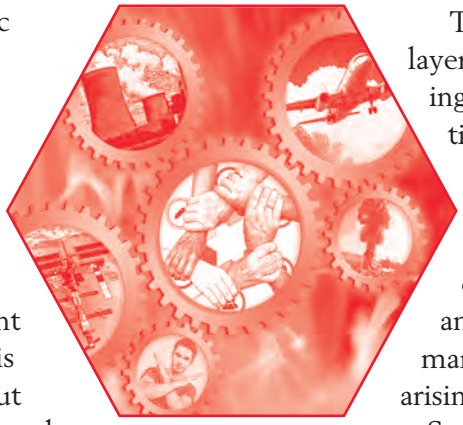
Chasing the Black Swan

by Malcolm Jones
Reading., U.K.

The term "Black Swan" is a familiar concept in the context of high-consequence operations. There is the continual concern that there may be an "as yet" undiscovered flaw or lack of understanding in the design of a product, process or facility that could lead to a catastrophic event. The potential incompleteness in understanding any design concept, implementation and associated assessment is of concern. Given that "absolute confidence" may never be possible, the question becomes how best to continue to search for such possible flaws with a view to subsequent removal or mitigation. At first sight, this appears to be a process without end, but the level of commitment must be balanced against any detrimental consequence that could ensue should a Black Swan exist. But when is "enough is enough?" In this paper, this subject is covered in the context of nuclear warheads, where the Black Swan could indeed be catastrophic should it exist. The paper is framed around what can be learned from the general literature associated with "Black Swan" thinking.



The contents of this document represent the views of the author and not necessarily those of the Atomic Weapons Establishment (AWE).

Introduction

Aspects of this subject have been covered in previous papers by this author [Refs. 1 & 2] which have taken different perspectives of an independent strength in depth approach to safety and its assessment. The first paper concentrated on organizational structures in relation to necessary organizational levels, each having independent responsibilities for safety ensurance and assurance, coupled with final decision-making responsibilities and, of course, each having the appropriate level of technical capability and experience. The overall organizational responsibility is that of ensuring a safe product, process or facility with appropriate technical and evidential support, with a level of scrutiny proportional to the potential consequence of getting it wrong. The first organizational layer is responsible for ensurance, making the safety case with a supporting evidence base. The second independent structure scrutinizes and challenges this case for appropriate depth and completeness, including assuring that the appropriate level of expertise has been applied and that the evidence and analysis offered is complete and not flawed.

The third independent organizational layer is responsible for final decision making, having taken into account and scrutinizing the ensurance and assurance evidence provided by the first two layers and, in turn, adding its own independent assessment based on a complementary fund of knowledge and experience. This third layer also manages and resolves any disagreement arising from the views of the first two layers. Some fundamental competencies needed for correct operation of this structure were identified in that paper, as were some of the difficulties of this structure.

However, this overall approach alone may not be sufficient for ensuring the absence of a Black Swan in a product such as a nuclear warhead. For this reason, a fourth layer is introduced (see Figure 1). This somewhat undefined layer continues to probe into the product, process or facility with the intention of taking a "what if?" and "ex-

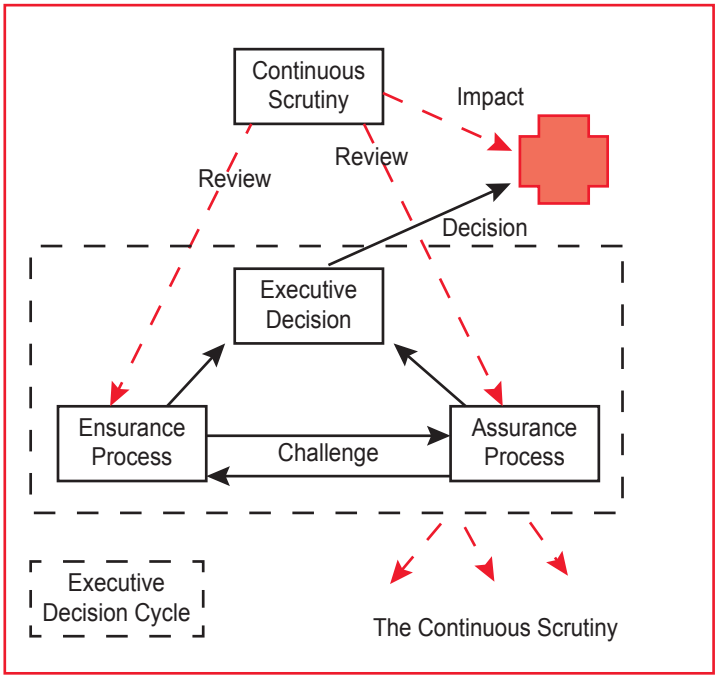


Figure 1 —The Continuous Assessment Process.

pecting the unexpected" mindset. In essence, this layer's job is never finished and it sets its target in the broadest sense. More about this is said later.

