(published in: Kleinman, Cloud-Hansen, Matta, and Handelsman (editors), *Controveries in Science and Technology Volume 2*, Mary Ann Liebert Press, 2008.)

Technical and Managerial Factors in the NASA Challenger and Columbia Losses: Looking Forward to the Future

Nancy G. Leveson, MIT

The well-known George Santayana quote, "Those who cannot learn from history are doomed to repeat it"[1] seems particularly apropos when considering NASA and the manned space program. The Rogers Commission study of the Space Shuttle Challenger accident concluded that the root cause of the accident was an accumulation of organizational problems.[2] The commission was critical of management complacency, bureaucratic interactions, disregard for safety, and flaws in the decision-making process. It cited various communication and management errors that affected the critical launch decision on January 28, 1986, including a lack of problem-reporting requirements; inadequate trend analysis; misrepresentation of criticality; lack of adequate resources devoted to safety; lack of safety personnel involvement in important discussions and decisions; and inadequate authority, responsibility, and independence of the safety organization.

Despite a sincere effort to fix these problems after the Challenger loss, seventeen years later almost identical management and organizational factors were cited in the Columbia Accident Investigation Board (CAIB) report. These are not two isolated cases. In most of the major accidents in the past 25 years (in all industries, not just aerospace), technical information on how to prevent the accident was known and often even implemented. But in each case, the potential engineering and technical solutions were negated by organizational or managerial flaws.

Large-scale engineered systems are more than just a collection of technological artifacts.[3] They are a reflection of the structure, management, procedures, and culture of the engineering organization that created them. They are also, usually, a reflection of the society in which they were created. The causes of accidents are frequently, if not always, rooted in the organization—its culture, management, and structure. Blame for accidents is often placed on equipment failure or operator error without recognizing the social, organizational, and managerial factors that made such errors and defects inevitable. To truly understand why an accident occurred, it is necessary to examine these factors. In doing so, common causal factors may be seen that were not visible by looking only at the direct, proximal causes. In the case of the Challenger loss, the proximal cause[4] was the failure of an O-ring to control the release of propellant gas (the O-ring was designed to seal a tiny gap in the field joints of the solid rocket motor that is created by pressure at ignition). In the case of Columbia, the proximal cause was very different—insulation foam coming off the external fuel tank and hitting and damaging the heat-resistant surface of the orbiter. These

---

[1] George Santayana, *The Life of Reason*, 1905.

[2] William P. Rogers (Chair), *Report of the Presidential Commission on the Space Shuttle Challenger Accident*, U.S. Government Accounting Office, Washington, D.C., 1986.

[3] Nancy G. Leveson, *Safeware*, Addison-Wesley Publishers, 1995.

[4] Although somewhat simplified, the *proximal* accident factors can be thought of as those events in a chain of linearly-related events leading up to an accident. Each event is directly related in the sense that if the first event had not occurred, then the second one would not have, e.g., if a source of ignition had not been present, then the flammable mixture would not have exploded. The presence of the ignition source is a proximal cause of the explosion as well as the events leading up to the ignition source becoming present. The *systemic* factors are those that explain why the chain of proximal events occurred. For example, the proximal cause might be a human error, but that does not provide enough information to avoid future accidents (although it does provide someone to blame). The systemic causes of that human error might point to poor system design, complacency, inadequate training, the culture values around work, etc.

proximal causes, however, resulted from the same engineering, organizational and cultural deficiencies, and they will need to be fixed before the potential for future accidents can be reduced.

This essay examines the technical and organizational factors leading to the Challenger and Columbia accidents and what we can learn from them. While accidents are often described in terms of a chain of directly related events leading to the loss, examining this chain does not explain why the events themselves occurred. In fact, accidents are better conceived as complex processes involving indirect and non-linear interactions among people, societal and organizational structures, engineering activities, and physical system components.[5] They are rarely the result of a chance occurrence of random events, but usually result from the migration of a system (organization) toward a state of high risk where almost any deviation will result in a loss. Understanding enough about the Challenger and Columbia accidents to prevent future ones, therefore, requires not only determining what was wrong at the time of the losses, but also why the high standards of the Apollo program deteriorated over time and allowed the conditions cited by the Rogers Commission as the root causes of the Challenger loss and why the fixes instituted after Challenger became ineffective over time, i.e., why the manned space program has a tendency to migrate to states of such high-risk and poor decision-making processes that an accident becomes almost inevitable.

One way of describing and analyzing these dynamics is to use a modeling technique, developed by Jay Forrester in the 1950s, called System Dynamics. System dynamics is designed to help decision makers learn about the structure and dynamics of complex systems, to design high leverage policies for sustained improvement, and to catalyze successful implementation and change. Drawing on engineering control theory, system dynamics involves the development of formal models and simulators to capture complex dynamics and to create an environment for organizational learning and policy design.[6]

