Risk Awareness:
A New Framework for Risk Management in Flight Test

Col Douglas "Beaker" Wickert, PhD (AF)
Air Force Test and Evaluation
Headquarters U.S. Air Force

"*The first principle is that you must not fool yourself—and you are the easiest person to fool.*"[1]

Richard Feynman

A robust, risk-aware culture is essential for flight test organizations to effectively manage risk in flight test. Despite apparently robust risk management, the flight test community continues to experience a high rate of accidents and mishaps. A review of decades of safety literature reveals several new risk management frameworks that have not been widely adopted in flight test. The flight test community needs an updated accident model and a practical risk management framework that recognizes the challenges of uncertainty and complexity inherent in flight test. ***Risk awareness****--the perception of uncertainty and the potential, projected outcomes resulting from uncertainty--is an attempt at such a framework and is based on four principles: 1) understanding the type of uncertainty; 2) reducing reducible ignorance; 3) democratizing safety; and 4) resisting drift. This framework regards accidents as a phase transition with knowledge as the primary control parameter. The risk awareness framework includes a distinction between risk and uncertainty and argues that different cognitive and risk management tools are applicable to these different domains. The risk awareness framework attempts to explain the primary contribution to mishaps in complex systems and offers practical assessment tools for the doers, reviews, and approvers in the flight test process to make risk-informed decisions.*

1 Introduction

Flight test is dangerous. By its very nature, flight test probes the unknown, revealing previously unexpected cliffs, model flaws, system deficiencies, flawed human factor designs, and inaccurate design assumptions. Not infrequently, the cost for obtaining the data that reduces uncertainty is realized in catastrophe. Over multiple decades and through the accumulated wisdom of mishaps in flight test, the flight test community has developed a relatively robust safety and risk management process: Test Hazard Analysis (THAs), General Minimizing Procedures/Conditions (GMPs/GMCs), Safety Review Boards (SRBs), First Flight Readiness Reviews (FFRRs), Test Readiness Reviews (TRRs), Threat Hazard Analysis Worksheets (THAWs), Safety Significant Event Reports (SSERs), Crew Resource Management (CRM), Operational Risk Management (ORM), and formal Risk Management Plans (RMPs). The Society for Experimental Test Pilots (SETP), the Society of Flight Test Engineers (SFTE), and individuals and organizations from across the military and civilian flight test enterprise routinely discuss lessons learned, highlight best practices, and collectively study accidents to avoid recurrences. To be a professional flight tester is to be a professional risk manager. Lessons learned and new ideas for risk management are routinely shared among the community.[2-12]

The Flight Test Safety Committee (FTSC) devoted the entire 2018 Flight Test Safety Workshop to risk management fundamentals.[13]

Although the flight test community takes test safety very seriously, there remains significant room to improve risk management in flight test. Since 2011, we have experienced at least 15 total-loss mishaps in flight test, including 24 fatalities in ten separate incidents.[14] Two of the test-related Class As, including one fatality, occurred under the author's operational command. In the aftermath of those accidents, the author undertook a comprehensive review of existing safety literature and accident investigations seeking better ideas for risk management.[15] Many of the ideas in the literature review were underwhelming or overly academic with little practical utility for making concrete decisions about risk, however there were several useful perspectives. A complex system perspective and the importance of organizational culture is widely recognized and emphasized.[16-19] Unfortunately, many of the existing accident models are merely descriptive. Operational leaders and risk managers need prescriptive, practical methods for fostering robust and safe operations as well as concrete steps to guide decision making under uncertainty. The risk awareness framework offered in this paper is a perspective that embraces complexity and pulls together the most relevant aspects of various safety models.[20] *Risk awareness* is not a final, definitive solution to risk management. Due to uncertainty, risk will always be an inherent aspect of flight test. However, by recognizing distinctions in different types of uncertainty, better manage risk is possible.

1.1 Literature Review of Accident Models and Risk Management Frameworks

Over the years, academics and safety professionals have introduced various models or frameworks for explaining how and why accidents occur. A common approach to risk management in many fields, including flight test, has been to start by examining accidents or undesired, negative outcomes; identify the factors that led to the mishap; and then design "safety controls" to either prevent the occurrence or reduce the severity should it occur again.[21,22] In theory, the approach is not without merit. Just as medical doctors spend a significant part of their training understanding the nature of disease, risk managers should understand the underlying sources and reasons for mishaps.

Usually considered the first organizational safety model, Heinrich's 1931 *Industrial accident prevention: a scientific approach* introduced the "domino theory" of accidents which views accidents as the culmination of a linear chain of events, often the result of a worker's carelessness or error.[23] Under the domino framework, accidents are prevented by a focus on training and procedural compliance. The single-sequence, chain-of-event description of mishaps is now generally regarded as too simplistic, though echoes of it exist in the well-known swiss cheese model.[24] In the swiss cheese model, accidents result from the unfortunate "lining up" of the holes in the various layers of defense in safety controls. Just as removing one of the dominoes is sufficient to "break the mishap chain" and prevent an accident, adding another layer in the defense, assuming that it is not hole-ridden, should prevent an accident by interrupting the sequence of events. The shortcoming of these models is that they do not capture the unpredictable interactions and other important features of

complex systems. As we will see, in the context of complexity, adding additional layers of defenses, additional complexity, can actually make mishaps more likely. This feature of many safety strategies is well-recognized by researchers: "Well-intentioned, commonplace solutions to safety problems often fail to help, have unintended side effects, or exacerbate problems."[25]

The first attempt to represent the nature of complex interactions in accidents was Turner's 1978 *Man-Made Disasters*.[26] In Turner's framework, disasters are the consequential release of energy due to imperfect or unappreciated information. The greater the energy in the system, the greater potential consequences of the disaster. An important aspect of Turner's framework was an incubation period during which an "accumulation of an unnoticed set of events which are at odds with the accepted beliefs about hazards" goes unrecognized.[27] This is a key point which we will leverage in developing our hybrid framework for risk management. Shortly after Turner, Perrow introduced the concept of the *system accident* which is an early recognition of the importance of complexity in explaining accident origins.[28]

Rather than a sequential chain of events as in the domino theory, systems theory posits that accidents result from the *interaction* of human decisions, machines, and the environment.[29] According to system theory, accidents are a failure of safety controls or constraints at a system-interaction level. Perrow further suggests that accidents in complex, tightly-coupled systems are inevitable because they are essentially uncontrollable by humans. Leveson's Systems-Theoretic Accident Model and Processes (STAMP) is firmly in the systems framework. STAMP explains accidents as the result of "external disturbances, component failures, and/or dysfunctional interactions among system components [that] are not adequately controlled, i.e., accidents result from inadequate control or enforcement of safety-related constraints on the development, design, and operation of the system."[30] According to Leveson, "preventing future accidents requires designing a control structure that will enforce the necessary constraints."[31] Per the system accident model, understanding why an accident occurred requires determining why the control structure was ineffective. System perspectives are necessary for design and reliability engineering and are a clear improvement to the single cause, chain-of-event accident model, but the notion of *control* is problematic, particularly in the context of uncertainty: it is difficult to control that which is not known.

A different approach to accident prevention is to look at organizations that are reliably accident resistant for clues to desirable organizational traits and practices for encouraging or even designing a "safety culture." For example, Pidgeon and O'Leary highlighted attributes of safety culture such as senior management commitment, shared care and realistic norms for hazards, and continual reflection of practices, i.e., organizational learning.[32] Weick emphasizes organizational learning as the key component of "high-reliability organizations."[33,34] There is a somewhat spirited debate among safety researchers over the usefulness of high-reliability organization and safety culture frameworks to preventing mishaps.[35,36] The details are academic and beyond the scope of this paper,

but the recognition that organizational culture is foundational to safety is important and a key component of the risk awareness framework.

Common to all of the accident models outlined thus far is a general idea that safety measures will inevitably erode over time:

> *Defenses therefore tend to degenerate systematically over time. When a larger view is taken, most accidents in complex systems can be seen to result from a migration to states of increasing risk over time. Once a system has migrated to an unsafe state, accidents are inevitable unless appropriate efforts are made to bring the system to a safe state.*[37]

The erosion of safety margin is an organizational behavior identified as "drift" by Jens Rasmussen.[38] Preventing drift will be an important component of our *risk awareness* framework for flight test.

More recently, Dekker has combined organizational complexity and the drift model to explain and argue that robust safety culture is an emergent phenomena.[39] Dekker's argument is fairly complicated, but it reflects a contemporary perspective that has become dominant in much of the safety literature: accidents, or by converse, safety, are emergent properties in the context of complexity theory. Hollnagel offers a similar argument and a prescription for moving to pro-active safety management which puts the organizational focus on doing things right rather than preventing things from going wrong. Hollnagel describes the goal of traditional safety models, which he calls "Safety I," as minimizing the number of things that can go wrong. By contrast, he argues, "Safety II" should seek to foster adjustments to operations before negative events occur rather than react to events to prevent them from happening.[40] "From a Safety-II perspective, the purpose of safety management is to ensure that as much as possible goes right."[41] Accidents, according to Hollnagel, are the consequence of unexpected combinations ("functional resonance") of normal performance variability as workers make efficiency-thoroughness tradeoffs.[42] The functional resonance accident model (FRAM) seeks to prevent accidents by identifying potential sources of resonance and designing countermeasures. There is much sound theory in both Dekker's and Hollnagel's descriptions of accidents. The challenge with both is that they are unnecessarily complicated and unlikely to be adopted in their existing form for flight test.

1.2 The practical challenge of existing frameworks

The principal problem with the conceptual frameworks for risk management and accident prevention introduced above is that they are too academic, too complicated, and too theoretical to be easily and practically applied to flight test. This may partly explain why both military and civilian flight test has largely failed to adopt any of the newer risk management frameworks. Test hazard analysis (THAs), the core risk management tool that informs SRBs, FFRRs, etc., has remained unchanged for decades despite the increasing complexity of the systems undergoing test. Properly used, THAs are powerful and robust; a practical example of what Gary Klein refers to as "pre-mortems".[43] But THAs are typically based on a linear fault tree or sequence and often assume single failures only. As systems under test become more complex and complicated, unpredictable "system failures" become more likely. System engineering and system safety design tools such as FMEA and

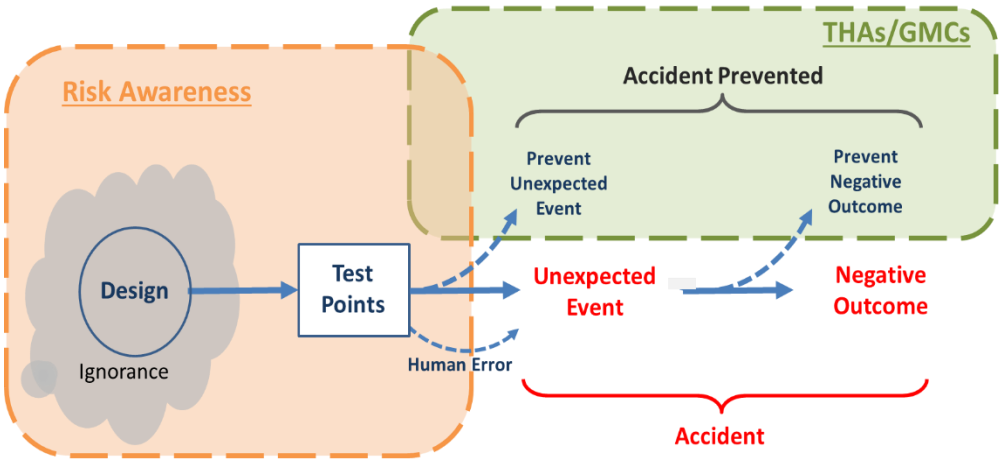


Figure 1: Accidents are a sudden, unexpected event that results in an unwanted, negative outcome. Preventing the event or mitigating the outcome, the domains of THAs/GMCs, will prevent/mitigate the accident. Risk awareness is focused on managing uncertainty and the potential projected outcomes of uncertainty.

FMECA which seek to eliminate or mitigate effects of component failures are valuable but ultimately insufficient for complex systems. As systems and systems under test become more complex and complicated, unpredictable "system failures" become more likely.

Fortunately, flight test is not resistant to change when the need is clear. New flight test planning methods and flight test techniques are routinely introduced to address or reduce exposure to specific hazards.[44] For example, following Dave "Cools" Cooley's 2009 fatal F-22 mishap, the test community was quick to develop and adopt time-safety margin (TSM) as a best practice.[45,46] However, despite a willingness to accommodate new test planning, flight test is still experiencing an unsettlingly high accident rate. Perhaps a fundamental shift in frameworks is required.