The second paper covered two independent technical strategies for safety ensurance and assessment. The first was based on an independent strength in depth technical analogue of the organizational structure. This is based on a deterministic approach, founded on fundamental and independent levels of technical defense, and is intended to include a hedge against uncertainty. The second strategy complements this with a realistic, but conservative, evidence-based numerical risk assessment looking for compliance with risk standards. Of course, any risk assessment will contain an element of foundational judgment supporting such an analysis. In principle, this judgment could be in error and for this reason, an acceptable level of conservatism must be included to address any remnant uncertainty. Of course, this risk assessment relates directly to a particular (or set of) design solution, as opposed to the deterministic approach based simply on "how to design for safety." The latter risk approach aims more for a demonstration of meeting acceptable risk standards, given the design solution. The two approaches have elements of fundamental difference but, in concert, compensate for individual potential limitations and statutory requirements. From a philosophical point of view, the deterministic approach does offer the better of the two approaches in protecting against the potential of a Black Swan because of its greater potential for protecting against any limits in our "complete knowledge." The risk assessment is necessary because society demands to understand "what the level of risk is" and whether it is acceptable. However, the risk assessment inevitably bases its approach on an "accepted level" of understanding, with some conservatism included to cover any potential deficiency in knowledge.

There is some value in the argument that, given a history of a product or process performing satisfactorily over a given period, this adds to the confidence that a Black Swan may well not exist. However, the value of such evidence must be set against the level of potential detrimental consequence. For example, if the requirement is that a major consequence should not have a frequency of greater than 1 in 100 years, then a 10-year successful history only gives additional confidence that all seems well. In fact, if the frequency requirement for a catastrophic event is far more demanding, then typical design history perspectives may be somewhat less comforting. History also shows us that past success does not always support future success.

Historical Experience

History shows us that many, if not most, catastrophic events occur though an "unfortunate" sequence of linked events rather than from a single technical cause. Such previously undiscovered sequences are often related to incomplete knowledge of the range of possible states leading to the propagation of the failure sequence or perhaps, more correctly, the level of dependency between the states in such a sequence. For example, an initial technical failure will not propagate if all states potentially influenced by such a failure have been fully identified, characterized and configured through independence to terminate such a sequence. The problem arises when there is incomplete knowledge about the range of states that may be mutually or dependently influenced. So, incomplete information is the enemy of safety assurance of termination of a mishap occurrence. This is often pictorially displayed in the "Swiss Cheese" model.

In general, the states involved may include a wide range of types. For example, in a simple technical product, an initial analysis may appear relatively simple, i.e., that of tracking the well-known and fully defined influenced states together with their known mitigating/independent resilience. However, the response of these states may in turn be influenced by external actors; for example, in certain environments at the time of the initial flaw and without this information the response robustness of the argument for terminating the sequence prior to the mishap will be questionable. Of course, nuclear warheads need to remain safe, given a wide range of potentially severe environments. To treat any real case, it is necessary to establish, as much as possible, a complete definition of the "system" of all states and influences. This is perhaps analogous to the goal in fundamental physics of seeking a complete understanding of all of nature's subatomic particles and forces, and their potential interactions. As the "system and influences grow," so does the difficulty of being able to fully identify and characterize the response of such states given any initiating event. In essence, complexity can become the enemy of safety and enhances the difficulty of identifying a Black Swan, should it indeed exist.

This complexity is perhaps best illustrated by a fictitious, but not improbable, example given in the form of a somewhat dated film, *Fate is the Hunter (1964)*, which was shown during the 2017 International System Safety Conference (ISSC) by John Rankin who used this film as an example. It is paraphrased as follows: Following an aircraft take-off, one of the engines was struck by a bird that was rarely found in that part of the world — an unlikely initiator. The plane suffered a jolt due to the loss of that

engine and, as a result, a cup of coffee just supplied to the Captain spilled onto an equipment box. Because the box was unsealed, coffee leaked onto the underlying equipment. The coffee was somewhat conducting in nature and, as a result, this interfered with underlying critical circuits, leading to the loss of the other engine. Because of aircraft congestion at the airfield at that time, the Captain took the reasonable option of landing on a nearby suitable beach. It so happened, unknown to the captain, that a pier on that beach, which had been scheduled for removal the previous week, was still in place because of a decision by the contractor that there was no real need to hurry. Hence, the fateful collision.

Although the initial verdict was judged to be human failure, careful analysis showed this not to be the case. The sequence is fictitious, but it does exemplify the need to fully understand what the "system and the influences" really consists of, along with all the potential interacting events given a somewhat unexpected fault initiator. Who would initially have identified the potential presence of a cup of coffee, the importance of an unsealed equipment box, congested airport conditions at the time and the continued existence of a pier that should have been absent? Simplicity in the "bounding of the system" is an important goal for minimizing the opportunity for failing to fully characterize it.