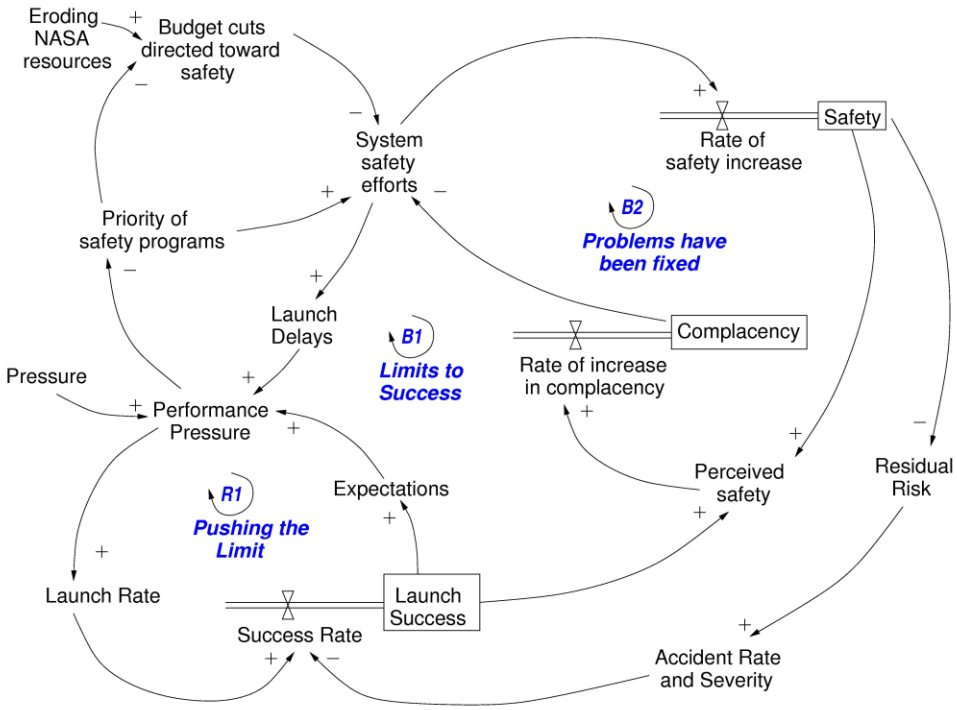


Figure 1. A Simplified Systems Dynamics Model of the NASA Manned Space Program

[5] Nancy G. Leveson, *System Safety: Back to the Future*, unpublished book draft, downloadable from http://sunnyday.mit.edu/book2.html , 2006

[6] John Sterman, *Business Dynamics: Systems Thinking and Modeling for a Complex World*, McGraw-Hill, 2000.

Figure 1 shows a simplified system dynamics model of the NASA manned space program. Although a simplified model is used for illustration in this paper, we have a much more complex model with several hundred variables that we are using to analyze the dynamics of the NASA manned space program.[7] The loops in Figure 1 represent feedback control loops where the "+" and "–" on the loops represent the relationship (positive or negative) between state variables: a "+" means the variables change in the same direction while a "–" means they move in opposite directions. There are three main variables in the model: safety, complacency, and success in meeting launch rate expectations. The model will be explained in the rest of the paper, which examines four general factors that played an important role in the accidents: the political and social environment in which decisions were made, the NASA safety culture, the NASA organizational structure, and the safety engineering practices in the manned space program.

Political and Social Factors

All engineering efforts take place within a political, social, and historical context that has a major impact on the technical and operational decision-making. Understanding the context in which decisions were made in the manned space program helps in explaining why bright and experienced engineers made what turned out to be poor decisions and what might be changed to prevent similar accidents in the future.

In the case of the Space Shuttle, political and other factors contributed to the adoption of a vulnerable design during the original approval process. Unachievable promises were made with respect to performance in order to keep the manned space flight program alive after Apollo and the demise of the cold war. While these performance goals even then seemed unrealistic, the success of the Apollo program and the *can-do* culture that arose during it—marked by tenacity in the face of seemingly impossible challenges—contributed to the belief that these unrealistic goals could be achieved if only enough effort were expended. Performance pressures and program survival fears gradually led to an erosion of the rigorous processes and procedures of the Apollo program as well as the substitution of dedicated NASA staff with contractors who had dual loyalties.[8] The Rogers Commission report on the Challenger accident concluded:

> The unrelenting pressure to meet the demands of an accelerating flight schedule might have been adequately handled by NASA if it had insisted upon the exactingly thorough procedures that were its hallmark during the Apollo program. An extensive and redundant safety program comprising interdependent safety, reliability, and quality assurance functions existed during and after the lunar program to discover any potential safety problems. Between that period and 1986, however, the program became ineffective. This loss of effectiveness seriously degraded the checks and balances essential for maintaining flight safety.[9]

The goal of this essay is to provide an explanation for why the loss of effectiveness occurred so that the pattern can be prevented in the future.

The Space Shuttle was part of a larger Space Transportation System concept that arose in the 1960's when Apollo was in development. The concept originally included a manned Mars expedition, a space station in lunar orbit, and an Earth-orbiting station serviced by a reusable ferry, or Space Shuttle. The funding required for this large an effort, on the order of that provided for Apollo, never materialized, and the concept was scaled back until the reusable Space Shuttle, earlier only the transport element of a broad transportation system, became the focus of NASA's efforts. In addition, to maintain its funding, the Shuttle had to be sold as performing a large number of tasks, including launching and servicing satellites,

---

[7] Nancy G. Leveson, Nicolas Dulac, Betty Barrett, John Carroll, Joel Cutcher-Gershenfeld, and Stephen Friedenthal, *Risk Analysis of NASA Independent Technical Authority*, unpublished research report, downloadable from http://sunnyday.mit.edu/ITA-Risk-Analysis.doc

[8] Nancy Leveson, Joel Cutcher-Gershenfeld, Betty Barrett, Alexander Brown, John Carroll, Nicolas Dulac, Lydia Fraile, and Karen Marais, *Effectively Addressing NASA's Organizational and Safety Culture" Insights from Systems Safety and Engineering Systems*, Engineering Systems Division Symposium, MIT, March 29-31, 2004.

[9] Rogers, ibid, p. 152

that required compromises in the design. The compromises contributed to a design that was more inherently risky than was necessary. NASA also had to make promises about performance (number of launches per year) and cost per launch that were unrealistic. An important factor in both accidents was the pressures exerted on NASA by an unrealistic flight schedule with inadequate resources and by commitments to customers. The nation's reliance on the Shuttle as its principal space launch capability, which NASA sold in order to get the money to build the Shuttle, created a relentless pressure on NASA to increase the flight rate to the originally promised 24 missions a year.

Budget pressures added to the performance pressures. Budget cuts occurred during the life of the Shuttle, for example amounting to a 40% reduction in purchasing power over the decade before the Columbia loss. At the same time, the budget was occasionally raided by NASA itself to make up for overruns in the International Space Station program. The later budget cuts came at a time when the Shuttle was aging and costs were actually increasing. The infrastructure, much of which dated back to the Apollo era, was falling apart before the Columbia accident. In the past 15 years of the Shuttle program, uncertainty about how long the Shuttle would fly added to the pressures to delay safety upgrades and improvements to the Shuttle program infrastructure.