It is not that the existing practices in flight test risk management are wrong as incomplete. THAs, GMCs, TSM and other tools of experience, often learned through catastrophic loss, must remain part of the risk management toolbox (*Figure 1*). These are appropriate tools for managing the risk of foreseeable negative outcomes, a domain of future uncertainty that will be classified as the risk domain below. But flight test also needs a more comprehensive toolset to guide risk acceptance and go/no-go decisions. Risk management frameworks built around the notion of control are misleading (c.f. the swiss cheese model). Knowing how the holes lined up after an accident is of limited help. To inform go/no-go and other decisions in flight test, we need to know where the holes are before the test.

The international standard definition of risk is "the effect of uncertainty on objectives."[47] Likewise, ISO's definition for risk management is "coordinated activities to direct and control an organization with regard to risk."[48] The Air Force defines risk management as "a decision-making process to systematically evaluate possible courses of action, identify risks and benefits, and determine the best course of action (COA) for any given situation" and further explains that "the intent is to review all aspects of the test to ensure all identified hazards are controlled to an

acceptable level."[49,50] Both of these definitions are variously flawed and we adopt a different definition in section 2.1 below. For a risk management framework, the notion of *control* is particularly problematic. The widely-used swiss cheese framework describing accidents perpetuates the fallacy of control. Operations are either safe or they are not.[51] Safe and unsafe are two separate states just as liquid water is distinct from solid ice. Risk is not a control parameter. There is no closed loop mechanism to feedback risk and control operations. At best, controlling risk for particular programs or test points is an open loop, feed-forward system, but that model is equally flawed. According to the swiss cheese framework, adding additional safety barriers or layers should be sufficient to prevent accidents by reducing the likelihood that the "holes line up." In practice, adding additional layers is usually futile. It increases the complexity of the system and may actually obscure latent risks making risk management more challenging. Adding safety layers can even make a system less safe by introducing additional failure modes. For example, a flight termination system may fail causing inadvertent destruction of a system under test.

Flight test needs a risk framework that provides tools for managing uncertainty and ignorance. Test is the gradual process of revealing uncertainty.[52] At its core, risk management means making informed decisions in the face of uncertainty. Schedule and resource pressures will likely drive local-optimizations that erode safety margins, so an effective risk management framework must provide a practical means for assessing and countering drift towards unsafe operations. The proposed risk management framework rests upon three ideas:

1. *There exist different types of uncertainty; different cognitive tools are necessary depending on the nature of the unknown*
2. *Mishaps may be regarded as a phase change from safe to unsafe operations with knowledge as the critical control parameter*
3. *A phase change framework permits a reliable mechanism to assess and resist organizational drift into failure. Cultural attributes of organizations resistant to drift are measurable and may be cultivated*

2 Uncertainty

As in much of life, uncertainty is a pervasive and persistent fact in flight test. What we know is constrained by two primary factors: 1) the theoretical limits of knowledge; and 2) the inherent variability or randomness of some systems. First, there exist unknowns which are fundamentally unknowable beyond certain boundaries or limits.[53] Chaotic systems and the Heisenberg uncertainty principle are ready examples from physics.[54] In complexity theory, emergent properties are not deterministic and hence cannot be predicted (more on this in section 3). Outcomes before an event—e.g., the flying qualities of a new aircraft before its first flight—cannot be known with certainty but may be predictable in a general sense with uncertain confidence intervals. By contrast, there are uncertain outcomes governed primarily by randomness. Predicting the roll of a fair die is not possible, but predicting the result within precise, knowable probabilities is straightforward. It is common in the field of decision making to distinguish between aleatory (random) and epistemic

(knowledge) uncertainty.[55] This distinction has a long history in decision science literature and is widely used in fields of decision support (e.g., policy analysis, economic policy, business strategy development, environmental impact assessment, etc.).[56-58] The distinction is important because different analytical tools and decision criteria are appropriate if the uncertainty is due to lack knowledge or if the uncertainty is due to variability.[59] This is discussed in more detail below.

2.1 Domains of Knowledge

Taking knowledge and variability as independent parameters naturally creates the four quadrants of knowledge given in *Figure 2*. The *certainty* domain in the lower-left quadrant is characterized by a state of high knowledge and low variability. Systems and decisions within the certainty domain are fully deterministic. The performance and flying qualities of a fielded aircraft with decades of data and experience falls within the certainty domain. The upper-left quadrant represents a state of high knowledge and high variability and is commonly called the *risk* domain. Stochastic systems, medical tests (with established accuracy and known false-positive rates), casino games, lottery tickets, and the failure rates of mature, well-tested flight system components are all representative exemplars of this domain where the "unknowns" are well-known and well-characterized. Risk is a slightly unfortunate term since it is often used interchangeably to refer to hazards in lay conversation. In flight test contexts, risk is technically defined as the combined severity and probability of a hazard or event. Since risk in the traditional test application of the term requires an estimation of probability of occurrence, it is not an inaccurate term in that context. The distinction of *risk* as a domain from the "risk" in risk management should be clear from the context.

The right side of the knowledge axis in *Figure 2* are the domains of *uncertainty*. The lower-right quadrant, characterized by low knowledge and low variability, is the domain of *pure uncertainty*. The systems studied in chaos theory, e.g. the famous Lorenz system depicted in the quadrant, are deterministic (non-random) yet unpredictable due to limited knowledge of the initial conditions. The upper-right quadrant, *random uncertainty*, is uncertain due to both randomness and low knowledge. Markov chains are one example of random uncertainty. Quantum mechanics includes several examples of uncertainty with random components including the radioactive decay annoying Schrodinger's cat.[60] Taken together, these two domains of uncertainty represent ignorance, i.e., a lack of knowledge. Some of the more difficult and challenging questions in flight test occur in this domain and is thus a key feature of *risk awareness*.

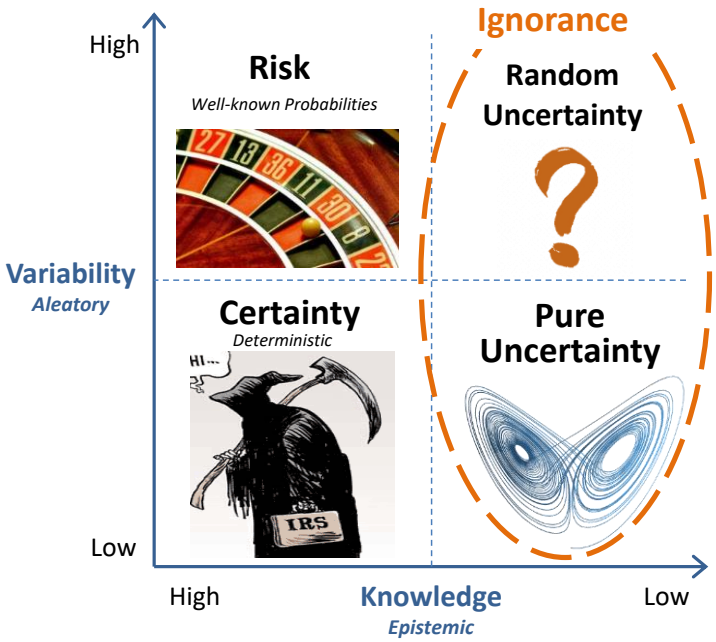


Figure 2: The domains of ignorance and uncertainty.

Many other disciplines—monetary policy, regulatory actions, epidemiology, hydrological management, environmental planning, program and project management, etc.—are also faced with the challenge of making decision with incomplete knowledge and under uncertainty.[61-69] A common approach in these fields is to describe a spectrum of ignorance that ranges from high knowledge to low knowledge: determinism, stochastic, ambiguous scenarios, recognized ignorance, and complete ignorance.[70] It is straightforward to map the spectrum of ignorance to the corresponding regions in the aleatory/epistemic quadrants of uncertainty (*Figure 3*).

2.2 Cognitive Biases and Heuristics

The distinction between the different types of uncertainty or domains of ignorance is important because different cognitive tools and decision making methods are appropriate in the different domains. Decades of extensive research have explored and catalogued the effects of cognitive biases and heuristics on decision making. Various consistent flaws in human reasoning are well-established and have led to the creation of new fields, e.g. behavioral economics. A rich body of work documents cognitive biases, flawed heuristics, and influence of biases in a wide-range of applications.[71-73] However, there is also an equally rich body of work describing the reliability and accuracy of heuristics in wide-ranging applications including medicine, finance, business decisions, and flight

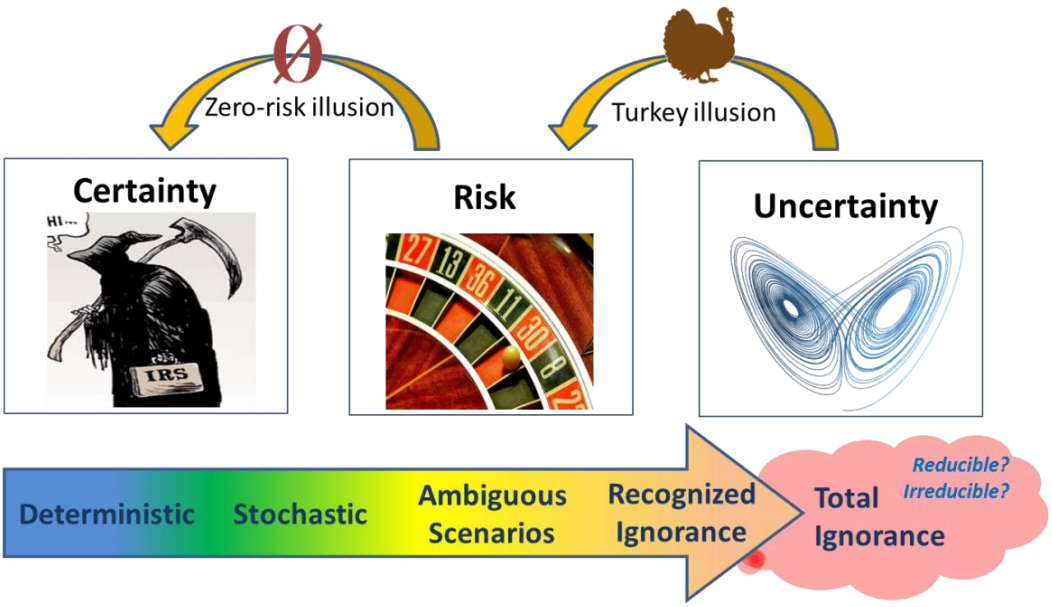


Figure 3: Spectrum of ignorance

test.[74-77] Kahneman and Gigerenzer, as the primary researchers and most prolific writers on these two perspectives, have become associated with the different camps. As with most things, the truth lies in the combination of ideas. Details of the debate between the cognitive-biases-are-bad and heuristics-are-good camps are beyond the scope of the current paper. At its heart, much of the supposed debate between Kahneman, the Nobel-laureate, and Gigerenzer, the spoiler, is not actually a debate. Both of them would agree that different cognitive tools are appropriate when applied in different domains.

Cognitive biases are an ever-present threat to decision makers. Psychology now counts well over 100 different cognitive biases.[78] Some of the more prominent that may be encountered in flight test are given in *Table 1*. Biases represent a blind spot in our rational reasoning. They are reliably consistent, cognitive flaws that can lead to poor decisions. In the author's experience, confirmation bias is the most common bias encountered in flight test. Due to confirmation bias, we tend to selectively emphasize or perceive data that supports (confirms) our working hypothesis and to discount data or information that disagrees with it. In a go/no-go decision, a go-oriented team will find ample justification to fly and is more likely to discount information that warns against flying. Confirmation bias was clearly a factor in the 1994 friendly fire shoot down of two UH-60 Black Hawks following the erroneous identification by two F-15s. The two F-15 pilots expected to see a hostile aircraft based on incorrect IFF information and their two fly-bys for visual identification simply confirmed their expectation. Confirmation bias is present in many failures of judgment. To counter confirmation bias, you must actively seek data and perspectives that negate your hypothesis. Actively soliciting dissenting opinions at Safety Review Boards is an effective technique. Empowering anyone in the control room with the ability to call "Knock It Off" is another technique to ensure all perspectives are considered.

