Reliability Engineering and System Safety 134 (2015) 83–91

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

# Implications of black swans to the foundations and practice of risk assessment and management

Terje Aven*

*University of Stavanger, Stavanger, Norway*

ARTICLE INFO

*Article history:*
Received 20 January 2014
Received in revised form
24 September 2014
Accepted 2 October 2014
Available online 30 October 2014

*Keywords:*
Black swans
Risk
Uncertainties

ABSTRACT

In this article, we discuss how to deal with black swans in a risk context. A black swan is here understood as a surprising extreme event relative to one's knowledge/beliefs, and can be of different types: a) unknown unknowns, b) unknown knowns (we do not have the knowledge but others do) and c) events that are judged to have a negligible probability of occurrence and thus are not believed to occur. In the article, we review the current approaches for confronting black swans, the aim being to gain new insights by addressing the three types of black swans separately, motivated by the fact that they require different types of measures. The main conclusions of the article are that there is a need to i) extend the current risk conceptualisation and treatment frameworks to include the black swan risk, ii) develop a new generation of risk assessment and decision support methods that place more emphasis on the black swan risk and iii) better understand what analysis captures and what lies within the management domain.


## 1. Introduction

Three major strategies are commonly used to manage risk: risk-informed, cautionary/precautionary, and discursive [33]. In most cases the appropriate strategy would be a mixture of these three. The risk-informed strategy involves the treatment of risk – avoidance, reduction, transfer and retention – by the use of risk assessments. The cautionary/precautionary strategy is also referred to as a strategy of robustness and resilience, and highlights features such as containment, constant monitoring, research to increase knowledge and development of substitutes. In the discursive strategy, measures are employed to build confidence and trustworthiness, through the reduction of uncertainties, clarification of facts, involvement of affected people, deliberation and accountability.

In this article, we focus on risk related to black swan type of events, here understood as surprising extreme events relative to one's knowledge/beliefs [5]. How should we confront this risk?

Taleb proposes "to stand our current approaches to prediction, prognostication, and risk management on their heads" ([41], p. 4, 5). When looking at much of the current thinking about risk assessment and management, with its focus on probability modelling and estimations, Taleb's view is understandable. However, risk management is required in order to find the proper measures to confront the occurrence of potential events. There are always limited resources available for this purpose, and the risk assessment provides decision support. The decision makers need to be informed about issues related to important precursors, the uncertainties, the knowledge available and so on. In a particular case, a decision maker may need to choose between investments in some measures that are effective in the case of some events but not in others, and investments in other measures with the reverse effect. Accurate predictions and estimates cannot be provided but, in most cases, informative risk descriptions can. This perspective seems to be the one adopted by Paté-Cornell [31] in her thought provoking analysis of how to confront the black swan risks. She takes an engineering risk analysis perspective, which highlights information gathering and analysis in support of proactive risk management decisions; due weight is given to the reinforcement of relevant systems, and thoughtful response strategies are developed.

In this article, we review current approaches to confronting black swans, having a special focus on the risk analysis approach and robust/resilient thinking. By distinguishing between different types of black swans, the article seeks to gain new insights on how to best deal with black swans. A specific issue raised is the use of risk assessments in such settings. What form and role can and should they take to provide useful decision support?

The article is organised as follows: First, in Section 2, we provide a detailed analysis of the concept of black swans. To provide a meaningful study on how to confront black swans, it is essential to have a clear understanding of this concept and the different types of black swans. Then, in Section 3, we perform the aforementioned review and discussion of some current approaches to confronting these

* Tel.: 47832267; fax: 4751831750.
*E-mail address:* terje.aven@uis.no

http://dx.doi.org/10.1016/j.ress.2014.10.004

events. This review is followed up in Section 4 with a discussion of how to meet the different types of black swans. Finally, Section 5 provides some conclusions.

The most novel part of the article is Section 4, where the analysis relates to a recent categorisation of black swans. However, also the other sections present to a large extent new and original material, by integrating and reflecting upon existing work, providing conceptual clarifications (as in Section 2), as well as providing guidance on how to best confront the black swans.

## 2. What is a black swan?

The metaphor and concept of the black swan has gained a lot of attention recently and is a hot topic in many forums that discuss safety and risk. In the scientific community it has also been a focus in the aftermath of Nassib Taleb's *The Black Swan* [40]. Taleb refers to a black swan as an event with the following three attributes. Firstly, it is an outlier, as it lies outside the realm of regular expectations, because nothing in the past can convincingly point to its possibility. Secondly, it carries an extreme impact. Thirdly, despite its outlier status, human nature makes us concoct explanations for its occurrence after the fact, making it explainable and predictable. Other definitions of black swans have also been suggested. Aven [5] refers to a black swan as a surprisingly extreme event relative to one's belief/knowledge, and in Aven and Krohn [8] three main types of black swan events have been identified based on this definition:

a) Events that were completely unknown to the scientific environment (unknown unknowns)
b) Events not on the list of known events from the perspective of those who carried out a risk analysis (or another stakeholder), but known to others (unknown knowns – unknown events to some, known to others)
c) Events on the list of known events in the risk analysis but judged to have negligible probability of occurrence, and thus not believed to occur.