Budget cuts without concomitant cuts in goals led to trying to do too much with too little. NASA's response to its budget cuts was to defer upgrades and to attempt to increase productivity and efficiency rather than eliminate any major programs. By 2001, an experienced observer of the space program described the Shuttle workforce as "The Few, the Tired".[10]

NASA Shuttle management also had a belief that less safety, reliability, and quality assurance activity would be required during routine Shuttle operations. Therefore, after the successful completion of the orbital test phase and the declaration of the Shuttle as "operational," several safety, reliability, and quality assurance groups were reorganized and reduced in size. Some safety panels, which were providing safety review, went out of existence entirely or were merged.

One of the ways to understand the differences between the Apollo and Shuttle programs that led to the loss of effectiveness of the safety program is to use the system dynamics model in Figure 1.[11] The control loop in the lower left corner of the model, labeled R1 or *Pushing the Limit*, shows how as external pressures increased, performance pressure increased, which led to increased launch rates and success, which in turn led to increased expectations and increasing performance pressures. The larger loop B1 is labeled *Limits to Success* and explains how the performance pressures led to failure. The upper left loop represents part of the safety program dynamics. The external influences of budget cuts and increasing performance pressures reduced the priority of system safety practices and led to a decrease in system safety efforts.

The safety efforts also led to launch delays, which produced increased performance pressures and more incentive to reduce the safety efforts. At the same time, problems were being detected and fixed, which led to a belief that *all* the problems would be detected and fixed (and that the most important ones had been) as depicted in loop B2 labeled *Problems have been Fixed*. The combination of the decrease in system safety program priority leading to budget cuts in the safety activities along with the complacency denoted in loop B2, which also contributed to the reduction of system safety efforts, eventually led to a situation of (unrecognized) high risk where despite effort by the operations workforce, an accident became almost inevitable.

One thing not shown in the simplified model is that delays can occur along the arrows in the loops. While reduction in safety efforts and lower prioritization of safety concerns may eventually lead to

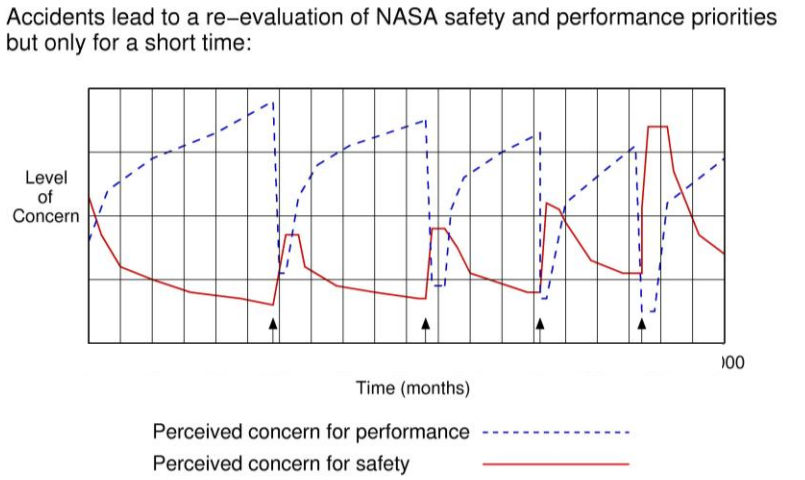
[10] Harold Gehman (Chair), *Columbia Accident Investigation Report*, U.S. GAO, August 2003.

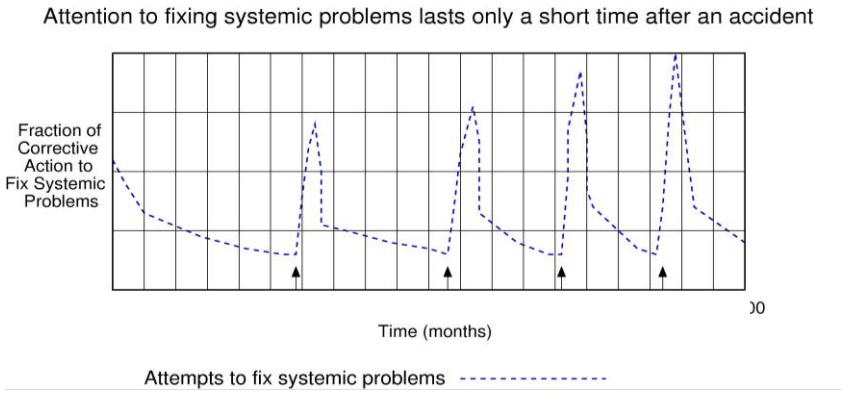
[11] Many other factors are contained in the complete model, including system safety status and prestige; shuttle aging and maintenance; system safety resource allocation; learning from incidents; system safety knowledge, skills, and staffing; and management perceptions. More information about this model can be found in Nancy Leveson, Nicolas Dulac, David Zipkin, Joel Cutcher-Gershenfeld, John Carroll, and Betty Barrett, "Engineering Resilience into Safety Critical Systems" in Erik Hollnagel, David Woods, and Nancy Leveson (eds.), *Resilience Engineering*, Ashgate Publishing, 2006.

accidents, accidents do not occur for a while so false confidence is created that the reductions are having no impact on safety. Pressures increase to reduce the safety program activities and priority even further as the external performance and budget pressures mount, leading almost inevitably to a major accident.

Figure 2 shows the outputs from the simulation of our complete NASA system dynamics model. The upward pointing arrows on the X-axis represent the points in time when accidents or serious incidents occur during the simulation. Despite sincere attempts to fix the problems, the dysfunctional dynamics return very quickly after an accident using our model (and appear to also return quickly in the actual Shuttle program) if the factors underlying the drift toward high risk are not countered. In the top graph, the arrows represent the occurrence of accidents over time. Note that while safety becomes of higher priority than performance for a very short time after an accident, performance quickly resumes its position of greater importance. The middle graph shows that concern about fixing the systemic problems that led to an accident also lasts only a short time after the accident. Finally, the bottom graph shows that the responses to accidents do not reduce risk significantly due to the first two patterns illustrated plus others in the model.

One of the uses for such a model is to hypothesize changes in the organizational dynamics that might prevent this type of cyclical behavior. For example, preventing the safety engineering priorities and activities from being subject to the performance pressures might be achieved by anchoring the safety efforts outside the Shuttle program, e.g., by establishing and enforcing NASA-wide safety standards and not providing the Shuttle program management with the power to reduce the safety activities. This independence did not and still does not exist in the Shuttle program although the problem is recognized and attempts are being made to fix it.





