ISSN 2754-1118 (Online)
ISSN 2753-6599 (Print)

# Chasing the Black Swan

**Malcolm Jones**

Atomic Weapons Establishment (AWE), Aldermaston, Berkshire, UK

## Abstract

*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. Concern lies in the potential incompleteness of understanding of any design concept, implementation and associated assessment. Given that 'absolute confidence' may never be possible, the question arises as to how best to continue to search for such a possible flaw with a view to subsequent removal or mitigation. This at first sight appears to be a process without end but the level of commitment must be balanced against the detrimental consequence that could ensue given that a Black Swan might exist. But when is 'enough-enough'? This subject is covered in the context of the ownership of nuclear warheads where the Black Swan can indeed be catastrophic should it exist. The paper is framed somewhat in terms of what can be learned from the general literature associated with Black Swan thinking.*

## 1 Introduction

Aspects of this subject have been covered in previous papers (Jones 2016) (Jones 2017), which have taken different perspectives of an independent strength in depth approach to safety and its assessment. The first paper concentrated on organisational structures in relation to the necessary organisational levels, each having independent responsibilities for safety ensurance and assurance[1], coupled with final decision-making responsibilities and of course each having the appropriate level of technical capability and experience. The overall organisational responsibility is that of ensuring a safe product, process or facility with appropriate technical and evidential support, with the level of scrutiny proportional to the potential consequence of getting it wrong.

The first organisational layer is responsible for ensurance through making the safety case with a supporting evidence base. The second independent structure scrutinises 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 organisational layer is responsible for final decision making, having taken into account and scrutinising the ensurance and assurance evidence provided by the

[1] We use the term "ensurance" to mean the demonstration that a design, etc., has met its requirements; whereas "assurance" is the independent challenge of whether the ensurance is valid.

first two layers and in turn, adding its own independent assessment based on a complementary fund of knowledge and experience. This third layer will also manage and resolve any disagreement arising from the views arising from the first two layers. Some of the fundamental competences needed for correct operation of this structure were identified in Jones (2016) together with some of the remaining difficulties.

*However, our historical experience suggests this overall approach alone may not be sufficient for ensuring the absence of a Black Swan for a product such as a nuclear warhead*. For this reason, a fourth layer is introduced, see Figure 1 (Jones 2016). That is a somewhat undefined layer which continues to probe into the product, process or facility with the intention of taking a 'what if' and 'expecting the unexpected' mind-set. This layer's job in essence is never finished and sets its target in the broadest sense. More is said later.