Table 1: Predominant cognitive biases in flight test

Bias	Explanation	Example
Anchoring	Tendency to rely too heavily, to 'anchor', on the first estimate or piece of information (or on one aspect). In a group setting, anchoring can readily lead to groupthink.	Overreliance on first estimate of risk of a test or on the first prediction of performance of a system. A group may get 'anchored' on assessing the safety of test event undercutting the wisdom of multiple, independent perspectives
Availability	Tendency to overestimate the likelihood of events with greater 'availability' in memory; can be influenced by the recency of memories or by particularly notable events	The risk of inadvertent flight during high-speed taxi is probably generally overestimated by the example of the YF-16 "flight 0"
Confirmation	Tendency to more readily see or interpret evidence that support prior beliefs (confirm existing preconceptions). Tendency to discount evidence that contradicts expectation. Confirmation bias is likely the strongest, most susceptible, and most misleading cognitive bias.	"Go-oriented" teams or individuals will find data to support/reinforce a "go" decision. In the case of Columbia, after "Program managers learned about the foam strike, their belief that it would not be a problem was confirmed (early, and without analysis) by a trusted expert who was readily accessible and spoke from experience." (CAIB, p172)
Framing	Drawing different conclusions from the same information depending on how that information is presented. See prospect theory below.	Emphasizing the gain in a low probability but catastrophic event drives a tendency to seek risk (example given below).
Escalation	Justifying increased investment in a decision, because of cumulative prior investment, despite new evidence suggesting that the decision was probably wrong. Also known as the sunk cost fallacy.	A program's commitment to a particular propulsion design despite mounting evidence that the original design was fundamentally flawed. Delayed decision for redesign. A team's decision to continue testing even after primary objectives cannot be achieved because of the prior work in getting ready to test.
Base rate fallacy	Tendency to ignore base rate information (population likelihood) and focus only on specific information (likelihood pertaining only to a certain class). I.e., failure to calculate conditional probability (Bayes' Theorem)	When calculating the reliability of a redundant flight control computer, it is necessary to account for both the failure rate and the false failure rate to obtain accurate predictions of possible failed states in flight test
Valence effect	Optimism bias (over-estimate the probability of positive outcomes relative to negative outcomes); valence is intrinsic 'attractiveness,' (how we feel about something); research suggests that unrealistic optimism is greater for negative than positive valence	A first flight team's pride in their months of work preparing for first flight and the expected reward in achieving first flight can make the team too optimistic when considering likelihood of negative outcomes, contributing to not adopting all possible measures for safe execution.
Planning fallacy	The tendency to underestimate task-completion times	Can result in schedule pressure that cause drift and reduces safety margin
Hindsight bias	The inclination to see past events as being more predictable than they actually were (outcome bias: evaluating a decision when the outcome of that decision is already known)	Post-accident analysis reveals many accidents to be "inevitable" based on latent failures. Combined with valence effect leads to thinking that "we're smarter and it can't happen to us"
Validity illusion	Belief that our judgments are accurate, especially when available information is consistent or inter-correlated.	With the Challenger launch decision, managers reviewed O-ring erosion events as a function of temperature (*Figure 8*), but the data was sampled on the dependent variable resulting in an invalid perspective on likelihood that temperature was at the root of O-ring erosions (see below).
Turkey illusion	Treating the *pure uncertainty* domain as if it were governed by the *risk* domain; result is tendency to extrapolate the past to predict the future (inappropriate application of Bayesian approach)	Challenger treatment of O-ring erosion and blow-by as acceptable since there were two O-rings and complete blow-by of the second O-ring had never occurred. Repeated success or safe results of risky test maneuvers reduces concern or estimate of likelihood of hazardous consequences
Zero-risk illusion	Treating the *risk* domain as if it were *certain* (ignores the chance of false positives or Type II errors); prefer the total elimination of minor risks to the significant reduction of large ones	Safety systems (e.g. warning lights, flight termination systems), have false positive error rates. Two failed 'fail-safe' relief valves at Three Mile Island contributed to the disaster. Engineers in the aftermath of Gus Grissom's Liberty Bell 7 mishap largely discounted the possibility that the "hatch just blew" despite several possible failure mechanisms. The failure to replicate many scientific studies should be expected from the combination of Type I and Type II error prevalence.

Another bias that is particularly relevant in flight test is the framing effect. How a question is

answered or decision is made is strongly influenced by how it is framed. Prospect theory is a particular example of the framing effect.[79] Prospect theory readily explains why people buy insurance and lottery tickets and why countries prolong a devastating war long after it is clear they stand no chance of winning. For high-likelihood events representing an almost certain gain, there is a pronounced tendency for people to accept a somewhat smaller but certain gain. This risk aversion against a slight chance of "loosing" (gaining nothing) is not irrational. In the case of an almost certain, high-likelihood loss, the opposite tendency occurs. People tend take the chance, to "roll the die," in the slight possibility of avoiding a certain loss; i.e., they tend to be risk seeking. With respect to program management and flight test, prospect theory explains the sometimes bewildering difference in perspective between program offices and flight test organizations. In flight test, we are generally concerned with low probability but catastrophic loss (the lower row of *Table 2*). The author has repeatedly observed program offices take aggressive positions with the readiness to field a yet unproven system or to rapidly proceed to a test point that the flight test squadron is wary of. In one particularly vivid example, an experimental rotary wing aircraft developed excessive vibrations following a deployment; vibrations that were sufficiently severe that the pilots refused to fly the aircraft. The program office and manufacturer wanted to perform a test flight with strap-on accelerometers to characterize the reported vibrations. The flight test organization preferred a more gradual approach with finite element modal analysis and ground vibration testing before proceeding to flight test. The program office saw only the gain (a test demonstrating an airworthy system) and preferred the risk-seeking approach to the test. The flight test squadron was focused on the potential to lose the aircraft and were naturally more risk averse.

Table 2: Prospect theory matrix[80]

Certainty	Gain	Loss
High probability (certainty effect)	**Risk Averse** *Fear of disappointment*	**Risk Seeking** *Hope to avoid loss*
Low probability (possibility effect)	**Risk Seeking** *Hope of large gain*	**Risk Averse** *Fear of large loss*

Kahneman's System I, fast thinking, is the primary source of cognitive biases. The flaws and influence of cognitive bias can be overcome by slow and deliberate thinking, that is, System II. Heuristics are also the product of System I, but we want to draw a fundamental distinction between heuristics and biases. Biases are blind spots. Biases often result in decision errors or demonstrably poor judgment. But you can only be blind to that which it is possible to see. Stated differently, you cannot be blind to that which it is impossible to see. When making a decision in the domain of *pure uncertainty*, it is by definition impossible to see the unknown and hence it is impossible to be blind. Applying heuristics with humility is recommended in the domain of *pure uncertainty* which is where many of the decisions of flight test are encountered.

2.3 Heuristics for flight test

In the *risk domain*, calculated, reliable decisions are possible because of the high state of knowledge. It is in this domain that the tradeoff between Kahneman's two systems of thinking, i.e., fast (System I) and slow (System II) is most relevant. System II thinking is slow, purposeful, and requires concentration and effort. Applying System I (fast thinking) to decisions within the *risk domain* is inappropriate because better decisions are possible through calculation since the principal unknowns are due to known variation. System II (slow thinking) is the recommended cognitive approach to decision within the *risk domain*. THAs, a form of scenario planning for foreseeable, negative outcomes, are an example and appropriate use of slow thinking for *risk domain* decisions. Unfortunately, most of the "hard" decisions and surprises in flight test reside in the domain of *pure uncertainty*. With *pure uncertainty*, heuristics result in better decisions.

Heuristics—rules of thumb—are demonstrably more reliable in the domains of *pure uncertainty*. Tens of thousands of years of evolution have sculpted our minds into cognitive instruments capable of making decisions based on gut feelings. As Gigerenzer explains: "Heuristics are efficient cognitive processes that ignore information. In contrast to the widely held view that less processing reduces accuracy, the study of heuristics shows that less information, computation, and time can in fact improve accuracy."[81] Reliable heuristics in environments of incomplete information include the recognition heuristic, the take-the-best heuristic, the tallying heuristic, and the satisficing heuristic.[82] Whereas biases are sources of error in domains where better decisions result from deliberate, careful thinking, heuristics are useful in domains where decisions cannot be made due to uncertainty. Thus, in the domain of *pure uncertainty* heuristics generally produce better overall judgments. The reliability of heuristics for certain types of judgments has been repeatedly demonstrated and is at least partly explained by the bias-variance tradeoff in the field of data science and machine learning. A detailed explanation of the mathematics is beyond the scope of this paper, but reducing absolute error of a sample with high noise often leads to an over-fit model. Better predictive fits for sparse data sets result from simpler (high bias) models than over-fit (high variance) models. Simply put, with respect to unknowable unknowns, you are better off trusting your gut. Most test pilots and flight test professionals are likely to readily agree with this conclusion.[83]

Test pilots and flight test engineers develop and hone their heuristics through the cauldron of experience in flight test: *good judgment comes from experience, and experience comes from bad judgment*. Sharing lessons learned in flight test is a powerful and vital means for passing on collective wisdom within the test enterprise. It was primarily for this reason that the Society of Experimental Test Pilots was established. Every test pilot and flight test engineer should reflect and collect their own set of heuristics to be used in making judgments in flight test. For example:

1) *Keep it simple*: when deciding between multiple options for obtaining data, the simplest method is preferred

2) *Slower is faster*: methodical and deliberate build-up with schedule margin for review and understanding is usually faster in the long-run; reject short-cuts or rushed analysis that may need to be repeated or re-flown
3) *Seek contrary data*: solicit outside perspectives and fully consider alternate hypotheses before deciding on a course of action; necessary to reveal and avoid blind spots and counter confirmation bias during planning and execution
4) *Surprises are warning*: understand unexpected results before continuing

Before making a judgment or decision in flight test, characterize the nature of the uncertainty in the decision. For decisions in the *risk domain*, "slow" thinking (System II) will give reliably better decisions, but be wary of cognitive biases. In the domain of *pure uncertainty*, heuristics provide reliably better decisions. In short, flight testers should "trust their gut" and apply heuristics with humility.

3 Complexity and Drift Towards Mishaps

Complexity is likely to be one of the primary sources of uncertainty described above. A complex system is a system made up of a large number of nodes or components which interact with each other in unpredictable ways. Examples of complex systems include networks, ecosystems, social structures, political and business organizations, financial markets, hormone and biochemical interactions, the brain, the climate, military campaigns, cellular automata (e.g., Conway's "Game of Life"), and flight test programs. Complex systems are inherently difficult to analyze because of the number of connections, interactions, and dependencies between all the nodes in the system (*Figure 4*). A feature of complex systems is that, due to the sheer number of interactions and possible states (i.e. the complexity) of the system, system-wide outcomes are difficult or impossible to predict from the local actions or attributes of individual nodes.[84] Yet, stable patterns often emerge from complex systems. Complexity theory is a branch of mathematics that studies these patterns and other phenomena which emerge from collections of interacting objects.[85] Though often conflated with chaos theory, complexity theory is fundamentally different from chaos with a primary distinction being the phenomena of emergence.[86] Emergence is a system-wide property that is not predictable from an individual node. Dekker offers a compelling argument that describes accidents and other large-scale failures as an emergent property of complex organizations.[87] Nancy Leveson makes a similar argument:

> *We are designing systems with potential interactions among the components that cannot be thoroughly planned, understood, anticipated, or guarded against. The operation of some systems is so complex that it defies the understanding of all but a few experts, and sometimes even they have incomplete information about its potential behavior. Software is an important factor here: it has allowed us to implement more integrated, multi-loop control in systems containing large numbers of dynamically interacting components where tight coupling allows disruptions or dysfunctional interactions in one part of the system to have far-ranging rippling effects. The problem is that we are attempting to build systems that are beyond our ability to intellectually manage: Increased interactive complexity and coupling make it difficult for the designers to consider all the potential system states or for operators to handle all normal and abnormal situations and disturbances safely and effectively.*[88]

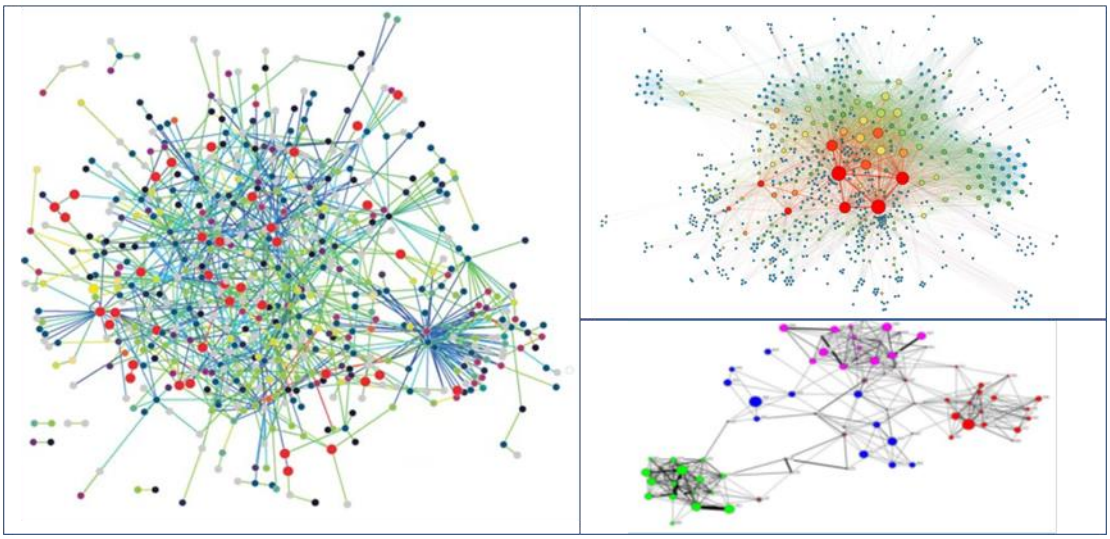


Figure 4: Examples of typical complex system networks (left: protein interaction network of Treponema pallidum[89]; top right: social network organization[90]; bottom right: Alex "Sandy" Pentland's MIT Humans Dynamics Laboratory mapping of face-to-face interaction within and among teams in a small organization[91])

3.1 Emergence and Phase Changes

Emergence is the appearance or description of system-wide phenomena or system-wide order that is unexpected or not predictable from the attributes of a single node or component of the system. It is the "arising of novel and coherent structures, patterns, and properties during the process of self-organization in complex systems."[92] For example, there is nothing in physics that predicts the macro properties of water from the well-understood properties of individual water molecules. What we describe as the "wetness" of water is an emergent property of the interaction of many, many water molecules. Similarly, consciousness has been widely explained as an emergent phenomena of a critical connection of individual neurons.[93,94] Self-organization, e.g., as in flocks of birds, swarms of insects, and traffic flow patterns, is also typically categorized as emergent behavior.[95,96] The concept of emergence also explains phase transitions in which the macro properties of a system change from one state to another, e.g., the transition from ice to liquid water with no fundamental change to the underlying water molecule.