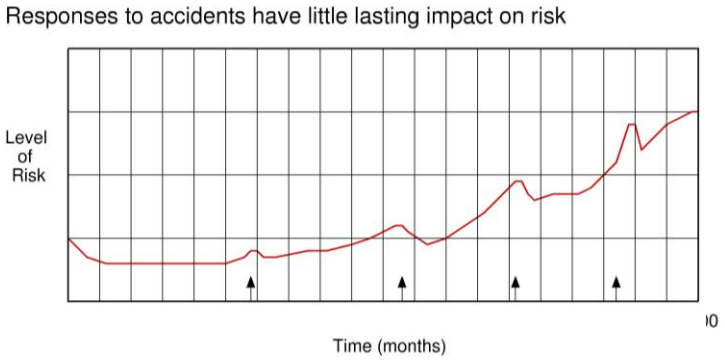


Figure 2. Results from Running the NASA System Dynamics Model

Safety Culture

A significant change in NASA after the Apollo era has been in the safety culture. Much of both the Rogers' Commission Report and the CAIB report are devoted to flaws in the NASA safety culture. The cultural flaws at the time of the Challenger accident either were not fixed or reoccurred before the Columbia loss. Many still exist in NASA today.

A *culture* can be defined as a shared set of norms and values. It includes the way we look at and interpret the world and events around us (our mental model) and the way we take action in a social context. *Safety culture* is that subset of an organizational or industry culture that reflects the general attitude and approaches to safety and risk management. It is important to note that trying to change culture and the behavior resulting from it without changing the environment in which it is embedded is doomed to failure. Superficial fixes that do not address the set of shared values and social norms, as well as deeper underlying assumptions, are likely to be undone over time.[12] Perhaps this partially explains why the changes at NASA after the Challenger accident intended to fix the safety culture, like the safety activities themselves, were slowly dismantled or became ineffective. Both the Challenger accident report and the CAIB report, for example, note that system safety was "silent" and ineffective at NASA despite attempts to fix this problem after the Challenger accident. Understanding the pressures and other influences that have twice contributed to a drift toward an ineffective NASA safety culture is important in creating an organizational infrastructure and environment that will resist pressures against applying good safety engineering practices and procedures in the future.

Risk and occasional failure has always been recognized as an inherent part of space exploration, but the way the inherent risk is handled at NASA has changed over time. In the early days of NASA and during the Apollo era, the belief was prevalent that risk and failure were normal aspects of space flight. At the same time, the engineers did everything they could to reduce it.[13] People were expected to speak up if they had concerns, and risks were debated vigorously. "What if" analysis was a critical part of any design and review procedure. Some time between those early days and the Challenger accident, the culture changed drastically. The Rogers Commission Report includes a chapter titled the "Silent Safety Program." Those on the Thiokol task force appointed to investigate problems that had been occurring with the O-rings on flights prior to the catastrophic Challenger flight complained about lack of management support and cooperation for the O-ring team's efforts. One memo started with the word

[12] Edgar Shein, *Organizational Culture and Leadership*, 2nd Edition, Safe Publications, 1986.
[13] Howard McCurdy, *Inside NASA: High Technology and Organizational Change in the U.S. Space Program*, Johns Hopkins University Press, October 1994.

"HELP!" and complained about the O-ring task force being "constantly delayed by every possible means;" the memo ended with the words "This is a red flag."[14]

The CAIB report notes that at the time of the Columbia loss "Managers created huge barriers against dissenting opinions by stating preconceived conclusions based on subjective knowledge and experience, rather than solid data." An indication of the prevailing culture at the time of the Columbia accident can be found in the reluctance of the debris assessment team—created after the launch of Columbia to assess the damage caused by the foam hitting the wing—to adequately express their concerns. Members told the CAIB that "by raising contrary points of view about Shuttle mission safety, they would be singled out for possible ridicule by their peers and managers."[15]

In an interview shortly after he became Center Director at the NASA Kennedy Space Center after the Columbia loss, Jim Kennedy suggested that the most important cultural issue the Shuttle program faces is establishing a feeling of openness and honesty with all employees where everybody's voice is valued.[16] Statements during the Columbia accident investigation and anonymous messages posted on the NASA Watch web site document a lack of trust leading to a reluctance of NASA employees to speak up. At the same time, a critical observation in the CAIB report focused on the managers' claims that they did not hear the engineers' concerns. The report concluded that not hearing the concerns was due in part to the managers not asking or listening. Managers created barriers against dissenting opinions by stating preconceived conclusions based on subjective knowledge and experience rather than on solid data. In the extreme, they listened to those who told them what they wanted to hear. Just one indication of the atmosphere existing at that time were statements in the 1995 Kraft report that dismissed concerns about Shuttle safety by labeling those who made them as being partners in an unneeded "safety shield" conspiracy."[17] This accusation of those expressing safety concerns as being part of a "conspiracy" is a powerful demonstration of the attitude toward system safety at the time and the change from the Apollo era when dissent was encouraged and rewarded.

A partial explanation for this change was that schedule and launch pressures in the Shuttle program created a mindset that dismissed all concerns, leading to overconfidence and complacency. This type of culture can described as a *culture of denial* where risk assessment is unrealistic and credible risks and warnings are dismissed without appropriate investigation. Managers begin to listen only to those who provide confirming evidence that supports what they want to hear. Neither Thiokol nor NASA expected the rubber O-rings sealing the joints to be touched by hot gases during motor ignition, much less to be partially burned. However, as tests and then flights confirmed damage to the sealing rings, the reaction by both NASA and Thiokol was to increase the amount of damage considered "acceptable." At no time did management either recommend a redesign of the joint or call for the Shuttle's grounding until the problem was solved. The Rogers Commission found that the Space Shuttle's problems began with a faulty design of the joint and increased as both NASA and Thiokol management first failed to recognize the problem, then failed to fix it when it was recognized, and finally treated it as an acceptable risk.