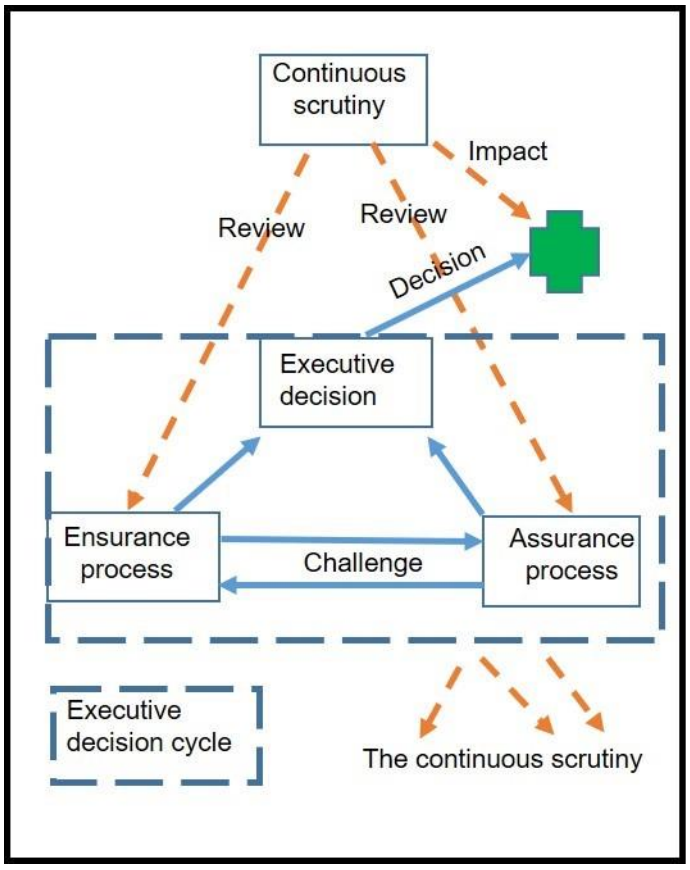


**Figure 1 ~ The Continuous Assessment Process**

Note that the Green Cross in the diagram is the potentially revised executive decision given what the fourth layer has unearthed as a result of continued scientific investigations; "Impact" is the impact on the executive decision given new information unearthed by the fourth layer.

The second paper covered two independent technical strategies for safety ensurance and assessment (Jones 2017). The first was based on an independent strength in depth technical analogue of the organisational structure. This is based on a deterministic approach founded on fundamental and independent levels of technical defence which is intended to include a hedge against uncertainty. The second approach complements this with a realistic, but conservative, evidential based numerical risk assessment looking for compliance with risk standards. Of course, any risk assessment will contain an element of foundational judgment for supporting such an analysis. In principle, this judgement could

be in error and for this reason an acceptable level of conservatism has to be included in order to cater for any remnant uncertainty. This risk assessment relates directly to a particular (or set of) design solution as opposed to the deterministic approach which is 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 for protection 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 of the societal demand to understand 'what the level of risk is' and whether it is tolerable in meeting accepted standards. 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 further 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 a 1 in 100 years, then a ten-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.

## 2 Historical Experience

History shows us that many, if not most, catastrophic events occur through an 'unfortunate' sequence of linked events rather than from a single technical cause. Such previously undiscovered sequences are often related to an 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 such 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, characterised 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 the general case the states involved may take on 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 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 far as possible, a complete definition of the 'system' of all states and influences. This is perhaps analogous to the goal of 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 characterise 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 in simple form by a fictitious, but not improbable, example given in the form of a somewhat dated cinema film, "Fate is the Hunter" (Nelson 1964), shown during the 35th International System Safety Conference (2017) by John Rankin, who used it by way of example. It is paraphrased as follows:

Following an aircraft take-off one of the engines was struck by a bird which 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, which led 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 and hence all of 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 the pier that should have been absent?

Simplicity in the 'bounding of the system' is an important goal for minimising the opportunity for failing to fully characterise 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 which have led to mishaps come from initial 'insult' initiators. This lesson is not lost on the strategy for minimising the opportunity of a Black Swan manifesting itself in nuclear warheads. The strategy is based on both the effort expended in both preventing such insults and the safety resilience given such an insult.

# 3 General Background to the Black Swan Term

## 3.1 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 at that time were white. The phrase "a Black Swan" was a metaphor for "that which could not exist". In the late 1600s, in Australia, black swans were discovered.

### 3.2 Current Understanding of the Phrase

An event judged through best established inductive/inferred logic to be assessed as rare — but this process does not remove the possibility of occurrence.

1. That the knowledge that we base our inductive assessment on is not quantitatively or logically correct. Its impact on risk assessment.
2. Or, more importantly that the scope of our inductive process is not complete, i.e. there are some possibilities we have not yet visualised. *Its impact on both risk assessment and defence 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 and 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:

1. The incompleteness in the quantification of the probability of mishap occurrence may be small but this delta does matter.
2. That the true nature and impact of the mishap may itself not be understood and underestimated. Although there may be a reasonable description of the mishap, the real follow-on 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 a mishap which is assessed to be somewhat improbable, but cannot be discounted, and that 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 and this is typified by the lessons from economic/financial traumas.