Phase transitions (phase changes) have been used to describe and model diverse phenomena including the spread of wildfires, innovation, protein folding, epidemic spreading, the origins of life, cancer dynamics, collective intelligence (the wisdom of the crowd), social dynamics, contrail formation, ecological systems, and the behavior of the mosh pit at a heavy metal concert.[97-100] With regards to the present topic of risk management in flight test, phase transitions are useful as they easily explain cascade failures, a failure in which a failure or decision at one node rapidly leads to cascading failures ultimately resulting in a catastrophic, system-wide failure.[101]

An important aspect of the phase-change model of mishaps is that a node failure does not have to occur to create a system-wide failure. The precipitate cause of the system-wide failure can be a

completely logical decision at an individual node. For example, the decisions to not test the blow-out preventer or to reinterpret unclear test results from a third negative-pressure test were logical, "locally" expedient decisions to avoid further, "unnecessary" delays on a project that was 43 days late and more than $21 million over budget.[102] But these decisions ultimately resulted in the blow-out, explosion, and fire that destroyed the Deep Water Horizon rig. An organizational phase transition to a failed state can quickly result from individual nodes making decisions that optimize local conditions without appreciating the system-wide cascade. Complex systems are more prone to phase transitions to system-wide failures because individual nodes, by definition of complexity, cannot understand the response of the entire system to local decisions. Contrast this with the traditional, simplistic view that an accident results from the chain-of-dominoes in which a single component breaks or an individual is guilty of human error.

Systems undergoing phase changes are controlled by one or more macro control parameters. Phase changes occur in systems in which two or more physical forces are driving the system towards different states. In the case of water at a given volume, phase changes are completely described by the control parameters of temperature and pressure. The phase change occurs when the force driving a system towards solid ice (intermolecular forces) are overwhelmed by the force driving the system to a state of higher entropy. Phase change models have been convincingly applied to financial markets to explain sudden collapses, e.g. the 2008 financial market liquidity crisis brought on by the collapse of subprime mortgage values and collateralized debt obligations.[103]

We have introduced the phase-transition model because it uncannily explains unexpected mishaps when the conditions leading to the mishap might have existed for years without harm. Reason's swiss cheese model does the same, i.e., the holes 'finally' line up, but the phase change framework offers the advantage and opportunity to identify the forces and control parameters. In the flight test enterprise and risk awareness management framework, the forces of schedule pressure and limited resources are arrayed against the force to provide a safety margin in flight test operations. In a complex, uncertain system and against the strong and persistent gradients of schedule and resource constraints, it is not surprising that organizations will experience a *drift* towards reduced safety margins and potentially experience catastrophic phase change, i.e. a mishap.

3.2 Drift

Organizational drift was introduced in a landmark 1997 *Safety Science* paper by Jens Rasmussen. Organizational drift is "the effect of a systematic migration of organizational behavior under the influence of pressure toward cost-effectiveness in an aggressive, competitive environment."[104] If the drift is significant enough, an accident or mishap occurs. A phase transition occurs when one of the

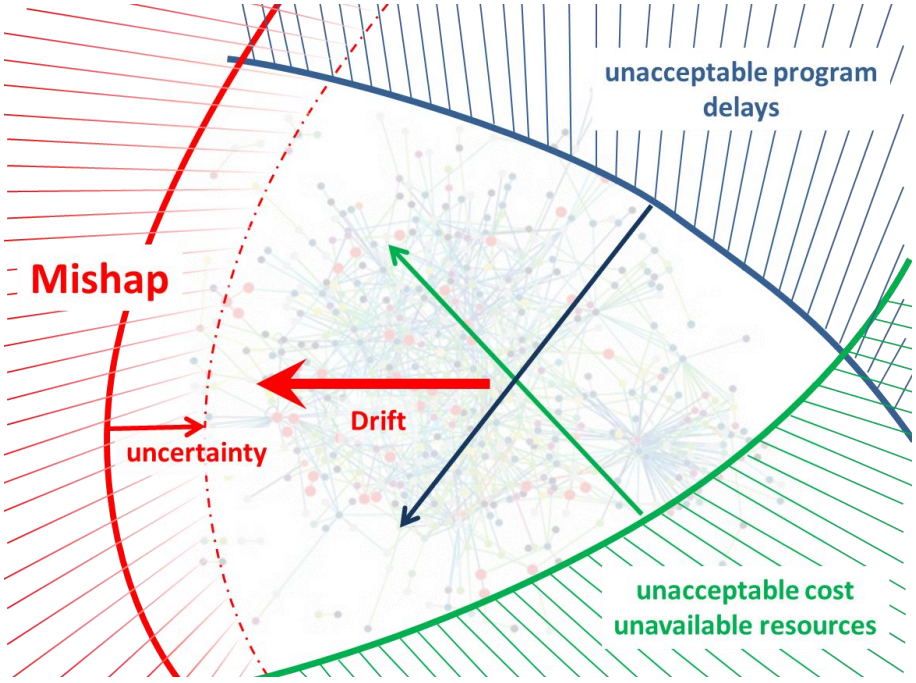


Figure 5: Drift and the gradients experienced in flight test

opposing forces driving a system towards different state dominates. These opposing forces are the control parameters and the point at which one of them emerges as dominant over the other is known as the critical threshold. In the flight test drift framework, the gradients of schedule pressure and limited resources for acquiring knowledge of the system under test are opposed to the tester's desire to understand the system in the face on uncertainty (*Figure 5*). Due to uncertainty, the boundary where a mishap will occur is also unknown.

With liquid water, sufficiently decreasing the control parameter (temperature) will cause the water to freeze. The addition of salt or another solute may delay the phase change, but the water will still freeze at the critical temperature. In flight test, the primary control parameter is knowledge of the system under test. If you reduce the ability of testers to understand the system, either by time (due to aggressive schedule) or by limiting build-up test and characterization (due to resource limits) below a critical threshold, an accident will result as certain as water freezes below the freezing point. The particular challenge of flight test is that because it inherently operates in the domain of uncertainty, the critical threshold of the control parameter (system knowledge) is unknown.

We briefly examine two examples of drift to explain accidents: Alaska Airlines Flight 261 and the loss of the Space Shuttle Columbia.

3.3 Example of Drift – Alaska 261

On 30 Jan 2000, an MD-83 operated by Alaska Airlines took off from Puerto Vallarta enroute to Seattle with a planned stop in San Francisco. After leveling off at FL310, the horizontal stabilizer jammed and the crew began preparing for a divert into LAX. Attempts to clear the jammed stab using

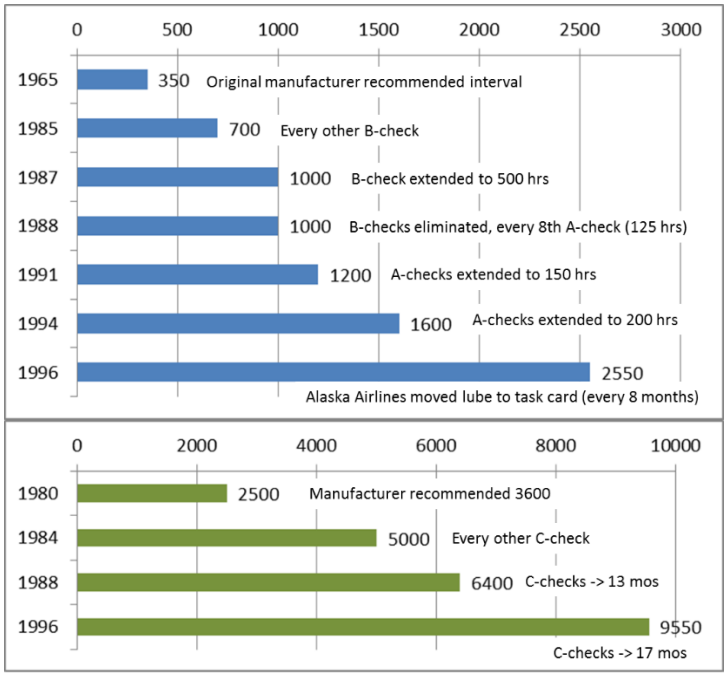


Figure 6: Gradual time interval extension (hrs) for lubricating the MD-83 jackscrew (top) and end-play check (bottom)

primary and alternate trim were initially unsuccessful. When the jam finally cleared after another trim input, the aircraft rapidly pitched nose down but the pilots were able to recover level flight around 24,000 ft. Ten minutes later in conjunction with an audible thump recorded by the cockpit voice recorder, the MD-83 pitched nose down again and impacted the Pacific Ocean in a near vertical dive off the coast of Port Hueneme. All 88 people on board were killed.[105]

The National Transportation Safety Board (NTSB) readily determined that the threads of the acme nut on the horizontal stabilizer trim jackscrew had failed (the stripped threads from the nut were wrapped around the jackscrew recovered from the ocean floor). The NTSB concluded that excessive wear of the nut and jackscrew caused by inadequate lubrication was the primary causal factor. The jackscrew functioned as primary structure supporting the horizontal stab. When the acme nut failed, airloads pushed the stab leading edge up resulting in the violent nose down pitching moment. The design and certification of the MD-83 trim design was based on the DC-9 (certified in 1965) and was common to all MD-80/90 aircraft as well as the 717. The original jackscrew lubrication interval for the DC-9 was 350 hours (*Figure 6).* In 1985, Alaska Airlines received FAA approval to lubricate the jackscrew every other B-check equating to a 700-hr lubrication interval. Two years later, B-checks were extended to 500 hours resulting in a 1000-hr lubrication interval. When B-checks were eliminated a year later, the lubrication task was moved to every 8th A-check, still a 1000-hr lubrication interval. In 1991, the A-check interval was extended resulting in a jackscrew lubrication interval of 1200 hrs. Three years later, the A-check interval was extended again resulting in a lubrication interval of 1600 hrs. Two years later, in 1996, Alaska Airlines moved the lubrication interval to a task card that was accomplished every 8 months. This equated to an average jackscrew lubrication interval of 2550 hrs. Thus, over the course of 30 years, the lubrication

interval for the jackscrew—a primary structure component—was gradually extended from the originally certified 350 hours to 2550 hours! This is the essence of drift. No single, small change to the interval seemed unreasonable at the time—these were rational, locally-optimal decisions—but the combined, cumulative effect of the drift over time resulted in insufficient safety margins.