NASA and Thiokol accepted escalating risk apparently because they "got away with it last time."[18] Morton Thiokol did not accept the implication of tests early in the program that the design had a serious and unanticipated flaw. NASA management did not accept the judgment of its engineers that the design was unacceptable, and as the joint problems grew in numbers and severity, they were minimized in management briefings and reports. Thiokol's stated position was that "the condition is not desirable, but it is acceptable."[19] As Feynman observed, the decision making was

---

[14] Rogers, ibid.
[15] Gehman, ibid, p. 169
[16] Leveson, ibid, March 2004
[17] Christopher Kraft, Report of the Space Shuttle Management Independent Review, Feb. 1995, available online at http://www.fas.org/spp/kraft.htm
[18] Rogers, ibid.
[19] Leveson, 1995. ibid.

> "a kind of Russian roulette… [The Shuttle] flies [with O-ring erosion] and nothing happens. Then it is suggested, therefore, that the risk is no longer so high for the next flights. We can lower our standards a little bit because we got away with it last time."[20]

Every time an incident occurred that was a narrow escape, it confirmed for many the idea that NASA was a tough, can-do organization with high intact standards that precluded accidents.[21] The exact same phenomenon occurred with the foam shedding, which had occurred during the life of the Shuttle but had never, prior to the Columbia loss, caused serious damage.

A NASA study report in 1999 concluded that the Space Shuttle Program was using previous success as a justification for accepting increased risk.[22] The practice continued despite this and other alarm signals. William Readdy, head of the NASA Manned Space Program, for example, in 2001 wrote that "The safety of the Space Shuttle has been dramatically improved by reducing risk by more than a factor of five."[23] It is difficult to imagine where this number came from as safety upgrades and improvements had been deferred while, at the same time, the infrastructure continued to erode. The unrealistic risk assessment was also reflected in the 1995 Kraft report, which concluded that "the Shuttle is a mature and reliable system, about as safe as today's technology will provide."[24] A recommendation of the Kraft report was that NASA should "restructure and reduce overall safety, reliability, and quality assurance elements."

The CAIB report identified a perception that NASA had overreacted to the Rogers Commission recommendations after the Challenger accident, for example, believing that the many layers of safety inspections involved in preparing a Shuttle for flight had created a bloated and costly safety program. Reliance on past success became a substitute for sound engineering practices and for accepting increasing risk. Either the decision makers did not have or they did not use inputs from system safety engineering. "Program management made erroneous assumptions about the robustness of a system based on prior success rather than on dependable engineering data and rigorous testing."[25]

Many analysts have faulted NASA for missing the implications of the Challenger O-ring trend data. One sociologist, Diane Vaughan, went so far as to suggest that the risks had become seen as "normal."[26] In fact, the engineers and scientists at NASA were tracking thousands of potential risk factors.[27] It was not a case that some risks had come to be perceived as normal (a term that Vaughan does not define), but that they had come to be seen as *acceptable* without adequate data to support that conclusion.

Edwin Tufte, famous for his visual displays of data, analyzed the way the O-ring temperature data were displayed at the meeting where the Challenger launch decision was made, arguing that they had minimal impact because of their physical appearance.[28]

While the insights into the display of data are instructive, it is important to recognize that both the Vaughan and the Tufte analyses are easier to do in retrospect. In the field of cognitive engineering, this common mistake has been labeled "hindsight bias:"[29] it is easy to see what is important in hindsight. It is much more difficult to achieve this goal before the important data has been identified as critical after the accident. Decisions need to be evaluated in the context of the information available at the time the decision is made along with the organizational factors influencing the interpretation of the data and the decision-making process itself. Risk assessment is extremely difficult for complex, technically advanced systems such as the Space Shuttle. When this engineering reality is coupled with the social and political

---

[20] Rogers, ibid, p. 148

[21] Leveson, 2003, ibid.

[22] Harry McDonald (Chair), Shuttle Independent Assessment Team (SIAT) Report, NASA, February 2000.

[23] Gehman, ibid, p. 101.

[24] Kraft, ibid

[25] Gehman, ibid, 184.

[26] Diane Vaughan, *The Challenger Launch Decision*, University of Chicago Press, Chicago, 1997.

[27] There are over 20,000 critical items on the Space Shuttle (i.e., items whose failure could lead to the loss of the Shuttle) and at the time of the Columbia loss over 3000 waivers existed.

[28] Edward Tufte, *The Cognitive Style of PowerPoint*, Graphics Press, 2003.

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

pressures existing at the time, the emergence of a culture of denial and overoptimistic risk assessment is not surprising.

Shuttle launches are anything but routine, so that new interpretations of old data or of new data will always be needed, that is, risk assessment for systems with new technology is a continual and iterative task that requires adjustment on the basis of experience. At the same time, it is important to understand the conditions at NASA that prevented an accurate analysis of the data and the risk and the types of safety culture flaws that contribute to unrealistic risk assessment.

Why would intelligent, highly educated, and highly motivated engineers engage in such poor decision-making processes and act in a way that seems irrational in retrospect? One view of culture provides an explanation. Social anthropologists conceive of culture as an ongoing, proactive process of reality construction.[30] In this conception of culture, organizations are socially constructed realities that rest as much in the heads of members as in sets of rules and regulations. Organizations are sustained by belief systems that emphasize the importance of rationality. Morgan calls this the *myth of rationality* and it helps in understanding why, as in both the Challenger and Columbia accidents, leaders often appear to ignore what seems obvious in retrospect. The myth of rationality "helps us to see certain patterns of action as legitimate, credible, and normal, and hence to avoid the wrangling and debate that would arise if we were to recognize the basic uncertainty and ambiguity underlying many of our values and actions."[31]

In both the Challenger and Columbia accidents, the decision makers saw their actions as rational at the time although hindsight suggests otherwise. Understanding and preventing poor decision making under conditions of uncertainty requires providing environments and tools that help to stretch our belief systems and overcome the constraints of our current mental models, i.e., to see patterns that we do not necessarily *want* to see.[32]

A final common aspect of the dangerous complacency and overconfidence seen in the manned space program is related to the use of redundancy to increase reliability. One of the rationales used in deciding to go ahead with the disastrous Challenger flight despite engineering warnings was that there was a substantial safety margin (a factor of three) in the O-rings over the previous worst case of Shuttle O-ring erosion. Moreover, even if the primary O-ring did not seal, it was assumed that a second, redundant one would. During the accident, the failure of the primary O-ring caused conditions that led to the failure of the backup O-ring. In fact, the design changes necessary to incorporate a second O-ring contributed to the loss of the primary O-ring.