### 3.3 Some Terms and Definitions

**Unexpected** — The 'best analysis', suggests that the event has a low probability of occurrence. Such analysis cannot claim to be absolutely free of incompleteness and as such cannot assure us it cannot happen. There is often a quantitative definition of what is meant by unexpected.

**Inconceivable** — One cannot imagine how it can possibly happen, or believes that there is some fundamental reason why it cannot happen. This in principle is a human construct, and does not completely remove the possibility of a high consequence event.

**Consequence** — This relates to the range of impacts (detrimental in this case) given the event and in the case of safety this will take many forms:

- Death or harm
- Environmental damage
- Loss of asset
- Financial loss
- Reputational loss
- Political harm
- *Et cetera*.

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

# 4 Some Differing Categories of Black Swan and AWE's Culture

## 4.1 Known — But Not To The Degree Necessary

Knowledge about such a possibility is available but is not sufficient for correct judgement 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 which is necessary for elimination of the Black Swan's possibility.

**With regards to detrimental event occurrence:** The warhead programme has strived to avoid falling into this category. It has taken a very much '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 out, it is subject to continued scrutiny and of course this may give rise to a necessary change of design. We continue to strive to establish a more comprehensive understanding of the underlying characteristic of the known issues, whether they be of technical, human factors, 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 there is always more to be done — a culture of continued in depth assessment[2].

## 4.2 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 regards 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 such unknowns are converted into 'fully characterized', and their impact 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 which can have an advantage in enhancing our safety. Of course, in this respect one covers the whole associated subject area 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, which can be incorporated to our advantage. Our close collaboration with our US colleagues plays a major mutual role in this.

[2] Note that the fourth layer is not specifically responsible for assessing safety, but has a more open remit for probing further and further into the scientific and technical aspects that underpin nuclear science and, if something is yet to be discovered that could be a 'Black Swan', it is here that it will be found.

In addition, one needs to keep abreast of 'state of the art' developments in the range of methods for safety assessment and assurance, much of which are bound up in the general subject area of System Safety and as such where we need to be expert practitioners. Of course this again is a constantly moving and evolving activity with an acknowledgement that there will be no perfection. Any 'judgement' 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 does provide a strong defence against this form of failure. Of course, the primary element of protection relating to 'known by others' comes through collaboration with our US 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.

### 4.3 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 regards 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 of 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 organisational aspects[3]. The best overall solution we are able to come up with in this respect, given no clear guidance of where further to look, is a continued commitment to ensure that we have a cadre of high calibre staff working at the cutting edge of all of the relevant technologies, processes and assessment methodologies and giving them sufficient freedom and encouragement 'to probe' into 'looking for the unexpected'. This activity is not solely restricted to 'in house' activity but also includes keeping a close look at developments of interest in the external world and how best to take advantage of such developments. Such 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. This Continuous Scrutiny process is identified in Figure 1. *AWE's culture strongly supports and encourages this approach.*

## 5 5 Lessons from the Open Literature

### 5.1 Nassim Nicholas Taleb

Taleb is well known for his work in this area and describes a Black Swan (Taleb 2008) in terms of:

1. It is unpredictable and as such cannot be truly sentenced as low probability;
2. It causes an extreme or catastrophic impact; and
3. It is 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*

[3] Beware of 'groupthink' in which people strive for consensus; the claim of independence is always subject to credible challenge

> *actions we undertake — we need to be clear that we really do have the correct post event analysis — otherwise it can happen again.*

Taleb's interest in unpredictability comes from his association with economics and the financial industries where the activity of 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 aspect is somewhat like weather prediction where longer term prediction is also fraught with uncertainty. This can manifest itself via so-called Chaos Theory. In both cases predicting the future is bedevilled by the vast number of interacting contributors of, not always perfectly defined, knowledge and starting conditions. In the case of the 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. *It is salutary to note that a nuclear weapon (or even a warhead) design, including its processing and ownership, is itself a somewhat complex subject area with inherent complicated relationships. However, a key lesson from Taleb's experience is that we should strive to limit complexity*[4] *(in respect to its relationship to safety) because of its impact on the 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 mind set of 'what if' — is it 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 and the tails, with all their uncertainties, can be represented by the longer term 'judgements'. Because of the complex nature of the financial industries, these 'tails' often turn out to be completely misleading. Distributed approaches are sometimes applied to warhead safety assessments. Such distributions usually have far more evidential support near the peaks and less in the tails, where the 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.*