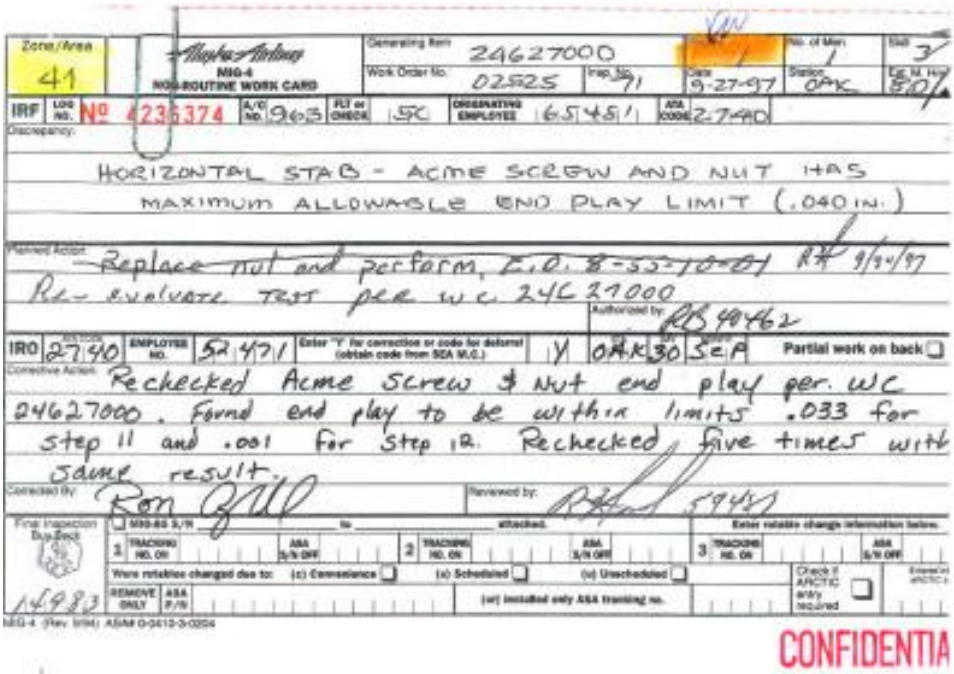


Figure 7: Alaska Airlines Sep 1997 work card for the aircraft involved in the Jan 2000 mishap.

Acme nut wear was a known issue with the jackscrew design. "In 1966, one year after the DC-9 went into service, the discovery of several [jackscrew] assemblies with excessive wear resulted in the development and implementation of an on-wing end-play check procedure to measure the gap between the acme screw and nut threads as an indicator of wear."[106] The acme nuts were designed with a softer metal than the jackscrew so that the nut would wear first and could be replaced when the end-play check exceeded a gap of 0.040 in. Similar to the gradual drift observed with the jackscrew lubrication interval, the end-play check was gradually extended over 26 years. The last end-play check of the Flight 261 aircraft was in 1997, almost three years before the accident. The measured end play was 0.040 in, right at the limit for replacement (*Figure 7*). The corrective action on the task card originally called for the acme nut to be replaced since it was just at the gap limit. Delays in getting the replacement part led the maintenance team to repeat the end-play check at which point they measured 0.033 in. Since this was in limits, the aircraft was returned to service without replacing the acme nut. Once again, this was locally-optimal decision that was rational from the maintenance standpoint, but ultimately represented gradual drift beyond the margin of safe operations.

3.4 Example of Drift – Columbia

The design specifications and criteria for the shuttle were explicit and clear: the shuttle "shall be designed to preclude the shedding of ice and/or other debris from the Shuttle elements during prelaunch and flight operations."[107] Despite this design requirement, foam shedding from the external tank during launch was a recurring problem. Of 79 shuttle launches for which photographic evidence was available, foam shedding was observed on 65 of the launches. At least 143 divots on the shuttle body were identified following shuttle missions, the majority of which were presumed to have been caused by foam strikes during launch. Debris strikes to the shuttle during launch were originally regarded as a safety issue. But over time, "NASA and contractor personnel came to view foam strikes not as a safety of flight issue, but rather a simple maintenance, or 'turnaround' issue."[108] The recharacterization of foam strikes from a safety of flight to a maintenance issue is a distinct example of drift. There was no real change in knowledge or reduction in uncertainty regarding the potential effect of foam strikes on the integrity of the fragile shuttle tiles. However, the organization drifted over time due to a combination of continuous schedule pressure and no investment in knowledge to more fully understand the potential danger of debris strikes. Instead, in a clear demonstration of the turkey illusion, "with each successful landing, it appears that NASA engineers and managers increasingly regarded the foam-shedding as inevitable, and as either unlikely to jeopardize safety or simply an acceptable risk."[109]

The Columbia Accident Investigation Board (CAIB) also cited organizational barriers to communication between engineers and managers. It is noteworthy that this was also a key finding of the Challenger disaster more than a decade earlier. Following the launch of Columbia, the Debris Assessment Team was sufficiently concerned of photographs showing that "out of family" foam shedding had struck the orbiter to initiate three independent requests for imagery of the leading edge.[110] Shuttle managers canceled the imagery requests, apparently "more concerned about the staff following proper channels than they were about the analysis."[111]

The CAIB concluded that "reliance on past success as a substitute for sound engineering practices (such as testing to understand why systems were not performing in accordance with requirements)" was a factor in the mishap.[112] Representative tests, such as shooting foam at high speed to impact shuttle tiles and leading edge structure which were done after the mishap, could have reduced the uncertainty regarding the foam strikes. Knowledge is the control parameter to resist drift and prevent mishaps. But knowledge in one part of a complex system does not necessarily percolate system wide. To ensure relevant system knowledge percolates sufficiently through an organization to prevent mishaps from occurring, system-wide risk awareness is required.

4 Risk Awareness

If situational awareness is the perception of the "elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future,"[113] then let us define risk awareness as the *perception of the elements of uncertainty and the potential, projected outcomes resulting from uncertainty*. In flight test, risk is often described as the

combination of consequence and probability of occurrence.[114] The absence of a "likelihood assessment" or "probability of occurrence" in the definition of risk awareness is deliberate. Since we are dealing with uncertainty, any probability assigned to a projected outcome will be unreliable due to the fact that we are dealing with fundamental ignorance. To put it bluntly, most probability estimates in traditional risk determinations are worth little more than wild guesses. The probability of a simultaneous loss of engine thrust and asymmetric leading edge slats—the root cause of a 1979 DC-10 crash at Chicago O'Hare, still the deadliest single airline accident on US soil—had been estimated to be less than one in a billion by McDonnel Douglas engineers.[115] And yet those exact failures occurred four times in four years before McDonnel Douglas introduced a fleet-wide fix for the DC-10 in 1982. Recognizing the possible consequences of uncertainty is far more important than providing a dubious assessment of probabilities of potential outcomes. All consequences of uncertainty should be considered regardless of their likelihood. The relevant question for safety boards should be "can it happen" instead of "how likely is it?" The crucial element of risk awareness is not the probability estimate of something going wrong but the awareness that there exist unknown outcomes that must be mitigated.

In risk management for flight test, knowledge is the control parameter. As certainly as decreasing temperature will lead to water freezing, decreasing knowledge, and by extension, risk awareness, of a system or program will eventually precipitate a mishap. To prevent mishaps in flight test, leaders and organizations must cultivate a culture of relentless knowledge seeking and risk awareness.

As noted above in the case of Columbia, knowledge does not necessarily percolate across a system. Assessing extent of system-wide knowledge is necessary for the risk manager to assess the risk awareness of the organization. Risk awareness is an emergent, system property of knowledge at each of the nodes in the system. Before Cools' fatal F-22 mishap, several planning methods and risk management practices for dive planning and negative-P_s test points had spread through some portions of the flight test community, but these methods were not widely adopted.[116,117] Had these flight test planning approaches percolated across the system, the F-22 negative-P_s points would have been planned differently. After Cools' accident, an improved appreciation of the risk of high-speed dives was more widely acknowledged and the adoption of TSM was quick. The unfortunate failure of knowledge to percolate through systems, organizations, and communities is not unusual.

During development of the DC-10, McDonnell Douglas was aware of a design flaw of the cargo door locking mechanism.[118] The cargo doors were designed to open outward requiring an elaborate locking mechanism to seal and secure the doors when the fuselage was pressurized. There were multiple independent warnings of the risk associated with the flawed design from various sources including a Dutch engineer in 1969, a subcontractor in 1970, a 1970 precertification pressurization ground test in which the aft cargo door blew off, and repeated entries in maintenance logs concerning difficulties in securing the door. Despite knowledge of the design flaw during development and a relatively simple and straightforward fix, no design changes were made. Schedule pressure from the direct competition with the Lockheed L-1011, also under development at the same time, is a likely

explanation. But knowledge is also a powerful antidote. When he took off in command of American Airlines Flight 96 on June 12, 1972, Captain Bryce McCormick knew that the DC-10 lacked a primary flight control manual reversion system. He was sufficiently concerned to have developed in the simulator, completely on his own initiative, a method for using differential throttles to control the aircraft in event of a total hydraulic failure.[119] After taking off out of Detroit, the incorrectly latched aft cargo door departed the aircraft. The resulting decompression buckled the floor and bound the flight control cables. McCormick's quick reaction with throttle prevented departure from a hard-over rudder. The crew safely landed the plane in Detroit with only ailerons, the right elevator, and differential throttle.[120] Less than two years later, an identical cargo door failure would result in the crash of Turkish Airlines Flight 981 outside of Paris and the loss of 346 passengers and crew. Knowledge can overcome fatal design flaws, but it must percolate through the system.

Leaders can use assessments of knowledge about the system under test to gauge risk awareness. What is known? What are confidence intervals and uncertainty bounds on what is known? What is unknown? Is the uncertainty aleatory or epistemic? Are there build-up tests or analysis that can reduce uncertainty? What is the state of knowledge across the organization? Does knowledge percolate across the system? We now offer guidelines for increasing risk awareness within an organization.

4.1 Cultivating Risk Awareness

Operations are either safe or they are not. Knowledge of the system under test is the control parameter governing the phase change between safe and unsafe operations. Uncertainty, fundamentally inherent in the nature of test, is the primary impediment hampering knowledge of the system and hence risk awareness. To cultivate risk awareness:

1) Identify and characterize the nature of the unknowns
2) Reduce the reducible ignorance
3) Democratize safety decision making
4) Resist drift

1) *Identifying and characterizing the nature of the unknowns*: Due to the different nature of aleatory and epistemic uncertainty and the different cognitive approaches appropriate in the different domains, it is necessary to not only identify what is unknown and uncertain but to also characterize the nature of uncertainty. THA development, ambiguity and fault-tree exploration, pre-mortems and pre-accident investigations are all techniques for exposing and exploring uncertainty. This is often best accomplished in group brainstorming sessions to harness collective intelligence and capture the wisdom of the crowd.[121,122] Scenario planning is a particularly useful technique used by several flight test organizations. Table top discussions and chair flying missions with various possible scenarios, outcomes, corrective actions, and debriefs with the crew, control room, design engineers, and maintenance personnel is a useful method for identifying where uncertainty exists. When making risk management judgments, recall the different cognitive approaches appropriate to

the nature of uncertainty. In the *risk domain*, be wary of cognitive biases and carefully apply slow thinking to arrive at risk-balanced decisions. In the *pure uncertainty* domain, beware of the fallacy of the turkey illusion and apply heuristics with humility.

2) *Reduce the reducible ignorance*: Within the *pure uncertainty* domain there is reducible and irreducible ignorance. Build-up tests, component tests, system tests, analysis, similar comparisons, and thorough data reviews are useful means for reducing reducible ignorance. The increase in knowledge about the system under test is obviously limited by programmatic resource and schedule constraints. By defining the minimum acceptable boundaries for knowledge and uncertainty about the system, the risk acceptance authority can use knowledge as an assessment and control parameter in the tradeoff between safety margin and schedule pressure. A general principle of reducing readily reducible ignorance is a sound investment of resources.

The Apr 2015 total declared loss of an AC-130J due to a 200% design limit load exceedance reinforces the importance of reducing reducible ignorance. The mishap aircraft departed controlled flight during a steading heading sideslip test point due to fin stall and aerodynamic locking of the rudder. Following the directional departure, during which sideslip may have reached as much as 56 deg of beta, the subsequent recovery from an inverted and extremely nose low attitude resulted in a load factor of 3.19 g's and a gross overspeed. The aircraft landed safely but was subsequently declared a complete loss. The test was designed to assess the flying qualities of the AC-130J which had several relevant modifications from the C-130J. Following a similar but less severe yaw departure in Feb 2014, the test team had requested predictive flight data from previous C-130J testing. For a variety reasons, including a perceived lack of government data rights to the prior test results, the team never obtained the sideslip performance predictions. "With a lack of predictive data, members of the team admitted they did not understand how much stability margin existed at the second special alert, but admitted it could potentially be very little."[123] The test team recognized and identified the reducible ignorance, but without actually reducing the ignorance the team was unable to avoid the mishap.