This, in fact, represents an important element in the safety approach to nuclear warheads: bounding the system (in the safety sense) as far as possible and aiming for simplicity, independence and clarity in the safety argument. Some of these aspects are covered in more detail later in this paper. The additional lesson learned from this fictitious example is that the sequence initiator was an abnormal environment, i.e., is an "insult" to the technical system rather than a design flaw, although one could argue that design flaws helped the sequence to propagate. In fact, history shows that many sequences that have led to mishaps come from initial "insult" initiators. This lesson is not lost on the strategy for minimizing the opportunity of a Black Swan to manifest in nuclear warheads. The strategy is based on the effort expended in preventing such insults and in the safety resilience given such an insult.

Human failure is often cited as the main cause leading to catastrophic failure and, for this reason, design safety has to be resilient to human failure, in addition to minimizing the potential for human failures.

General Background on the Term "Black Swan"

Origin of the Phrase "Black Swan"

A rare event, based on the belief widely held in England in the 1600s that swans could only be white. All swans in England were white. The phrase "a Black Swan" was a metaphor for "that which could not exist." As a side note, black swans were discovered in the late 1600s in Australia.

Current Understanding of the Phrase

An event judged through best established inductive/inferred logic as rare — but this process does not remove the possibility of occurrence. Black Swans may escape notice because:

a. The knowledge that we base our inductive assessment on is not quantitatively or logically correct (its impact on risk assessment).
b. Or, more important, the scope of our inductive process is not complete, i.e., there are some possibilities we have not yet visualized (its impact on both risk assessment and defense in depth).

The bottom line is, given the application of our best knowledge, a quantitative assessment may be broadly realistic, but does not discount the possibility of an occurrence. In most cases, this quantitative assessment is sufficient; in fact, the whole subject areas of "cost benefit and insurance risks" work on this principle. However, this approach may be inappropriate if the Black Swan event has a major or catastrophic consequence. In this case, there are two emerging issues:

a. The incompleteness in the quantification of the probability of mishap occurrence may be small but this delta matters.
b. The true nature and impact of the mishap may itself not be understood and may be underestimated. Although there may be a reasonable description of the mishap, the real consequences are usually less well understood and often turn out to be of a far greater nature than first anticipated.

Hence, the context of the Black Swan is that it applies to an unwanted mishap that is assessed to be somewhat improbable, but cannot be discounted. Additionally, its outcome can be severe or catastrophic, and not fully understood. In fact, in a general sense, it might be unclear as to what the mishap itself may be. This is typified by lessons from economic/financial traumas.

Some Terms and Definitions

Additional important terms and definitions include:

- **Unexpected —** The "best analysis" suggests that the event has a low probability of occurrence. Such

an analysis cannot claim to be absolutely free of incompleteness and, as such, cannot assure us that the event can't happen. There is often a quantitative definition of what is meant by "unexpected."

- **Inconceivable —** There is some fundamental reason why an event cannot happen or why one cannot imagine how the event can possibly happen. This in principle (particularly the latter) is a human construct and does not completely remove the possibility of a high-consequence event.
- **Consequence —** The range of impacts (detrimental in this case) given the event's occurance. In the case of safety, consequences will take many forms: death or harm, environmental damage, loss of asset, financial loss, reputational loss, political harm, etc.

Black Swans are generically spoken of in the context of major detriments and for nuclear warheads (NW) can be associated with a major issue in relation to performance or safety. In the latter case, this can be related directly to the occurrence of the worst-case catastrophic event itself, or to the realization that there may be a major flaw in the safety argument that requires urgent mitigating action. Of course, this is a key issue in the continued ownership and safety scrutiny of a stockpile of nuclear warheads.

Some Differing Categories of Black Swans and AWE Culture

Known, But Not to the Degree Necessary

Knowledge about such a possibility is available but is not sufficient for correct judgment on probability and/or consequence. Of course, there is always the judgmental problem of when "enough is enough" in relation to the depth of scrutiny necessary for elimination of the Black Swan's possibility.

With regard to detrimental event occurrence: The warhead program strives to avoid falling into this category, taking a "what if?" and "expecting the unexpected" approach. All known possibilities are subject to in-depth scrutiny. If there is any reason why an issue has not been completely closed, it is subject to continued scrutiny and, of course, this may give rise to a necessary change in design. We continue to strive to establish a more comprehensive understanding of the underlying characteristics of known issues, whether they be of a technical, human factor, implementation or governance nature. We retain a culture of assuming that we may not have reached the end of the road towards this goal and that there is always more to be done — a culture of continued in-depth assessment.

Unknowns, But Knowable Unknowns that Could Have Been Looked for or Known by Others

The problem here is that such unknowns may have been regarded (erroneously) as being of no consequence, e.g., for safety.

With regard to detrimental event occurrence: The relevance to warhead activity follows from the comments under the last heading in the sense that there should be no assumption, without sufficient evidence, that the knowable unknowns can be discounted. If the unknowns are knowable, then this gives us a direction in which to focus further effort and scrutiny with the goal of reaching a position where these unknowns are converted to "fully characterized" and their impact is determined. Included in this is the need to be aware of what's happening elsewhere in the world where similar technologies may be exercised and where there are technology enhancements taking place that can be advantageous in enhancing our safety. Of course, in this respect, one covers the associated subject areas included in STEM (Science, Technology, Engineering and Mathematics) where it is necessary to keep a close eye on worldwide developments in both industry and academia that can be incorporated to our advantage. Our close collaboration with our U.S. colleagues plays a major mutual role in this.