## 5.2 Taleb's View in Combatting Uncertainty

Because of the complexity associated with the financial industry, Taleb suggested that the effort expended in trying to prevent catastrophic occurrence was not well spent. He felt that catastrophes were inevitable and most effort should be best directed towards robustness in coping with Black Swan events when they do occur, rather than on predicting or preventing 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 on how to deal with them and mitigate the consequences. Given the assumption that the 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, rather than prioritising on the mitigation actions following such an event, if indeed this was realistically possible. For this reason, we concentrate on limiting complexity, particularly in the context of the principal safety*

[4] However measured

*arguments, in order to enhance transparency and confidence in the evidence to support an assessment of a very low occurrence probability.*

### 5.3 Charles Perrow

Perrow is well known for his work in accident theory (Perrow 2007). He provides another guide as to how Black Swans can arise, and what approaches should be put in place to avoid their occurrence. He has spent a great deal of his career in the field of historical accident assessment and the reasoning behind their occurrence. These have covered a wide range of subject areas, but their applicability are generally pertinent to nuclear warhead safety. The principles that clearly emerge from his work are again simplicity, clarity, and independence. These are evidenced by his observations and recommendations noted below.

1. *Complex, interconnected, and highly-coupled systems should be expected to fail*

**Response in warhead design:** Warheads by their very nature have to be interconnected but this is different in the sense of 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. As such 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 the safety ensurance and assurance sense. Design also aims to minimize complexity, and this helps towards unintended (undetected) safety interconnectivity. For example, safety systems strive to minimize component counts (and 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 being complex in nature. The more complex the safety argument the more challenges that can be realistically 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 to look 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 in relation to early detection of any flaws that might exist or may be developing over time, and is supported with a commitment to maintaining a strong cadre of experts probing in the relevant technology areas.

2. *Focus on those aspects with catastrophic consequences*

**Response in warhead design:** Although we focus on safety across the board, we nevertheless concentrate particularly on the worst-case events of Inadvertent Nuclear Yield and major radioactive material release, and consequently go to great lengths to prevent their occurrence both by way of design, associated processes and independent organizational assessment activity. Again, the prevention of Inadvertent Nuclear Yield and major release of radioactive material is based on a number of foundational safety principles, which 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 both for 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.

> *3. Strive for systems that are simple, easily understood, disconnected, and decoupled*

**Response in warhead design:** This is essentially a re-statement of point 1 above. We adopt this strategy for nuclear weapon design safety focusing on clarity, simplicity and strong independent arguments in the defence in depth approach. Safety is based on a limited number of very 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 analyse 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 design is confronted by constraints on available volume and mass and the need for the complementary requirement that these safety systems can be reliably removed when fully 'authorised' to do so, and such authorisation is based on a principle of uniqueness which effectively decouples it from the pre-authorisation safety argument.

# 6 Some Additional Factors

## 6.1 A Lesson from History — Predicting the Future based on the Past

In 1907, five years before he went down with his ship, the Captain of RMS Titanic, E. J. Smith, was quoted in a New York Times interview:

> *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.*

The clear lesson here is that past history (or perceived past history), whilst it may be a guide to the future, is only a guide and so gives *no* assurance about the future. Confidence (false) was also based on the safety built into the 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 an overestimate of the ship's capability to withstand such a collision. *This was an example of a 'Known — but not to the degree necessary'*. Similar arguments can be applied to the Columbia and Challenger tragedies where confidence was again 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 as such there is no guarantee they are fully understood with regards to the future. AWE however 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 UK's nuclear weapons ownership where there have been no major safety events, but one cannot take this as assurance in respect to the future. Certainly, the number of successful weapon lifetimes in our history comes nowhere near to proving compliance with the exacting safety standards that we strive to comply with. We certainly take heed of any issues that have arisen and have acted upon them with high priority.