3) *Democratize safety decisions*: If safety is an emergent property of an organization, then knowledge and the attributes of all the collected nodes determine the overall system safety. Every individual in an organization must be a risk manager. Some well-known, successful flight test organizations within the industry go so far as to askew separate safety officers and safety functions to underscore the mindset that "every individual is a safety officer." An analogous approach to distributed safety is the "shared space" concept of the late traffic engineer Hans Monderman. Somewhat paradoxically, pedestrian-vehicle accident rates in shared spaces are decreased by removing street signs, lane lines, curbs, cross walks, and marked pedestrian zones.[124] Removing explicit constraints to driving forces drivers to be more cognizant, aware, and pro-active around pedestrians and reduces overall accidents.[125] In a shared space, everyone is forced "to own" part of the overall safety of the system.

It is important to avoid the potential trap of democratization becoming an abdication of responsibility. Two examples of the "it is someone else's responsibility" illustrate the risk. In the case of the F-15 friendly fire shoot down of the two Black Hawk helicopters previously discussed, an AWACS plane full of air battle managers who had previously been aware of friendly UH-60s in the area as well as the fact that they were squawking incorrect IFF codes did not intervene to prevent the engagement. Similarly, the accident report for the 1966 XB-70/F-104 mid-air collision highlights the fact that at least six different officials in the leadership and execution chain of command had the knowledge and the authority to cancel the unapproved photo opportunity but failed to do so.

For democratized decisions to be effective, knowledge and appreciation for what is uncertain must freely percolate throughout the organization. Individual workers have to own as much of the safety of operations as they can. As a result of the inherent difficulty in understanding complex organizations and systems, this democratization of safety is most easily achieved with teams that are as small as possible. Every additional node in a complex system is an n-factorial increase in the number of potential interactions between nodes. Minimizing the number of people minimizes the organizational complexity. Likewise, information flow across an organization is enhanced by a reduction in the number of nodes and the number of hierarchal levels in an organization. "If one is seeking to build relationships, engage in risk conversations and focus on learning, then one needs good questions that promote dialogue and encourage conversations. Engagement, conversation and listening are transformative because they create learning and community."[126] Consider removing "safety controls" that only add complexity without controlling risk. Reward safety concerns that were organically discovered and mitigated. Encourage self-reporting. Do not punish mistakes or human errors unless they were of a deliberate, criminal nature.

4) *Resist Drift*: Recognize that there will be persistent gradients from schedule and resource constraints that will erode safety margins. At its heart, flight test is about understanding the unknown. Flight test matrices should be deliberately planned to reduce the reducible ignorance and then steadfastly defended against schedule pressure. Resist the temptation to eliminate test points because you think you understand the system in areas that have not yet been tested. Do not fall victim to the turkey illusion. An envelope expansion plan that has cleared 95% of the envelope is still as uncertain about the remaining 5% as it was before the first flight of the system.[127]

Building realistic and credible schedules (avoiding optimism bias) and defending test schedule is important in resisting drift. Since flight test operations normally pick-up during later stages of system development, and program delays in system development are common, there is always pressure to recapture program schedule during the flight test phase. This is an unfortunate conjunction since many of the greatest uncertainties are not realized until test. One Air Force Flight Test Squadron uses an established battle rhythm for pacing the test planning and execution: test plan working group (not later than 90 days prior to test), detailed test plan (NLT 60 days), test support plan (NLT 45 days), technical and safety review board (NLT 30 days), test cards (NLT 3 days), brief

(NLT 1 day). Routinely and repeatedly falling behind the pacing timeline is an indication and trend warning that schedule pressure is mounting and helps foster enhanced squadron risk awareness.

Surprises are warnings that either you do not understand the system, that you have uncovered previously unappreciated uncertainty, or that the organization has drifted. Understand surprises and update characterization of uncertainty before proceeding further. As Richard Feynman wrote in Appendix F to the Rogers Commission report on the Challenger disaster ("Feynman's Dissent"):

> *if the real probability [of failure] is not so small, flights would show troubles, near failures, and possible actual failures with a reasonable number of trials. and standard statistical methods could give a reasonable estimate. In fact, previous NASA experience had shown, on occasion, just such difficulties, near accidents, and accidents, all giving warning that the probability of flight failure was not so very small.*[128]

Because the circumstances and decisions leading to the Challenger disaster are so well known, it is a useful case study. With the fully acknowledged risk of hindsight and outcome bias, we examine the Challenger launch decision through the risk awareness framework.

4.2 Challenger Disaster in the Risk Awareness Framework

When STS-51-L launched on 28 Jan 1986, the launch had already been postponed three times and scrubbed once. At the time, NASA managers were under intense pressure to maintain an aggressive launch schedule. Christa McAuliffe, the first teacher in space, was a crew member and the President's State of the Union Address was scheduled for the night of 28 Jan.[129] A launch temperature of 30° F was forecast for the morning of 28 Jan, the coldest launch yet of the program and at the lower bound of "certified" temperature. O-ring erosion in the field joints of the solid rocket boosters had been a recurring problem and there was speculation by many engineers that the erosion was temperature related. Seven of the previous 24 shuttle flights had experienced some degree of O-ring erosion from exposure to combustion gases and Morton-Thiokol engineers were notably concerned (*Figure 8*). A previous mission launched at 53° F, STS-51-C, had experienced the worst erosion yet with problems at four joints. The 51-C data had been thoroughly reviewed a year before the Challenger launch during the Flight Readiness Review for STS-51-E in Feb 1985 during which the Thiokol engineers concluded that "low temperature enhanced probability of blow-by."[130] Nevertheless, by the time of Challenger's FRR shuttle managers had concluded that O-ring erosion was within "allowable limits." Without justification (knowledge of the system), they reasoned erosion was an acceptable risk because of "redundancy" in the design. The redundancy was the existence of two O-rings. In fact, the O-rings were not designed to erode nor should they have experienced any exposure to hot gas intrusion into the field joint (an insulating zinc-chromate putty between the internal combustion chamber and the O-ring was meant to prevent impingement of combustion gases on the O-ring).

As with the design flaw of the DC-10 cargo door and leading edge slats, knowledge was available about problems with O-ring erosion and blow-by. There was understandable uncertainty regarding the launch decision, but risk awareness of the problem was low. O-ring erosion was more thoroughly discussed at the STS-51C FRR the prior year (launch temperature 53° F) than at the STS-51L FRR (30° F). The first step of risk awareness, characterizing the uncertainty, was not done. To the extent

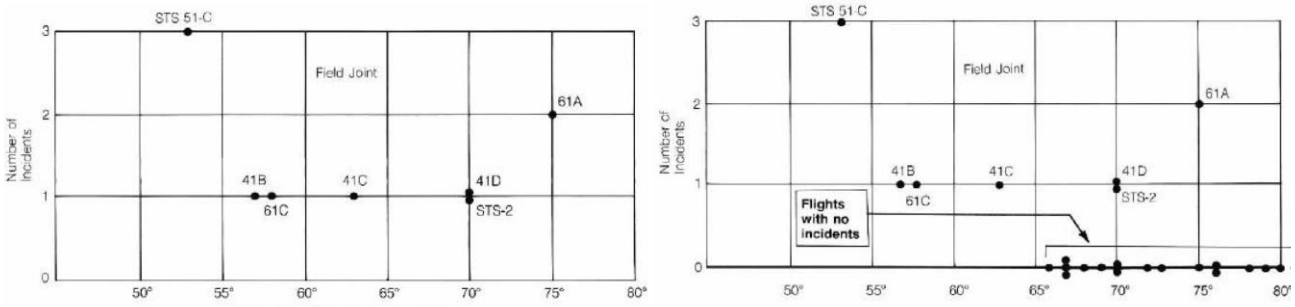


Figure 8: Incidents of O-ring "distress" (erosion) as a function of launch temperature. Left: this is the data considered by managers on the night before the launch. Presentation of data in this format which only includes missions experiencing O-ring erosion, is misleading (this is sampling on the dependent variable). Right: plot of all missions clearly shows the correlation and relationship with temperature.[117]

that it was considered, O-ring erosion was considered to be an aleatory, or *risk domain*, uncertainty. Erosion events from previous flights were improperly presented (the data was sampled on the dependent variable, a statistical sin) and used to assess the risk. In fact, uncertainty of O-ring erosion should have been classified as *pure uncertainty.* The origin and consequences of the erosion and blow-by were not understood. O-ring erosion was not part of the design and should have been a distinct clue that something was wrong. Partial erosion was not something from which safety can be inferred. As with foam strikes on Columbia, we see the turkey illusion in action. A deeper appreciation of the uncertainty of launching in the coldest temperatures ever may have spurred managers to seek additional data to understand the phenomena, to reduce the reducible ignorance. But in the free-body diagram of schedule and resource gradients aligned opposite to understanding the system, knowledge was the victim. The corresponding lack of risk awareness led to a decision to launch the Challenger.

NASA is far from a democratic organization. The immense complexity of shuttle missions and the large bureaucracy necessary to review all aspects of a complicated engineering system prior to launch make democratization of decisions at NASA particularly challenging. Nevertheless, reducing the number of layers and hierarchy in organizations can improve communication flow through the organization and the likelihood that the right knowledge percolates through the system. Organizational structural problems were cited in the Rogers Commission report. The report noted that "project managers, because of a tendency at Marshall [Space Flight Center] to management isolation, failed to provide full and timely information bearing on the safety of flight 51-L to other vital elements of Shuttle program management." The commission further recommended that the "flight crew commander, or a designated representative, should attend the Flight Readiness Review, participate in acceptance of the vehicle for flight, and certify that the crew is properly prepared for flight."[132] It is remarkable that the flight crew did not participate in the FRR. The lack of democratized safety at NASA has been well documented. In his personal appendix to the Rogers Commission report, Feynman highlights some of the cultural problems: "why do we find such an enormous disparity between the management estimate and the judgment of the engineers? It would appear

that, for whatever purpose, be it for internal or external consumption, the management of NASA exaggerates the reliability of its product, to the point of fantasy."[133]

Planning fallacy and optimism bias likely contributed to the shuttle program's ambitious launch schedule at the time of the Challenger disaster. An overestimate of the level of understanding of the design and performance after 24 successful launches, the turkey illusion, would have made it difficult to resist drift against the program pressure and resource constraints. The Rogers Commission found that

> *In establishing the schedule, NASA had not provided adequate resources for its attainment. As a result, the capabilities of the system were strained by the modest nine-mission rate of 1985, and the evidence suggests that NASA would not have been able to accomplish the 15 flights scheduled for 1986. These are the major conclusions of a Commission examination of the pressures and problems attendant upon the accelerated launch schedule.*[134]

Knowledge is the primary control parameter in the phase transition between safe and unsafe operations. Risk awareness is the perception of uncertainty and its potential outcomes. Assessing the level of system-wide knowledge is one of the best ways of assessing the risk awareness of an organization. There will always be tradeoffs between learning more about a system and limited resources for acquiring knowledge as well as the schedule pressure to move a program forward. Resisting drift is the final defense in risk awareness. Safety boundaries for given levels of uncertainty are established at the beginning of most programs and must be vigorously defended. Recognizing the critical difference between true knowledge and residual uncertainty throughout a program is a key mindset for making sound decisions under uncertainty. The decisions of the Challenger risk management team demonstrate significant deficiencies in all four areas of fostering robust risk awareness.

5 Conclusions

"You don't know you've lost SA until you start to get it back" is a fighter pilot truism. The same can be said for risk awareness. One can never be certain that they are risk aware as opposed to simply lucky until an accident exposes the lack of risk awareness. But just as situational awareness can be developed over time and improved with experience, risk awareness can be fostered through an understanding of uncertainty and cognitive tools. A distinction between different types of uncertainty lies at the heart of the risk awareness framework. Understanding if a decision is being made in the stochastic risk domain or the pure uncertainty domain is important if the decision maker is to avoid cognitive biases or errors in judgment. THAs and scenario planning tools are effective in the risk domain, but this is only a start. Most of the hard decisions in flight test occur in the domain of pure uncertainty. In this domain, flight testers need a deep respect for what is unknown. Heuristics in the domain of uncertainty are likely to provide better go/no-go decisions so long as they are applied with humility and respect for the unknown.

In the risk awareness framework, knowledge is the primary control parameter governing the phase transition between safety and accidents. Decreasing knowledge of a system will precipitate a

mishap just as surely as decreasing the temperature of water will cause it to freeze. The framework of knowledge as control parameter provides a ready assessment tool for the risk manager. What do we truly know? How well does the team know the system? What is unknown? What is the nature of the unknown? Are there build-up tests or analysis that can reduce the reducible ignorance? What is the state of knowledge across the organization? Is knowledge flowing between nodes? Are there nodes that are disproportionally lacking knowledge? As so many accidents have proven, locally-logical or even optimal decisions at a node can cause system-wide failures. This is the analog to sensitive-dependence-on-initial-conditions in complex systems. System knowledge, knowledge density at nodes, and knowledge flow between nodes is the antidote.