In addition, one needs to keep abreast of state-of-the-art developments in the range of methodologies for safety assessment and assurance, many of which are bound up in the general subject area of system safety and, as such, where we need to be expert practitioners. On course, this again is a constantly moving and evolving activity with an acknowledgement that there will be no perfection. Any "judgment" that we have reached a position of having identified and characterized such known unknowns is subject to constant challenge. Again the "what if?" attitude and "expecting the unexpected" culture at AWE provides a strong defense against this form of failure.

Of course, the primary element of protection relating to "known by others" comes through collaboration with our U.S. colleagues and other work, particularly in the nuclear industry at large. The aim here is to establish confidence that no potential influencing subject area is left untouched.

Unknown, but Unknowable Unknowns

No absolute basis for being able to make any assured prediction for occurrence expectation or consequence — *a true Black Swan!* Nevertheless, even here, "confident" predictions may well be made but, of course, with the potential for failure due to ignorance.

With regard to detrimental event occurrence: This presents the most difficult aspect with regard to warhead safety assurance in that, if such flaws exist in our knowledge, we have little or no guidance on where to look further. The best practical protection against this uncertainty comes from the independent strength in depth approach to design, ownership and assessment both in terms of the technical and organizational aspects. The best overall solution we are able to come up with in this respect, given no clear guidance of where to look further, is a continued commitment to ensuring that we have a cadre of high-caliber staff working at the cutting edge of all relevant technologies, processes and assessment methodologies — and that our staff has sufficient freedom and encouragement to probe into for the unexpected. This activity is not solely restricted to in-house activities but also includes keeping a close eye on developments of interest in the external world, along with how best to take advantage of such developments. Unknowns are not likely to be uncovered through standard process and routine. From a business perspective, this may appear to be an untargeted, unproductive and non-cost-effective overhead — but we fail here at our peril. The continuous scrutiny process is identified in Figure 1. AWE's culture strongly supports and encourages this approach.

Lessons from the Open Literature

Nassim Nicholas Taleb

Taleb is well known for his work in this area and describes a Black Swan as:

- Being unpredictable and, as such, unable to be truly sentenced as low probability
- Causing an extreme or catastrophic impact
- Retrospectively predictable, with the warning that there can be a tendency to simplify (possibly erroneously) the post-event analysis of the cause. There may be a telling lesson here for us in relation to the Review Learn and Improve (RLIs) we undertake — we need to be clear that we really do have the correct post-event analysis; otherwise, it can happen again.

Taleb's [Ref. 3] interest in unpredictability comes from his association with economics and the financial industries, where prediction has been shown on many occasions to be unwarranted, with major detrimental consequences. There may be some success in predicting the near future, but this gives no assured guidance for the longer term — and the longer term may not be too extensive in this context. This is somewhat like weather prediction, where longer-term prediction is fraught with uncertainty. This can manifest via so-called chaos theory. In both cases, predicting the future is bedeviled by the vast number of interacting contributors of not always perfectly defined knowledge and starting conditions. In the case of economic and financial industries, the situation is further complicated by the emotional response of people, whereas weather prediction is primarily governed by the laws of physics. *A nuclear weapon (or even a warhead) design, including its processing and ownership, is itself a somewhat-complex subject, with inherent complicated relationships. However, a key lesson from Taleb's experience is that we should strive to limit complexity (as it relates to safety) because of its impact on our ability to predict with confidence. As a result, we continue to strive to limit complexity and to gain as complete an understanding as possible of what we have, with the mindset of "what if" — but is our understanding complete?*

There have been many attempts to model economic and financial futures, some based on a statistical basis, for example, on the normal distribution for occurrence probability and range of consequence. In the context of financial industries, the "well-known" distribution peaks can be associated with the "currently understood" in the relatively short term while the "tails," with all their uncertainties, can be represented by longer-term "judgments." Because of the complex nature of the financial industries these "tails" often turn out to be completely misleading. Such distributed approaches are sometimes applied to warhead safety assessments. These distributions usually have far more evidential support near the peaks, with less in the tails where low-probability high-consequence predictions are made and where prediction can be in significant error. This is where the Black Swans, if they exist, may lie. The safety of warheads depends on the evidential confidence in the application of such distributions where such distributions need to be applied.