### 6.2 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, has a long history of being undermined; note the US Minot incident[5]. 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 principally on scientific, technical and engineering principled 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, in the limit 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.

### 6.3 Conservatism, Worst Case and Probability Function

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 somewhat '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 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. Somewhat less conservative but still supportable statistical approaches, Figure 2, are then deemed more appropriate. This tends towards a more onerous approach (in application) and can carry with it a smaller conservative margin in protecting against uncertainty.

As noted previously, distributions often have greater evidential support near their peaks and far less in the tails, which typically characterise 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 only accepted if detailed scrutiny provides the evidence noted above and that in turn this evidence stands

[5] An unauthorized transfer of nuclear warheads between Minot Air Force Base, North Dakota, and Barksdale Air Force Base, Louisiana, 29th — 30th August 2007.

up to independent challenge. This typically applies to assessment of explosive response to environmental challenge.

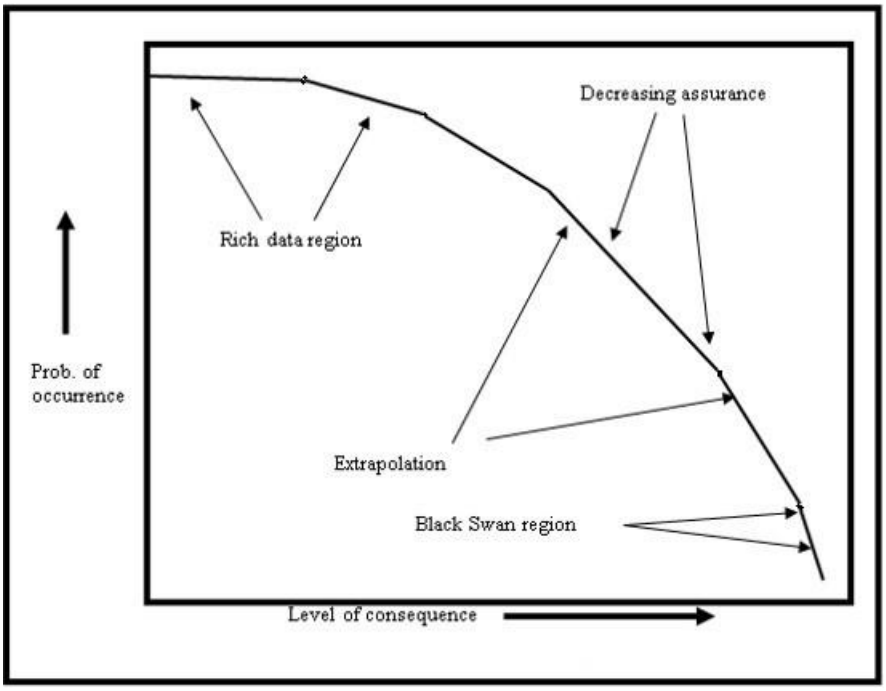


**Figure 2 ~ Distribution Approach to Safety Assessment**

## 7 The Importance of, and Achieving, Independence

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

- A full understanding of the inventory of possible hazardous sequences leading on from any given initiator and culminating in a mishap.
- 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 Jones (2017) in terms of the identification of the strength in depth of barriers or Lines of Defence (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 in order to meet the independence requirement. For example:

1. Fundamental physics or chemical properties based on the laws of nature, which directly support the LOD resilience
2. Uniqueness of response. In the sense that defeat of the LOD can only occur as a result of unique circumstances.
3. That microelectronics/firmware/software elements are clearly decoupled from the primary safety arguments.
4. Best practices engineered implementation of a concept to the extent that failure is unlikely even under challenging circumstances, e.g. accident conditions.