Awareness of the tendency to drift and erode safety margins is also absolutely essential for test organizations. Schedule and resource constraints will always create strong and persistent gradients countering uncertainty reduction of a system under test. Recognizing drift and resisting the pressure to give away safety margin in the interest of program schedule is another important aspect of the risk awareness framework. In many cases, information is ambiguous and the implication of tradeoffs are unclear, but framing the question in terms of knowledge helps distill the decision to objective criteria and provides a common perspective to challenge program managers and other test customers.

The risk awareness framework will not prevent all future accidents in flight test. Test is the exploration of uncertainty. By the very nature of pure uncertainty, there will be outcomes that were unpredictable. Risk awareness focuses attention of the organization and decision makers on what is truly known and not known. By fostering respect for the nature of uncertainty, approaching the unknown with humility and a desire to reduce reducible ignorance, the risk awareness framework seeks to—as wisely as possible—manage the inherent risk of flight test. Our value as flight test professionals is our ability to manage risk. Flight test professionals are ultimately professional risk managers. We must never forget that uncertainty lies at the heart of test. We must first not fool ourselves—and we are the easiest person to fool.

6 Acknowledgements & Disclaimer

The author wishes to thank Professor David Hofmann, University of North Carolina, for suggesting the utility of heuristics under uncertainty and for introducing the author to the work of Gigerenzer and Dekker. The seeds of many good ideas sprung from discussions with Professor Hofmann and his lectures on organizational decision making. The author also wishes to thank Col Matthew “Sieg” Higer, former Commandant of the USAF Test Pilot School, for his insightful and deep intellectual review of the risk management framework as well as recommendations for several additional examples. Finally, the author owes great thanks to Col Douglas “Crack” Creviston, Lt Col Matthew “KITT” Caspers, Lt Col Michael “HAVOC” Nielsen, and Lt Col Ryan “Hulk” Sanford for their review and suggestions. All are extraordinary Air Force leaders and the author learned much from observing their exemplary management of risk in flight test.

The views expressed are those of the author and do not reflect the official policy or position of the U.S. Air Force, the Department of Defense, or the U.S. Government. This material has been approved for public release and unlimited distribution.

---

[1] Feynman, Richard, "Cargo Cult Science," Caltech 1974 Commencement Address, Jun 1974.

[2] Kelly, N., "A Test Pilot's Thoughts on Aircraft Accident Prevention," *Cockpit* (Feb), 1967.

[3] Galipault, J., "An Investigation of Risk-Acceptance and Pilot Performance During Low Altitude Flight," *Cockpit*, May 1967.

[4] Lowry, R., "Risk Management Comparison and Similarities Between the Hollywood Stunt Business and Aircraft Flight Testing," SETP Paper Database, 2001.

[5] Hall, G.W., "The Test Pilot's Role in Risk Management," *Cockpit* (Jul/Aug/Sep), 2001.

[6] Wickert, D.P., "Mitigating Risk at the Edge of the Envelope: Revising 46 TW Basic Aircraft Limit Test Procedures," *Proceedings to the Society of Experimental Test Pilots* 50th Annual Symposium, Sep 2006.

[7] Tomassetti, A., "Saying Yes to the 'No Vote'," *Proceedings of the 52nd Annual SETP Symposium*, 2008.

[8] Prosser, K., "Lower Risk Surrogate UAV Flight Testing Lessons Learned from a UAV "Instructor" Pilot," SETP Paper Database, 2008-2009.

[9] Hehr, R., "Hidden Risk," *European 46th SETP and 25th SFTE Symposium, 15-18 June 2014, Luleå, Sweden*, 2014.

[10] Culbertson, D., Foster, B., McMahon, R., Schifferle, P., "Contact! Risk Management Lessons Learned from the Navy AAR Program," *Proceedings of the 58th Annual SETP Symposium*, 2014.

[11] Javorsek, Daniel, "Modernizing the Safety Process: Complexity Theory for Flight Test Professionals," *Proceedings of the 59th Annual SETP Symposium*, 23-26 Sep, 2015.

[12] Meier, Michael, "Lessons Learned and Murphy's Corollary," *Proceedings of the 59th Annual SETP Symposium*, 23-26 Sep, 2015.

[13] Recording from the workshop are available at: http://www.flighttestsafety.org/2018-arlington-tx

[14] Partial list compiled from various sources (fatalities in parenthesis): Gulfstream GVI, Apr 2011 (4); AS532 Cougar, Jul 2012 (6); Superjet 100, Jul 2013; VSS Enterprise, Oct 2014 (1); AC-130J, Apr 2015; A400M, May 2015 (4); AW609, Oct 2015 (2); Bell 525, Jul 2016 (2); unspecified, Dec 2016; Icon A5, May 2017 (2); Sikorsky S-97, Aug 2017; Cobalt Co50, Sep 2017; unspecified, Sep 2017 (1); F-2000A Typhoon, Sep 2017 (1); A-29 Super Tucano, Jun 2018 (1)

[15] A world catalog search for books on safety and risk management returns more than 10,000 titles in print; additionally, there are currently more than 20 peer-reviewed journals publishing research on safety science and risk management.

[16] Conklin, Todd, *Pre-Accident Investigations: An Introduction to Organizational Safety*, CRC Press, 2012.

[17] Alston, Greg, *How Safe is Safe Enough?: Leadership, Safety and Risk Management*, Routledge, 2017.

[18] Dekker, Sidney, *The Field Guide to Understanding "Human Error"*, CRC Press, 2017.

[19] Dekker, Sidney, *Safety differently: human factors for a new era*, CRC Press, Boca Raton, FL, 2015.

[20] Frameworks are models or compositions of concepts that help guide the understanding and explanation of phenomena. Frameworks provide working hypotheses that are necessary to guide decision making. E.g., see, Ravitch, and Riggan, *Reason and Rigor: How Conceptual Frameworks guide Research*, Thousand Oaks CA: Sage, 2017.

[21] Hollnagel, E. *Barriers and accident prevention,* Aldershot, UK: Ashgate, 2004.

[22] Merlin, Bendrick, & Holland, *Breaking the Mishap Chain: Human Factors Lessons Learned from Aerospace Accidents and Incidents in Research, Flight Test, and Development*, NASA, 2012.

[23] Heinrich, H.W., *Industrial Accident Prevention*, New York: McGraw-Hill, 1931.

[24] Reason, James, "The Contribution of Latent Human Failures to the Breakdown of Complex Systems," *Philosophical Transactions of the Royal Society of London*, Series B, Biological Sciences, 327 (1241): 475–484, 1990.

[25] Archetypes for Organizational Safety by Karen Marais and Nancy G. Leveson. Proceedings of the Workshop on Investigation and Reporting of Incidents and Accidents, September 2003.

[26] Turner, B.A., *Man-Made Disasters*, Wykeham Science Press, 1978.

[27] Turner, B.A. and Pidgeon, N.F., *Man-made Disasters (Second Edition),* Oxford: Butterworth Heinemann, p. 72, 1997.

[28] Perrow, C., *Normal Accidents: Living with High-Risk Technologies*, New York: Basic Books, 1984.

[29] Environments also feature in the epidemiological accident model. The epidemiological accident model describes accidents as the result of failed barriers or weakened defenses being infected by a pathogen from the environment. Latent conditions also feature in the epidemiological accident model.

[30] A New Accident Model for Engineering Safer Systems by Nancy Leveson. Safety Science, Vol. 42, No. 4, April 2004

[31] Leveson, N.G., *A New Approach To System Safety Engineering*, 2002

[32] Pidgeon, N.F., 1991. Safety culture and risk management in organizations. Journal of Cross-Cultural Psychology 22 (1).

[33] Weick, Karl E., K. Sutcliffe, and D. Obstfeld 1999 "Organizing for high reliability". Research in Organizational Behavior, 21: 81–123.

[34] Weick, Karl E. 1987 "Organizational culture as a source of high reliability". California Management Review 29(2): 112–127, Winter

[35] Sagan, Scott, *The Limits of Safety*, Princeton University Press, 1995.

[36] Leveson, N., et. al., "Moving Beyond Normal Accidents and High Reliability Organizations: An Alternative Approach to Safety in Complex Systems," *Organizational Studies*, Volume: 30 issue: 2-3, page(s): 227-249, 2009.

[37] Marais, Karen, and Nancy Leveson, "Archetypes for Organizational Safety," *Proceedings of the Workshop on Investigation and Reporting of Incidents and Accidents*, September 2003.

[38] Rasmussen, Jens, "Risk Management in a Dynamic Society: A Modelling Problem," *Safety Science*, Vol 27, No. 2/3, p 183-213, 1997

[39] Dekker, Sidney, *Drift into Failure: From Hunting Broken Components to Understanding Complex Systems*, Ashgate Publishing, Ltd., 2012

[40] Prof Dave Hofmann (Kenan-Flagler Business School at UNC Chapel Hill) uses the following analogy between physicians and personal trainers: physicians look for the absence of disease to define a state of health; trainers, by contrast, look at physical fitness to determine the state of health.

[41] Hollnagel, E., "A Tale of Two Safeties," Unpublished, Retrieved from http://www.erikhollnagel.com/A_tale_of_two_safeties.pdf, 2012

[42] Hollnagel, E. *Barriers and accident prevention,* Aldershot, UK: Ashgate, 2004.

[43] Klein, G., "Performing a Project Premortem," *Harvard Business Review,* Sep 2007.

[44] Wickert, D.P., "Flight Path Angle and Energy Height Planning for Negative-Ps Test Points," *Proceedings to the SETP 49th Annual Symposium*, Sep 2005.

[45] Gray, W.R., "Time Safety Margin: A generalized method for dive safety planning," *Proceedings to the 54th SETP Annual Symposium*, 2010.

[46] Gray, W.R., *Time Safety Margin: Theory and Practice*, 412TW-TIH-16-01 Technical Information Handbook

[47] ISO 31000:2018, *Risk Management Guidelines,* International Organization for Standardization, 2018.

[48] Ibid.

[49] AIR FORCE INSTRUCTION 90-802, *Special Management, Risk Management*, 11 Feb 2013, certified current on 15 May 2017.

[50] *The Air Force System Safety Handbook*, Air Force Safety Agency, Kirtland AFB NM, Revised July 2000.

[51] This is, admittedly, a binary over-simplification. Safety is not necessarily the absence of accidents, it is the elimination of unacceptable hazards or risks. During phase transitions between two states, it is possible to have both states exist simultaneously in various proportions. Similarly, hazards may exist at varying levels and severity in the transition between completely safe operations and mishaps. If a mishap occurs, the phase transition to the state of unsafe operations is complete.

[52] The Society of Experimental Test Pilots has an X for its logo to symbolize the unknown and the motto of the Air Force Test Center's is *Ad Inexplorata*, "into the unknown."

[53] Sautoy, Marcus du, *The Great Unknown: Seven Journeys to the Frontiers of Science*, Penguin Books, 2017.

[54] Chaotic systems exhibit sensitive dependence on initial conditions. Though deterministic, without infinite resolution of the initial condition, the future is indeterminable beyond a certain limit. The Heisenberg uncertainty principle states that the product of the variance of position and momentum of a particle must be greater than the reduced Plank constant.

[55] Beacher, G.B. and Christian, J.T., "Natural Variation, Limited Knowledge, and the Nature of Uncertainty in Risk Analysis," presented at *Risk-Based Decision making in Water Resources IX*, Oct. 15-20, 2000, Santa Barbara.

[56] This distinction between aleatory and epistemic uncertainty has a long history. The economist John Maynard Keynes made the same distinction in his 1937 "The General Theory of Employment":

> *By "uncertain" knowledge, let me explain, I do not mean merely to distinguish what is known for certain from what is only probable. The game of roulette is not subject, in this sense, to uncertainty. Even the weather is only moderately uncertain. The sense in which I am using the term is that in which the prospect of a European war is uncertain, or the price of copper and the rate of interest twenty years hence… About these matters there is no scientific basis on which to form any calculable probability whatever. We simply do not know.*

[57] Klein, Gary A, *Sources of Power: How People Make Decisions*, Cambridge, Mass: MIT Press, 1998.

[58] Loch, C.H., et. al., *Managing the Unknown: A New Approach to Managing High Uncertainty and Risk in Projects*, John Wiley & Sons, 2006.

[59] An additional, potential category of uncertainty is *equivocality of information*. Professor Karl Weick defines equivocality as ambiguity or the existence of multiple and conflicting interpretations of information. Equivocality results from not knowing how to interpret information that is available or from different expert opinions on the meaning of available information.

[60] Sen, D., "The uncertainty relations in quantum mechanics," *Current science*, 107(2): 203-218, 2014.