Taleb's View on Combatting Uncertainty

Taleb's view was that, because of the complexity associated with the financial industry, effort expended in trying to prevent catastrophic occurrence was not well spent. He said that catastrophes are inevitable and most effort is best directed toward robustness in coping with Black Swan events when they occur, rather than on trying to predict or prevent their occurrence. His experience was associated with economic and financial subject areas, where his assumption (based on substantial historical evidence) was that these events *will* occur and one should have plans in place for how to deal with them and mitigate the consequences. Given this

assumption that an event will occur, Taleb's approach appears correct. *However, the priority for nuclear warheads lies firmly in the direction of preventing the occurrence, and particularly in relation to the "worst case" occurrence of Inadvertent Nuclear Yield (INY), rather than prioritizing mitigation actions following such an event (if indeed this was realistically possible). For this reason, we concentrate on limiting complexity, particularly of principal safety arguments, to enhance transparency and confidence in the evidence supporting an assessment of a very low occurrence probability.*

Charles Perrow and Accident Theory

Perrow is well known for his work in accident theory (e.g., Ref. 4). He provides additional guidance to how Black Swans can arise and what approaches should be put in place to avoid their occurrence. He has spent much of his career in the field of historical accident assessment and the reasoning behind their occurrence. These assessments have covered a wide range of subjects, but their applicability to nuclear warhead safety is generally pertinent. The principles that clearly emerge from his work include simplicity, clarity and independence. These are evidenced by his observations and recommendations noted here:

> "Taleb's view was that, because of the complexity associated with the financial industry, effort expended in trying to prevent catastrophic occurrence was not well spent. He said that catastrophes are inevitable and most effort is best directed toward robustness in coping with Black Swan events when they occur, rather than on trying to predict or prevent their occurrence."

A. Complex, interconnected and highly coupled systems should be expected to fail.

Response in Warhead Design — Warheads, by their very nature, must be interconnected, but this is different from being highly coupled in the context of safety. For example, interconnectivity in the sense of arming safety requires a unique set of authorizing actions/conditions for "fault progression," and their absence prevents unintended fault progression. This principle also applies to other aspects of the warhead. In the safety sense, there is limited inadvertent connectivity. The same is true for the AWE safety organizational structure in that, although there are interconnections, they are designed to be independent in maintaining safety ensurance and assurance.

Design also aims to minimize complexity, and this helps towards unintended (undetected) safety interconnectivity. For example, safety systems strive to minimize component counts (particularly those that are safety related). In addition, principal safety arguments are based on a demanding requirement to demonstrate mutual independence between safety systems, including the fundamentals of principles and implementation. Coupled with this is a strategy aimed at making the principal safety argument transparent and simple, as opposed to complex in nature. The more complex the safety argument, the more challenges that can be raised and the greater the difficulty in providing the appropriate level of assurance.

Both simplicity and independence minimize the opportunity for failure through unintended coupled and undetected paths. In turn, simplicity and independence ease the burden of providing safety assurance, given that one is always tasked to continue looking for all the possibilities leading to mishap. It is certainly not a stance of taking more and more comfort from what we have achieved so far. A strong continuous probing, analysis, testing and surveillance culture is key to early detection of any flaws that might exist or may develop over time, and is supported by a commitment to maintaining a strong cadre of experts probing in the relevant technology areas.

B. Focus on those aspects with catastrophic consequences

Response in Warhead Design — Although we focus on safety across the board, we nevertheless concentrate on the worst-case events of Inadvertent Nuclear Yield (INY) and major radioactive material (RAM) release. We go to great lengths to prevent their occurrence by way of design, associated processes and independent organizational assessment activity. Again, the prevention of INY and RAM release is based on a number of foundational safety principles that are, in turn, scrutinized with respect to the probity of their underlying scientific, technical, engineering, implementation and evidential basis. This includes assurance that their implementation fully meets these principles for both normal and abnormal (accident) environments. Of course, the ownership approach also aims to minimize the occurrence of abnormal environments and the impact of human error.

C. Strive for systems that are simple, easily understood, disconnected and decoupled

Response in Warhead Design — Again, we adopt this strategy for NW design safety, focusing on clarity, simplicity and strong independent arguments in the defense in-depth approach. Safety is based on a limited number of strong and independent principles, arguments and technical implementations. These, together with the clarity of their safety intent and quality of implementation, enable a clear and comprehensive challenging process to be undertaken in order to fully test and analyze for overall safety completeness and compliance. The principal arguments supporting the safety case should be simple and clear, and should not arise as a result of a complex set of contributing arguments.

The strategy is based on the application of *a limited number* of clear, strong and independent safety arguments and implemented safety systems based upon ensuring maximum resilience against the undermining of safety that can occur through added complexity as the number of safety systems increases. Increasing the number inevitably leads to increasing complexity and increasing concerns about maintaining true independence between the implemented safety systems. Added complexity generally increases the difficulty of ensuring confidence of overall safety because of the multiplicity of potential sneak paths or fault sequences that can give rise to failure. In addition, a warhead is confronted by constraints on available volume and mass, as well as the complementary requirement that these safety systems can be reliably removed when fully "authorized" to do so. This authorization is based on a principle of uniqueness that effectively decouples it from the safety argument.

Some Additional Factors

A Lesson from History — Predicting the Future Based on the Past

"But in all my experience, I have never been in any accident … of any sort worth speaking about. I have seen but one vessel in distress in all my years at sea. I never saw a wreck and have never been wrecked nor was I ever in any predicament that threatened to end in disaster of any sort." — Captain E.J. Smith of *RMS Titanic* in 1907, quoted five years before he went down with his ship.