The design of the Shuttle solid rocket booster (SRB) was based on the U.S. Air Force Titan III, one of the most reliable ever produced. Significant design changes were made in an attempt to increase that reliability further, including changes in the placement of the O-rings. A second O-ring was added to the Shuttle solid rocket motor design to provide backup: If the primary O-ring did not seal, then the secondary one was supposed to pressurize and seal the joint. In order to accommodate the two O-rings, part of the Shuttle joint was designed to be longer than in the Titan. The longer length made the joint more susceptible to bending under combustion pressure, which led to the failure of the primary and backup O-rings. In this case, and in a large number of other cases,[33] the use of redundancy requires design choices that in fact defeat the redundancy at the same time that the redundancy is creating unjustified confidence and complacency.

In this case, the ineffectiveness of the added O-ring was actually known. An engineer at NASA Marshall Space Flight Center concluded after tests in 1977 and 1978 that the second O-ring was ineffective as a backup seal. Nevertheless, in November 1980, the SRB joint design was classified as redundant until November 1982. Its classification was changed to non-redundant on December 17, 1982 after tests showed the secondary O-ring was no longer functional after the joints rotated under 40 percent of the SRB maximum operating pressure. Why that information did not get to those making the

[30] Gareth Morgan, *Images of Organizations,* Sage Publications, 1986.
[31] Morgan, ibid, pp. 134-135.
[32] Leveson, 2004, ibid.
[33] Leveson, ibid, 1995.

Challenger launch decision is unclear, but communication and information system flaws may have contributed (see the section on safety engineering practices below).

Organizational Structure

Organizational change experts have long argued that structure drives behavior. Much of the dysfunctional behavior related to both accidents can be traced to flaws in the NASA organizational safety structure including poorly designed independence, ill-defined responsibility and authority, a lack of influence and prestige leading to insufficient impact, and poor communication and oversight.

Independence, Responsibility, and Authority

Both accident reports criticized the lack of independence of the safety organization. After the Challenger loss, a new independent safety office was established at NASA Headquarters, as recommended in the Rogers' Commission report. This group is supposed to provide broad oversight, but its authority is limited and reporting relationships from the NASA Centers are vague. In essence, the new group was never given the authority necessary to implement their responsibilities effectively and nobody seems to have been assigned accountability. The CAIB report noted in 2003 that the management of safety at NASA involved "confused lines of responsibility, authority, and accountability in a manner that almost defies explanation."[34]

The CAIB also noted that "NASA does not have a truly independent safety function with the authority to halt the progress of a critical mission element."[35] In essence, the project manager "purchased" safety from the quality assurance organization. The amount of system safety applied was limited to what and how much the project manager wanted and could afford. "The Program now decides on its own how much safety and engineering oversight it needs."[36]

The problems are exacerbated by the fact that the Project Manager also has authority over the safety standards applied on the project. NASA safety "standards" are not mandatory. In essence, they function more like guidelines than standards. Each program decides what standards are applied and can tailor them in any way they want.

There are safety review panels and procedures *within* individual NASA programs, including the Shuttle program. Under various types of pressures, including budget and schedule constraints, however, the independent safety reviews and communication channels within the Shuttle program degraded over time and were taken over by the Shuttle Program office.

Independence of engineering decision making also decreased over time. While in the Apollo and early Shuttle programs the engineering organization had a great deal of independence from the program manager, it gradually lost its authority to the project managers, who again were driven by schedule and budget concerns.

In the Shuttle program, all aspects of system safety are in the mission assurance organization. This means that the same group doing the system safety engineering is also doing the system safety assurance—effectively eliminating an independent assurance activity.

In addition, putting the system safety engineering (e.g., hazard analysis) within the assurance group has established the expectation that system safety is an after-the-fact or auditing activity only. In fact, the most important aspects of system safety involve core engineering activities such as building safety into the basic design and proactively eliminating or mitigating hazards. By treating safety as an assurance activity only, safety concerns are guaranteed to come too late in the process to have an impact on the critical design decisions. Necessary information may not be available to the engineers when they are making decisions and instead potential safety problems are raised at reviews, when doing something about the poor decisions is costly and likely to be resisted.

---

[34] Gehman, ibid, p. 186
[35] Gehman, ibid, p. 180
[36] Gehman, ibid, p. 181

This problem results from a basic dilemma: either the system safety engineers work closely with the design engineers and lose their independence or the safety efforts remain an independent assurance effort but safety becomes divorced from the engineering and design efforts. The solution to this dilemma, which other groups use, is to separate the safety engineering and the safety assurance efforts, placing safety engineering within the engineering organization and the safety assurance function within the assurance groups. NASA attempted to accomplish this after Columbia by creating an Independent Technical Authority within engineering that is responsible for bringing a disciplined, systematic approach to identifying, analyzing, and controlling hazards. The design of this independent authority is already undergoing changes with the result unclear as this time.

Influence and Prestige

The Rogers Commission report on the Challenger accident observed that the safety program had become "silent" and undervalued. A chapter in the report, titled *The Silent Safety Program*, concludes that a properly staffed, supported, and robust safety organization might well have avoided the communication and organizational problems that influenced the infamous Challenger launch decision.

After the Challenger accident, as noted above, system safety was placed at NASA Headquarters in a separate organization that included mission assurance and other quality assurance programs. For a short period thereafter this safety group had some influence, but it quickly reverted to a position of even less influence and prestige than before the Challenger loss. Placing system safety in the quality assurance organization, often one of the lower prestige groups in the engineering pecking order, separated it from mainstream engineering and limited its influence on engineering decisions. System safety engineering, for all practical purposes, began to disappear or became irrelevant to the engineering and operations organizations. Note that the problem here is different from that before Challenger where system safety became silent because it was considered to be less important in an operational program. After Challenger, the attempt to solve the problem of the lack of independence of system safety oversight quickly led to loss of its credibility and influence and was ineffective in providing lasting independence (as noted above).