[61] Stirling, Andy, "Keep it Complex," *Nature*, Vol 468, Dec 2010.

[62] Lacroix, Kelly M., et. al. "Using Scenario Planning to Prepare for Uncertainty in Rural Watersheds," T*he University of Arizona Cooperative Extension*, Dec 2015

[63] Hansen, Steffen Foss, "Multicriteria mapping of stakeholder preferences in regulating nanotechnology," *Journal of Nanoparticle Research*, 12:1959–1970, 2010.

[64] Van Exel NJA, G de Graaf. "Q methodology: A sneak preview," 2005 [available from www.jobvanexel.nl]

[65] Höllermann, B., Evers, M., "Integration of uncertainties in water and flood risk management," *Proceedings of the International Association of Hydrological Sciences,* 370, 193–199, 2015.

[66] Willows, R., Reynard, N., Meadowcroft, I., and Connell, R., "Climate adaptation: Risk, uncertainty and decision-making," *UKCIP Technical Report*, UK Climate Impacts Programme, Oxford, 41-88, 2003.

[67] Leber, D., Herrmann, J., "Incorporating Attribute Value Uncertainty into Decision Analysis," *Proceedings of the 2012 Industrial and Systems Engineering Research Conference,* 2012.

[68] Herrmann, J., *Engineering Decision Making and Risk Management*, Wiley, 2015.

[69] Schoemaker, P., "When and How to Use Scenario Planning: A Heuristic Approach with Illustration," *Journal of Forecasting,* Vol 10: 549-564, 1991.

[70] Walker, W.E., et. al., "Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support", *Integrated Assessment*, Vol 4., No 1, 2003.

[71] Kahneman, Daniel, *Thinking, Fast and Slow*, Farrar, Straus and Giroux, 2011.

[72] Tversky, Amos, and Kahneman, Daniel, "Judgment under Uncertainty: Heuristics and Biases." *Science,* 185: 1124-1131, 1974.

[73] Tversky, Amos, and Kahneman, Daniel, "On the Reality of Cognitive Illusions." *Psychological Review,* 103: 582-591, 1996.

[74] Gigerenzer, G., *Calculated Risks: How to Know When Numbers Deceive You*, Simon and Schuster, 2002.

[75] Gigerenzer, G., *Risk Savvy: How to Make Good Decisions*, Penguin, 2014.

[76] Gigerenzer, G., & Goldstein, D.G., "Reasoning the fast and frugal way: Models of bounded rationality," *Psychological Review*, 103: 650-669, 1996.

[77] Hall, G.W., "The Test Pilot's Role in Risk Management," *Cockpit Magazine* (Jul/Aug/Sep), 2001.

[78] Cognitive Bias Codex, 2016, https://betterhumans.coach.me/cognitive-bias-cheat-sheet-55a472476b18.

[79] Kahneman, D., & Tversky, A., "Prospect theory: An analysis of decision under risk," *Econometrica*, 47(2): 263-291, 1979.

[80] Kahneman, Daniel, *Thinking, Fast and Slow*, Farrar, Straus and Giroux, 2011.

[81] Gigerenzer, G., and Brighton, H., "Homo Heuristicus: Why Biased Minds Make Better Inferences," *Topics in Cognitive Science I*, 2009.

[82] Todd, P, and Gigerenzer, G., "Environments that make us smart: Ecological Rationality," *Current Directions in Psychological Science,* Vol 16, No 3, p 167-171, 2007.

[83]

[84] Watts, D. J., "The 'New' Science of Networks," *Annual Review of Sociology,* 30: 243-270, 2004.

[85] Johnson, Neil F., *Simply complexity: A clear guide to complexity theory*, Oneworld Publications, 2009.

[86] The mathematical origins of chaos theory date back to 1880 and Poincare's study of the three-body problem. Chaos theory developed further in the 1960s and grew rapidly in the late 1970s and 80s with the increase and availability of computing power. The equations leading to chaos are nonlinear and deterministic, with a primary attribute of sensitive dependence on initial conditions. By contrast, the mathematics of complexity theory are governed by connections between nodes of a network. It is generally non-deterministic with the primary attribute being the emergence of large-scale, system-wide patterns. Nevertheless, there are features shared by both chaos theory and complexity theory, e.g. scale-free, system-wide characteristics; wide-spread nonlinearity and feedback; apparent order from disorder; bifurcation phenomena; and sensitivity to small changes of the system (initial conditions in the case of chaos, cascades in the case of complexity).

[87] Dekker, Sidney, *Drift into Failure: From Hunting Broken Components to Understanding Complex Systems*, Ashgate Publishing, Ltd., 2012

[88] Leveson, Nancy, "A New Accident Model for Engineering Safer Systems," *Safety Science*, Vol. 42, No. 4, April 2004.

[89] Titz, B., et al., "The Binary Protein Interactome of Treponema pallidum – The Syphilis Spirochete," *PLoS ONE* 3(5): e2292, 2008.

[90] Grandjean, M., "Introduction à la visualisation de données, l'analyse de réseau en histoire," *Geschichte und Informatik 18/19*, pp. 109-128, 2015.

[91] Kim, T., Pentland, A., et al. "Sensor-Based Feedback Systems in Organizational Computing," *'09 International Conference on Computational Science and Engineering,* 2009.

[92] Goldstein, J., "Emergence as a Construct: History and Issues," *Emergence,* 11, 49–72, 1999.

[93] Wilson, E.O., *Consilience: The Unity of Knowledge*, New York: Knopf, 1998.

[94] Godwin, D., Barry, R., and Marois, R., "Breakdown of brain's networks with awareness" *Proceedings of the National Academy of Sciences*, Mar 2015.

[95] Kerner, B., "Experimental Features of Self-Organization in Traffic Flow," *Physical Review Letters*, 81 (17): 3797-3800, 1998.

[96] Li, Daqing, et al. "Percolation transition in dynamical traffic network with evolving critical bottlenecks." *Proceedings of the National Academy of Sciences,* 112.3 (2015): 669-672.

[97] Schroeder, Manfred R., *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise*, W.H. Freeman, 1991.

[98] Sole, Richard. V., *Phase Transitions*, Princeton University Press, Princeton, 2011.

[99] Bahcall, Safi, *Loonshots: How to Nurture the Crazy Ideas That Win Wars, Cure Diseases, and Transform Industries*, St Martin's Press, 2019 (upcoming).

[100] Silverberg, J., et. al., "Collective Motion of Humans in Mosh and Circle Pits at Heavy Metal Concerts," *Physical Review Letters,* Vol. 110-22, May 2013.

[101] Buldyrev, S. V., et.al., "Catastrophic cascade of failures in interdependent networks," *Nature,* Vol 464, 2010.

[102] "US Chemical Safety and Hazard Investigation Board Report, Executive Summary," *Drilling Rig Explosion and Fire at the Macondo Well,* Report No 2010-10-I-OS, 04/12/2016.

[103] Yalamova, R., and Bill McKelvey, "Explaining What Leads Up to Stock Market Crashes: A Phase Transition Model and Scalability Dynamics," *Journal of Behavioral Finance*, 12:3, 169-182, 2011.

[104] Rasmussen, Jens, "Risk Management in a Dynamic Society: A Modelling Problem," *Safety Science*, Vol 27, No. 2/3, p 183-213, 1997

[105] NTSB/AAR-02/01: *Aircraft Accident Report, Loss of Control and Impact with Pacific Ocean Alaska Airlines Flight 261 McDonnell Douglas MD-83, N963AS, January 31, 2000,* National Transportation Safety Board, Dec 2002.

[106] Ibid. p 22.

[107] Columbia Accident Investigation Board, Report Volume I, Aug 2003.

[108] Ibid. p 122.

[109] Ibid.

[110] Ibid. Ch 6.

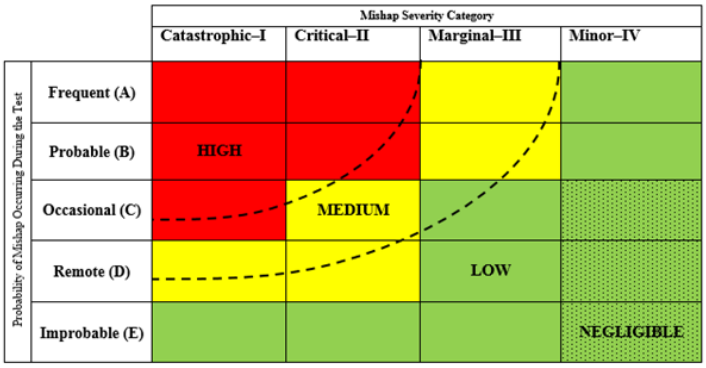
[111] Ibid. F6.3-19, p 172.

[112] Ibid. p 9.

[113] Endsley, M. R., "Design and evaluation for situation awareness enhancement," *Proceedings of the Human Factors Society 32nd Annual Meeting*, Santa Monica, CA, 1988.

[114] This is in accordance with MIL-STD-882E, 11 May 2012, which defines risk as the "combination of the severity of the mishap and the probability that the mishap will occur." The MIL-STD-882E risk assessment matrix is on the left below. The AFTCI 91-202 risk assessment matrix is on the right.

RISK ASSESSMENT MATRIX				
SEVERITY / PROBABILITY	Catastrophic (1)	Critical (2)	Marginal (3)	Negligible (4)
Frequent (A)	High	High	Serious	Medium
Probable (B)	High	High	Serious	Medium
Occasional (C)	High	Serious	Medium	Low
Remote (D)	Serious	Medium	Medium	Low
Improbable (E)	Medium	Medium	Medium	Low
Eliminated (F)	Eliminated			



[115] Perrow, C., *Normal Accidents: Living with High-Risk Technologies*, New York: Basic Books, 1984.

[116] Wickert, D.P., "Flight Path Angle and Energy Height Planning for Negative-Ps Test Points," *Proceedings to the SETP 49th Annual Symposium*, Sep 2005.

[117] George, B., "Dive Planning and Energy Management Techniques for Negative Ps Test Points," *Proceedings to the 37th SFTE Annual Symposium*, Reno, 2006.

[118] The DC-10 had two relevant design flaws: 1) the cargo door could be latched but unlocked due to poorly-designed "fail-safe" locking pins; 2) the cabin floor lacked vents between the passenger and cargo holds to equalize pressure differential in the event of cargo hold de-pressurization. Both flaws were ultimately fixed by McDonnel Douglas.

[119] Chittum, Samme, "A Tale of Two DC-10s," *Air & Space Magazine,* Dec 2017, retrieved online 19 Aug 2018, https://www.airspacemag.com/flight-today/book-excerpt-flight-981-disaster-180967121/.

[120] Aircraft Accident Report: American Airlines, Inc. McDonnell Douglas DC-10-10, N103AA. Near Windsor, Ontario, Canada. 12 June 1972," National Transportation Safety Board, 28 February 1973.

[121] Surowiecki, James, *The Wisdom of Crowds,* New York: Anchor Books, 2005.

[122] Tetlock, Philip, and Dan Gardner, *Superforecasting: The Art and Science of Prediction*, Crown Publishing, 2015.

[123] United States Air Force Aircraft Accident Investigation Report, AC-130J, T/N 09-5710, Eglin AFB, FL, 21 April 2015.

[124] Auttapone, K., Wilson, D. & Dunn, R. "A Review of the Evolution of Shared (Street) Space Concepts in Urban Environments," *Transport Reviews*, 34:2, 2014.

[125] Hamilton-Baillie B., "Towards Shared Space," *Urban Design International,* 13(2): 130-138, 2008.

[126] Long, R.D., *Real risk: human discerning and risk,* Scotoma Press, 2014.

[127] This statement does not preclude a Bayesian approach to updating expectations and building confidence in the models as the envelope is expanded. The point is merely that the *a priori* estimate for the confidence in the remaining, untested portion of the envelope should not change. It is still as uncertain as before testing began.

[128] Feynman, R., *Report of the Presidential Commission on the Space Shuttle Challenger Accident, Appendix F: Personal Observations on Reliability of Shuttle*.

[129] Though widely reported as a factor in the launch decision, the Rogers Commission found no evidence or plans that a live telecast with Christa McAuliffe was planned as part of the State of the Union address. Three live telecasts with the teacher in space were planned. (Rogers Commission Report, p 177). Though the Rogers Commission found no evidence that managers felt outside pressure from the White House, NASA shuttle managers had long emphasized launch schedule in an effort to meet the expectation and claim that the shuttle was "operational."

[130] Rogers Commission report, p 148

[131] Ibid. p 147.

[132] Ibid. p 200

[133] Feynman, R., *Report of the Presidential Commission on the Space Shuttle Challenger Accident, Appendix F: Personal Observations on Reliability of Shuttle*.

[134] Rogers Commission report, p 165.