The clear lesson here is that history (or perceived history), although it may be a guide to the future, nevertheless is *only* a guide and, as such, gives no assurance about the future. Confidence — that is, false confidence — was based on the safety built into *Titanic*, where all potential threats were deemed to have been considered and catered for. The ship was judged "unsinkable," but, in fact, suffered a catastrophe on its maiden voyage. This resulted from a combination of failure to prevent the collision and overestimation of the ship's capability to withstand such a collision. *This was an example of a "known but not to the degree necessary" Black Swan.* Similar arguments can be applied to the *Columbia* and *Challenger* space shuttle tragedies where confidence was based on "past successes and inadequate technical assessment of known threats."

Warnings from the past come in the form of failures and "near misses" and, if acted upon, can improve the situation. However, near misses are not always recognized for what they are and, therefore, there is no guarantee they are fully understood with regard to the future. AWE scrutinizes issues of this nature to ensure that the fundamental reasons for their cause are fully understood and removed. We *may* take some comfort from the past history of the U.K.'s nuclear weapons ownership, where there have been no major safety events, but one cannot take this as assurance of the future. Certainly, the number of successful weapon lifetimes in our history comes nowhere near to proving compliance with the exacting safety standards with which we strive to comply. We certainly take heed of any issues that arise and act on them with high priority.

Human Factors and Independence

Humans are not exempt from failure. Assumptions of true independence in human actions, which is often used as a major redundancy element in preventing a failure, have a long history of being undermined — note the incident at Minot Air Force Base in the U.S., where six Advanced Cruise Missiles (ACMs) were misplaced for 36 hours. This type of failure has been noted in dual "independent" checking processes, where enhanced confidence is based on an assumption that true independence is assured in the process. Humans are seldom truly independent in this context. For this reason, the safety of nuclear warheads is based mainly on scientific, technical and engineering principle foundations. Nevertheless, "independent" checking and dual control, together with human error assessment, do have an important place in nuclear warhead safety in preventing and quantifying inadvertent actions. Of course, even safeguards based on technical principles are not absolutely free from human influence. For example, engineering-based safeguards need human actions for design manufacture, inspection and maintenance for ensuring and maintaining design intent compliance.

Conservatism, Worst Case and Probability Function

Overall, the approach to nuclear warhead safety design and assessment is conservatively based. This conservatism is applied to combat any possible remnant uncertainty in occurrence probability and consequence, and usually takes the form of a "worst case" rather than "best estimate" approach to risk assessment. This less onerous (in application) approach is well understood and provides a sensible hedge against uncertainty — but it will not always be usable. In some cases, worst-case assumptions can be too conservative and lead to difficulty in meeting exacting risk safety standards. Therefore, somewhat less conservative but still supportable statistical approaches are deemed more appropriate (Figure 2). These tend toward a more onerous approach (in application) and can carry with them a smaller conservative margin in protecting against uncertainty. As noted previously, distributions often have greater evidential support near their peaks and far less in their tails, which typically characterizes the low probability associated with potentially catastrophic events — Black Swans. The safety assessment of nuclear warheads sometimes puts us in these regions where we need to apply such tails of distributions with suitable caution and with a sound evidential base founded on substantial supporting technical and historical arguments. Such distributions are accepted only if detailed scrutiny provides the evidence noted previously and if, in turn, this evidence stands up to independent challenge. This typically applies to assessments of explosive response to environmental challenge.

The Importance of and Achieving Independence

The occurrence of mishaps — and even Black Swans — is often the result of an initiator setting off a related sequence that is not truncated. Protection against this occurrence relies heavily on:

(a) A full understanding of the inventory of possible hazardous sequences leading from any given initiator to culmination in a mishap

(b) Ensuring a suitable level of independence between such events in the sequence so that the sequence is truncated before reaching the mishap stage. Some of these aspects were covered in Reference 2 in terms of the identification of the strength in depth of bar-