In the testimony to the Rogers Commission, NASA safety staff, curiously, are never mentioned. No one thought to invite a safety representative to the hearings or to the infamous teleconference between Marshall and Thiokol. No representative of safety was on the mission management team that made key decisions during the countdown to the Challenger flight.

The Columbia accident report concludes that, once again, system safety engineers were not involved in the important safety-related decisions although they were ostensibly added to the mission management team after the Challenger loss. The isolation of system safety from the mainstream design engineers added to the problem:

> "Structure and process places Shuttle safety programs in the unenviable position of having to choose between rubber-stamping engineering analyses, technical errors, and Shuttle program decisions, or trying to carry the day during a committee meeting in which the other side always has more information and analytical ability."[37]

The CAIB report notes that "We expected to find the [Safety and Mission Assurance] organizational deeply engaged at every level of Shuttle management, but that was not the case."

One of the reasons for the lack of influence of the system safety engineers was the stigma associated with the group, partly resulting from the placement of an engineering activity in the quality assurance organization.

Safety was originally identified as a separate responsibility by the Air Force during the ballistic missile programs of the 1950's to solve exactly the problems seen here—to make sure that safety is given due consideration in decision making involving conflicting pressures and that safety issues are visible at all levels of decision making. Having an effective safety program cannot prevent errors in judgment in balancing conflicting requirements of safety and schedule or cost, but it can at least make sure that decisions are informed and that safety is given due consideration. However, to be effective the system

[37] Gehman, ibid. p. 187

safety engineers must have the prestige necessary to have the influence on decision making that safety requires. The CAIB report addresses this issue when it says that:

> "Organizations that successfully operate high-risk technologies have a major characteristic in common: they place a premium on safety and reliability by structuring their programs so that technical and safety engineering organizations own the process of determining, maintaining, and waiving technical requirements with a voice that is equal to yet independent of Program Managers, who are governed by cost, schedule, and mission-accomplishment goals."[38]

Both accident reports note that system safety engineers were often stigmatized, ignored, and sometimes actively ostracized. "Safety and mission assurance personnel have been eliminated [and] careers in safety have lost organizational prestige."[39] The author has received personal communications from NASA engineers who write that they would like to work in system safety but will not because of the negative stigma that surrounds most of the safety and mission assurance personnel. Losing prestige has created a vicious circle of lowered prestige leading to stigma, which limits influence and leads to further lowered prestige and influence and lowered quality due to the most qualified engineers not wanting to be part of the group. Both accident reports comment on the quality of the system safety engineers and the SIAT report in 2000 also sounded a warning about the quality of NASA's Safety and Mission Assurance efforts.[40]

Communication and Oversight

Proper and safe engineering decision-making depends not only on a lack of complacency—the desire and willingness to examine problems—but also on the communication and information structure that provides the information required. For a complex and technically challenging system like the Shuttle with multiple NASA Centers and contractors all making decisions influencing safety, some person or group is required to integrate the information and make sure it is available for all decision makers.

Both the Rogers Commission and the CAIB found serious deficiencies in communication and oversight. The Rogers Commission report noted miscommunication of technical uncertainties and failure to use information from past near-misses. Relevant concerns were not being reported to management. For example, the top levels of NASA management responsible for the launch of Challenger never heard about the concerns raised by the Morton Thiokol engineers on the eve of the launch nor did they know about the degree of concern raised by the erosion of the O-rings in prior flights. The Rogers Commission noted that memoranda and analyses raising concerns about performance and safety issues were subject to many delays in transmittal up the organizational chain and could be edited or stopped from further transmittal by some individual or group along the chain.[41]

A report written before the Columbia accident notes a "general failure to communicate requirements and changes across organizations" (ref. 9). The CAIB found that "organizational barriers … prevented effective communication of critical safety information and stifled professional differences of opinion." It was "difficult for minority and dissenting opinions to percolate up through the agency's hierarchy".[42]

As contracting of Shuttle engineering has increased over time, safety oversight by NASA civil servants has diminished and basic system safety activities have been delegated to contractors. According to the CAIB report, the operating assumption that NASA could turn over increased responsibility for Shuttle safety and reduce its direct involvement was based on the 1995 Kraft report that concluded the Shuttle was a mature and reliable system and that therefore NASA could change to a new mode of management with less NASA oversight. A single NASA contractor was given responsibility for Shuttle safety (as well

---

[38] Gehman, ibid, p.184
[39] Gehman, ibid, p. 181.
[40] McDonald, ibid.
[41] Edwin Zebroski, Sources of common cause failures in decision making involved in man-made catastrophes. In James Bonin and Donald Stevenson (eds.), *Risk Assessment in Setting National Priorities*, p. 443-454, Plenum Press, New York, 1989.
[42] Gehman, ibid, p 183.

as reliability and quality assurance), while NASA was to maintain "insight" into safety and quality assurance through reviews and metrics. In fact, increased reliance on contracting necessitates *more* effective communication and more extensive safety oversight processes, not less.

Many aerospace accidents have occurred after the organization transitioned from oversight to "insight".[43] The contractors have a conflict of interest with respect to safety and their own goals and cannot be assigned the responsibility that is properly that of the contracting Agency. In addition, years of workforce reductions and outsourcing had "culled from NASA's workforce the layers of experience and hands-on systems knowledge that once provided a capacity for safety oversight".[44]

System Safety Engineering Practices

After the Apollo fire in 1967 in which three astronauts were killed, Jerome Lederer (a renowned aircraft safety expert) created what was considered at the time to be a world-class system safety program at NASA. Over time, that program declined for a variety of reasons, many of which were described earlier. After the Challenger loss, there was an attempt to strengthen it, but that attempt did not last long due to failure to change the conditions that were causing the drift to ineffectiveness. The CAIB report describes system safety engineering at NASA at the time of the Columbia accident as "the vestiges of a once robust safety program."[45] The changes that occurred over the years include:

- *Reliability engineering was substituted for system safety*. Safety is a system property and needs to be handled from a system perspective. NASA, in the recent past, however, has treated safety primarily at the component level, with a focus on component reliability. For example, the CAIB report notes that there was no one office or person responsible for developing an integrated risk analysis above the subsystem level that would provide a comprehensive picture of total program hazards and risks. Failure Modes and Effects Analysis (FMEA), a bottom-up reliability engineering technique, became the primary analysis method. Hazard analyses were performed but rarely used. NASA delegated safety oversight to its operations contractor USA, and USA delegated hazard analysis to Boeing, but as of 2001, "the Shuttle program no longer required Boeing to conduct integrated hazard analyses."[46] Instead, Boeing performed analysis only on the failure of individual components and elements and was not required to consider the Shuttle as a whole, i.e., system hazard analysis was not being performed. The CAIB report notes "Since the FMEA/CIL process is designed for bottom-up analysis at the component level, it cannot effectively support the kind of `top-down' hazard analysis that is needed … to identify potentially harmful interactions between systems"[47] (like foam from the external tank hitting the forward edge of the orbiter wing).

- *Standards were watered down and not mandatory* (as noted earlier).

- *The safety information system was ineffective.* Good decision-making about risk is dependent on having appropriate information. Without it, decisions are often made on the basis of past success and unrealistic risk assessment, as was the case for the Shuttle. Lots of data was collected and stored in multiple databases, but there was no convenient way to integrate and use the data for management, engineering, or safety decisions.[48]

---

[43] Nancy G. Leveson, The Role of Software in Spacecraft Accidents, *AIAA Journal of Spacecraft and Rockets*, Vol. 41, No. 1, July 2004.
[44] Gehman, ibid, p. 181
[45] Gehman, ibid, p. 177
[46] Gehman, ibid, p. 188
[47] Gehman, ibid, p. 188
[48] Aerospace Safety Advisory Panel, The Use of Leading Indicators and Safety Information Systems at NASA, NASA Headquarters, March 2003.

Creating and sustaining a successful safety information system requires a culture that values the sharing of knowledge learned from experience. Several reports have found that such a learning culture is not widespread at NASA and that the information systems are inadequate to meet the requirements for effective risk management and decision-making.[49,50,51,52,53] Sharing information across Centers is sometimes problematic and getting information from the various types of lessons-learned databases situated at different NASA centers and facilities ranges from difficult to impossible. Necessary data is not collected and what is collected is often filtered and inaccurate or tucked away in multiple databases without a convenient way to integrate the information to assist in management, engineering, and safety decisions; methods are lacking for the analysis and summarization of causal data; and information is not provided to decision makers in a way that is meaningful and useful to then. In lieu of such a comprehensive information system, past success and unrealistic risk assessment are being used as the basis for decision-making.

- *Inadequate safety analysis was performed when there were deviations from expected performance.* The Shuttle standard for hazard analyses (NSTS 22254, *Methodology for Conduct of Space Shuttle Program Hazard Analyses)* specifies that hazards be revisited only when there is a new design or the design is changed: There is no process for updating the hazard analyses when anomalies occur or even for determining whether an anomaly is related to a known hazard.

- *Hazard analysis, when it was performed, was not always adequate.* The CAIB report notes that a "large number of hazards reports contained subjective and qualitative judgments, such as `believed' and `based on experience from previous flights' this hazard is an accepted risk." The hazard report on debris shedding (the proximate event that led to the loss of the Columbia) was closed as an accepted risk and was not updated as a result of the continuing occurrences.[54] The process laid out in the Shuttle standards allows hazards to be closed when a mitigation is *planned*, not when the mitigation is actually implemented.

- *There was evidence of "cosmetic system safety."* Cosmetic system safety is characterized by superficial safety efforts and perfunctory bookkeeping: hazard logs may be meticulously kept, with each item supporting and justifying the decisions made by project managers and engineers.[55] The CAIB report notes that "Over time, slowly and unintentionally, independent checks and balances intended to increase safety have been eroded in favor of detailed processes that produce massive amounts of data and unwarranted consensus, but little effective communication".[56]

Conclusions

Space exploration is inherently risky. There are just too many unknowns and requirements to push the technological envelope to be able to reduce the risk level to that of other aerospace endeavors such as

[49] Aerospace Safety Advisory Panel, Annual Report, NASA, January 2003.
[50] Aerospace Safety Advisory Panel, The Use of Leading Indicators and Safety Information Systems at NASA, NASA, March 2003.
[51] Government Accounting Office, Survey of NASA's Lessons Learned Process, GAO-01-1015R, September 5, 2001.
[52] MacDonald, ibid.
[53] Gehman, ibid.
[54] Gehman, ibid.
[55] Leveson, ibid, 1995.
[56] Gehman, ibid, p. 180

commercial aircraft. At the same time, the known and preventable risks can and should be managed effectively.

The investigation of accidents creates a window into an organization and the opportunity to examine and fix unsafe elements. The repetition of the same factors in the Columbia accident implies that NASA was unsuccessful in permanently eliminating those factors after the Challenger loss. The same external pressures and inadequate responses to them, flaws in the safety culture, dysfunctional organizational safety structure, and inadequate safety engineering practices will continue to contribute to the migration of the NASA manned space program to states of continually increasing risk until changes are made and safeguards are put into place to prevent that drift in the future. The current NASA administrator, Michael Griffin, and others at NASA are sincerely trying to make the necessary changes that will ensure the safety of the remaining Shuttle flights and the success of the new manned space program missions to the Moon and to Mars. It remains to be seen whether these efforts will be successful.

The lessons learned from the Shuttle losses are applicable to the design and operation of complex systems in many industries. Learning these lessons and altering the dynamics of organizations that create the drift toward states of increasing risk will be required before we can eliminate unnecessary accidents.

Acknowledgements: Many of the ideas in this essay were formulated during discussions of the "Columbia Group," an informal, multidisciplinary group of faculty and students at MIT that started meeting after the release of the CAIB report to discuss that report and to understand the accident. The group continues to meet regularly but with broader goals involving accidents in general. The author would particularly like to acknowledge the contributions of Joel Cutcher-Gershenfeld, John Carroll, Betty Barrett, David Mindell, Nicolas Dulac, and Alexander Brown.