human learning*. New York, NY: Routledge.

Kahneman, D. & Tversky, A. (1974). Judgment under uncertainty: Heuristics and biases, *Science 185*, 1124-1131.

Klein, G. A. (2003). *Intuition at work.* New York: Doubleday.

Klein, G. A. (1993). A recognition-primed decision (RPD) model for rapid decision making. In G. Klein, J. Orasanu, R. Calderwood, & C. E. Zsambok (Eds.), *Decision making in action: Models and methods* (pp. 138–147). Norwood, NJ: Ablex.

Klein, G. A. (1997). An overview of naturalistic decision making applications. In C. E. Zsambok & G. Klein (Eds.), *Naturalistic decision making* (pp. 49–69). Mahwah, NJ: Lawrence Erlbaum.

Lighthall, F. F. (2014). A high-tech "near miss": Organizational decision making up close. www.high-techdanger.com

Lighthall, F. F. (2015). *Disastrous High-Tech Decision Making: From Disasters to Safety*. Indianapolis, IN: Kilburn Sackett Press.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of personality and social psychology, 37*(11), 2098-2108.

McDonald, A. J. (1989). Return to flight with the redesigned solid rocket motor. AIAA Paper No. 89-2404, 25th Joint Propulsion Conference (AIAA/ASME/SAE/ASEE), Monterey, CA, 10-12 July, 1989.

McDonald, A. J. & Hanson, J. R. (2009). *Truth, lies, and O-Rings: Inside the space shuttle* Challenger *disaster.* Gainesville: University Press of Florida.

National Aeronautical and Space Administration (2003). Press roundtable STS-107 mission management. *Facts*, July 22, 2003 (hereafter, NASA *Facts*, 2003) 8-9.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology, **2*** (2): 175–220.

Paté-Cornell, M. E. (1990). Organizational aspects of engineering system safety: The case of offshore platforms. *Science, 250,* 1210–1217.

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

Roberts, R. C., Flin, R., & Cleland, J. (2015). Staying in the zone: Offshore drillers' situation awareness. *Human factors,* 57(4), 573-590.

Rogers, E. W. (2006). Introducing the pause and learn (PaL) process: Adapting the army after-action review process to the NASA Project World at the Goddard Space Flight Center. (White paper circulated at Goddard Space Flight Center, revised June 8, 2006). Available at: www.nasa.gov/centers/goddard/pdf/287922main_PaLwhitepaperV3.pdf

Schein, E. (1993). On dialogue, culture, and organizational learning. *Organizational dynamics, 22* (2), 40–51.

Swets, J. A. (Ed.). (1964). *Signal detection and recognition by human observers*. New York, NY: Wiley.

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological science in the public interest, 1*(1), 1-26.

Tompkins, P. K. (2005). *Apollo, Challenger, Columbia: The decline of the space program*. Los Angeles, CA: Roxbury.

Vaughan, D. (1996). *The* Challenger *launch decision: Risky technology, culture, and deviance at NASA.* Chicago, IL: University of Chicago Press.

Weick, K. E., & Sutcliffe, K. M. (2001). *Managing the unexpected: Assuring high performance in an age of complexity.* San Francisco, CA: Jossey-Bass.

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. Acta psychologica, *41*, 67–85.

Wickens, T. D. (2002). *Elementary signal detection theory*. New York, NY: Oxford University Press.

Wilpert, B. & Qvale, T. U. (Eds.) (1993/2013). *Reliability and safety in hazardous work systems: Approaches to analysis and design.* Mahwah, NJ: Lawrence Erlbaum.

Woods, D. D. (2006). Essential characteristics of resilience. In E. Hollnagel, D. D. Woods, & N. Leveson (Eds.), *Resilience engineering: Concepts and precepts* (pp. 21-34). Burlington, VT: Ashgate

Zelik, D., Patterson, E. S., & Woods, D. D. (2007). *Understanding rigor in information analysis.* Paper presented at the 8th International Conference on Naturalistic Decision Making, Pacific Grove, CA.

Zohar, D., & Luria, G. (2003). Organizational meta-scripts as a source of high-reliability: The case of an army armored brigade. *Journal of organizational behavior*, *24*, 837–859.

Zohar, D., & Luria, G. (2004). Climate as a social-cognitive construction of supervisory safety practices: Scripts as proxy of behavior patterns. *Journal of applied psychology*, *89***,** 322–333.