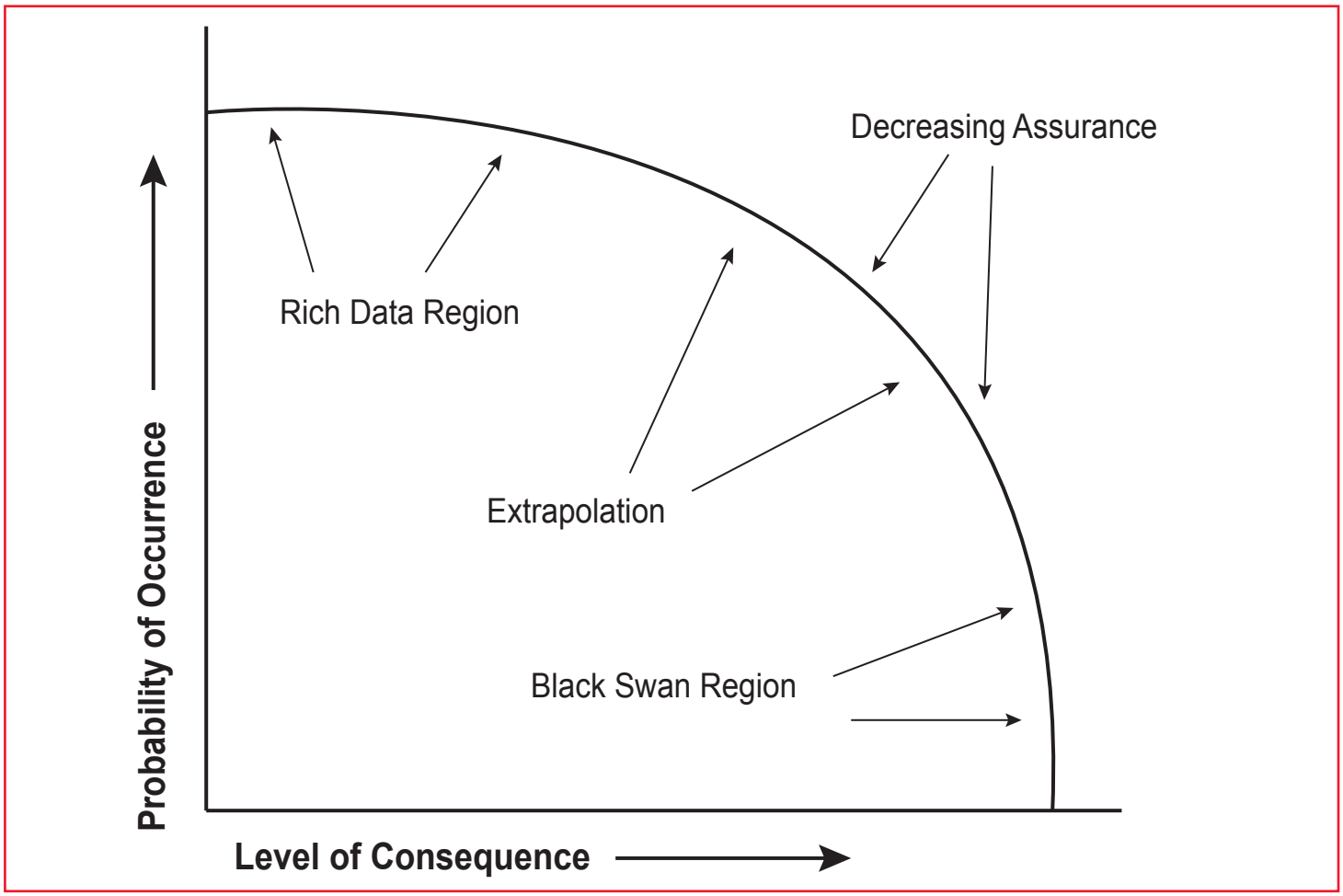


Figure 2 —Distribution Approach to Safety Assessment.

riers or Lines of Defense (LOD) in relation to their number and strength in preventing such a mishap — in particular, the importance of independence between such LOD. The choice of LOD is based on the widest range of categories possible to meet the independence requirement. For example:

- Fundamental physics or chemical properties based on the laws of nature which directly support the LOD resilience
- Uniqueness of response, in the sense that defeat of the LOD can occur only as a result of unique circumstances
- Decoupling of microelectronics/firmware/software elements from the primary safety arguments
- "Best practices engineered" implementation of a concept to the extent that failure is unlikely even under challenging circumstances — e.g., accident conditions
- Procedure based, relying on humans to do the correct thing at the correct time (and not doing the incorrect things at the wrong time). Examples are fully trained staff and application of the two-man checking principle.

These broad categories include the resilience properties of *isolation*, *incompatibility*, *inoperability* and *independence*. Of course, although correct procedure can have a significant influence in safety, it takes the lowest place in the hierarchy of attributes in the "strength in depth" argument.

The ability to make an LOD resilience case deterministically and to support the principle of LOD independence follows this hierarchy. A structured sequence protection argument based strongly on correct LOD-based procedures alone will not be acceptable. Although human LOD are least favorable in the overall choice, all LOD have some element of human involvement because of human relationship to judgment, inspection, testing, maintenance, etc.

In addition to the choice of LOD category, the overall strength in depth arguments for LOD in any particular fault sequence are enhanced by applying the following hierarchy of application for independence in as far as technical implementation allows:

- Fundamentally different concepts
- Fundamentally different applications of the same concept
- Different engineering implementations of the same concept
- Application of different — and differently sourced — materials in the same concepts or engineering implementation

All of these LOD attributes form the major basis of preventing sequences from propagating to mishap conditions.