5. Procedure based. This relies on humans doing 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 can be broadly categorised as having the resilience properties of: **Isolation, Incompatibility, Inoperability, 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 deterministically make an LOD resilience case and to support the principle of LOD independence will follow this hierarchy. A structured sequence protection argument based strongly on correct LOD based procedures alone will not be acceptable. Although human LOD are least favourable in the overall choice, all LOD have some element of human involvement because of human relationship to judgement, 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 is enhanced by applying the following hierarchy of application for independence in as far as technical implementation allows:

1. Fundamentally different concepts.
2. Fundamentally different applications of the same concept.
3. Different engineering implementations of the same concept.
4. 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 propagating to the mishap conditions.

# 8 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. As such 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 of the 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 occurrence 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 the prevention of such environments propagating to the warhead. Nevertheless, such events cannot be totally discounted and as such the safety characteristics of a warhead design must show appropriate resilience against such threats.

# 9 Conclusions

Hunting for the Black Swan in the context of safety can represent a somewhat uncertain activity in the sense that one may be looking for something that does not exist. One cannot prove a negative. However, there is a compelling history of disasters which evidence the existence of such Black Swans which 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 and, as a result, have established a sound culture aimed at minimising the potential for such an occurrence. In addition, we need to 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 of 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 organisational independent assurance challenge. The latter covers the fidelity and completeness of the: logic, evidence, implementation and a search for any aspects which might have been missed or not covered in sufficient depth. Both contributions are then considered at the organisational executive decision-making level where a further level of knowledge and experience is applied. These activities go a long way to eradicating the presence of Black Swans but of course this does not guarantee that one is not present. The final layer in the defence comes from the strong organisational support to a continuing probing[6] culture which applies a 'what if' and 'expecting the unexpected' mind-set. Consequently, the campaign for assuring safety never ends. This latter contribution is based on a strong and continuing organisational 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 potential Black Swan identification and early eradication before such can manifest itself. This is exercised both 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 may appear to be somewhat 'undirected' in its approach. However, it is still the best way we know of with regard to tackling the demanding challenge of Unknown Unknowns. The overall strategy is hence based on:

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

### Disclaimer

The contents of this document represent the views of the author and not necessarily those of AWE plc.

[6] i.e. questioning and challenging

**Correspondence Address**

Malcolm Jones, Room 110, AWE, Aldermaston, Berkshire RG7 4PR UK

Malcolm.Jones@awe.co.uk

**Acknowledgments**

The author acknowledges the helpful, interesting and informative discussions held with members of AWE and the International System Safety Society when preparing this paper.

**References**

Jones M. (2016). *Organisational Problems — Potential Causes — Unintentional Consequences.* In: 34th International System Safety Conference, Orlando, Florida, USA, 8 -12 August 2016. System Safety Society, Unionville.

Jones M. (2017). *Strength in Depth and Quantitative Risk Assessment in the Context of Low Frequency High Consequence Systems*. In: 35th International System Safety Conference, Albuquerque, New Mexico, 21-25 August 2017. System Safety Society, Unionville.

Nelson R. (1964). *Fate Is the Hunter*. Arcola Pictures Corporation, Los Angeles.

Perrow C. (1972). *Complex Organizations: A Critical Essay*. McGraw-Hill, New York.

Perrow C. (2007). *The next catastrophe: Reducing our vulnerabilities to natural, industrial and terrorist disasters*. Princeton University Press, New Jersey.

Smith E. J. (1907). *Captain, RMS Titanic Quotes*. Quotes.net. Retrieved 5th May 2023 from https://www.quotes.net/quote/35858.

Taleb N. N. (2008). *The Black Swan: The Impact of the Highly Improbable*. Penguin Books, London.

This collation page left blank intentionally.