The Influence of Environmental Factors on Black Swan Assurance

An initiator can take the form of a design flaw (failure) or an "insult." The latter represents an abnormal environment (accident) and this environment could, in turn, be applied simultaneously to all LOD in the "strength in depth" strategy. Therefore, it is important to avoid common mode failure in the application of the LOD. General historical evidence shows that many major mishaps have indeed arisen from common cause environmental-based failures of this manner. The potential for abnormal environments presents the greatest challenge in ensuring that all sequence paths are identified for any warhead design. In turn, one must show that there is sufficient resilience to ensure that such sequences, through design and testing, are truncated under these circumstances and that there is no common mode failure. Prevention of abnormal environment occurrences is, of course, also a major objective in the overall safety theme, and every effort possible is made to prevent/limit the potential for such insults/accidents and, in turn, to prevent such environmental occurrences from propagating to the warhead. Nevertheless, such events cannot be totally discounted and the safety characteristics of a warhead design must show appropriate resilience against such threats.

Conclusions

Hunting for the Black Swan in the context of safety can represent a somewhat uncertain activity as one may be looking for something that does not exist. One cannot prove a negative. However, there is a compelling history of disasters that evidence the existence of such Black Swans that were not identified in advance. The level of scrutiny applied to the search for Black Swans must be related to the potential consequence of failure. Of course, for nuclear warheads, we are well aware of the worst-case catastrophe represented by inadvertent nuclear yield (INY) and, as a result, have established a sound culture aimed at minimizing the

potential for such occurrences. In addition, we must provide a compelling level of evidence/assurance that the probability of such a (man-made) occurrence will be less than natural catastrophes of a similar magnitude. We do this based on an independent "strength in depth" approach coupled with state-of-the-art application of science, engineering and technology with all the associated requirements set by established best practice. Such designs are then subjected to an independent conservative numerical risk assessment based on a foundation of supporting evidence to demonstrate compliance with demanding national risk standards. Both fundamental design aspects and the risk assessment are, in turn, subject to organizational independent assurance challenges. The latter covers the fidelity and completeness of the logic, evidence and implementation, as well as a search for any aspects that might have been missed or not covered in sufficient depth. Both contributions are then considered at the organizational executive decision-making level, where a further level of knowledge and experience is applied. These activities go a long way toward eradicating the presence of Black Swans but, of course, this does not guarantee that one is not present. The final layer in the defense comes from strong organizational support to a continuing probing culture that applies a "what if?" and "expecting the unexpected" mindset. Therefore, the campaign for assuring safety never ends.

This latter contribution is based on a strong and continuing organizational commitment to engaging the best brains, which continue to probe into the various science, technology and design areas with the goal of achieving even greater depths of understanding in the cause of Black Swans. This is exercised through direct in-house activity and through what can be gleaned from the external world. One cannot claim that this last layer leads to perfection and, at times, it may appear to be somewhat undirected in its approach. However, it is still the best way we know to tackle the demanding challenge of *"Unknown Unknowns."* Thus, the overall strategy is based on:

- A sound approach for developing a safe product and process — ensurance
- A sound independent scrutinizing/challenging element — independent assurance
- A knowledgeable and experienced executive decision-making element — final accountability
- An organizational commitment to applying the best brains to a continuing probing approach into all the technology and procedural areas — the last possible hiding place of the Black Swan — with the mindset that safety is never finished

About the Author

Malcolm Jones has previously led the Distinguished Scientists group at the Atomic Weapons Establishment (AWE). He currently holds the position of Scientific Adviser to AWE's Chief Scientist and directly supports AWE's Chief of Product Assurance. His career at AWE has taken him through a wide range of scientific and engineering topics, but he has maintained a continuous association with nuclear weapon design and process safety and top-level nuclear safety standards. His interests extend to corporate safety cultures and the root-cause reasons for failures. He is a Fellow of the International System Safety Society and is an adviser to a number of senior U.K. Ministry of Defence and AWE safety bodies. He has been awarded an MBE in the Queen's Birthday Honours List for contributions to the U.K. defense industry and is a recipient of the John Challens' Medal, which is AWE's highest award for lifetime contributions to science, engineering and technology. He has also been honored by VNIIA in the Russian Federation for his work in fostering nuclear weapon safety collaboration between the U.K. and the R.F.

References

1. Jones, M. "Organisational Problems – Potential Causes – Unintentional Consequences," *Proceedings of the 34th International System Safety Conference*, Orlando, Florida, August 2016.
2. Jones, M. "Strength in Depth and Quantitative Risk Assessment in the Context of Low Frequency High Consequence Systems," *Proceedings of the 35th International System Safety Conference*, Albuquerque, New Mexico, August 2017.
3. Taleb, N. N. *The Black Swan: The Impact of the Highly Improbable*, Penguin Books, London, 2010.
4. Perrow, C. *The Next Catastrophe Reducing our Vulnerabilities to Natural, Industrial and Terrorist Disasters*, Princeton University Press, 2011.
5. Jones, M. "High Consequence Accident/Failure Theories," *Proceedings of the 19th International System Safety Conference*, Huntsville, Alabama, September 2001.
6. Perrow, C. *Complex Organizations: A Critical Essay*, Echo Point Books & Media, 1972 (reprinted in 2014).