The term “black swan” is used to express any of these types of events, tacitly assuming that it carries an extreme impact.

The first category of black swan type of events (a) is the extreme – the type of event is unknown to the scientific community. A good example is the effects of the thalidomide drug [38]. The drug was introduced in 1957 and not long after children were observed with gross limb malformations of an unusual form. In activities about which there is considerable knowledge, such unknown unknowns are likely to be rarer than in cases of severe or deep uncertainties.

The second type of black swans (b) is events that are not captured by the relevant risk assessments, either because we do not know them or we have not made a sufficiently thorough consideration. If the event then occurs, it was not foreseen. If a more thorough risk analysis had been conducted, the event could have been identified. The September 11 attack is a good example of this type of black swans.

The third category of black swans comprises events that occur despite the fact that the probability of occurrence is judged to be negligible. The events are known, but considered so unlikely that they are ignored – they are not believed to occur and cautionary measures are not implemented. An example is the event that an underwater volcano eruption occurs in the Atlantic Sea leading to a tsunami affecting, for example, Norway. The events are on the list of hazards and risk sources but then removed as their probability is judged as negligible. Their occurrence will come as a surprise. The tsunami that destroyed the Fukushima Daiichi nuclear plant was similarly removed from the lists due to the judgement of negligible probability.

The black swans' “surprising” aspect must always be understood in relation to by whom and when. Figs. 1–4 illustrate this. We consider an activity, for instance, the operation of an offshore installation, at a given future time period, for example, next year. We let C denote the consequences of the activity in relation to the values we are concerned about (life, health, environmental, assets). What C will be is unknown to us at time s; there are risks

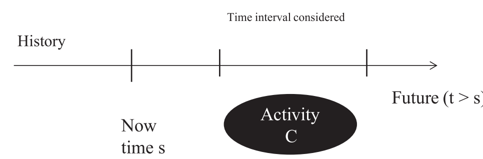


**Fig. 1.** Illustration of risk in relation to the time dimension. C: consequence of activity.

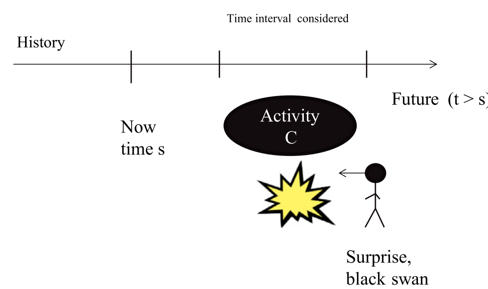


**Fig. 2.** Illustration of relationship between risk, black swan and the time dimension.

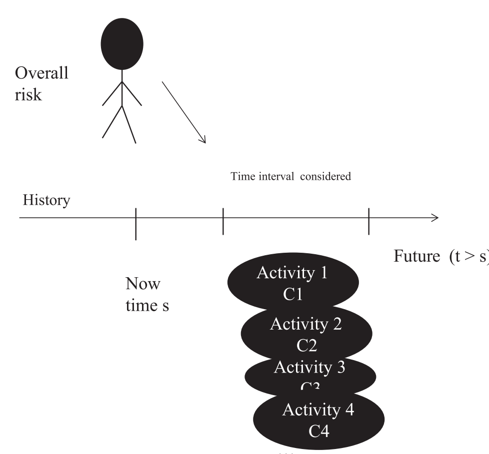


**Fig. 3.** Illustration of risk in relation to the time dimension when the perspective is macro, for example the whole oil and gas industry.

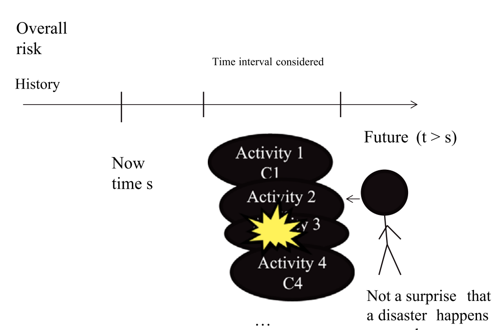


**Fig. 4.** Illustration of relationship between risk, black swan and the time dimension when the perspective is macro, for example the whole oil and gas industry.

present. We assume now that a risk assessment of the activity has been conducted at the time s. Time goes by, and C is realised, usually without a major accident occurring. But let us imagine that such an accident actually occurs, as shown in Fig. 2. Think about the Macondo accident as an example. The accident is a result of a combination of events and conditions occurring and comes as a surprise to those involved in the management of the activity. In the Macondo case, this combination includes [30]

- erroneous assessments of the results of pressure tests
- failure to identify that the formation fluid penetrated the well, despite the fact that log data showed that this was the case
- the diverter system was unable to divert gas
- the cutting valve (Blind Shear Ram) in the Blow Out Preventer did not seal the well

The accident is a black swan for them. It came as a surprise that such a sequence of events could occur; they had not thought of such a scenario. It is a black swan of type b). If they had thought of it, they would have concluded that it was so unlikely that one could ignore the possibility of the scenario occurring. It would have been a black swan of type c).

Now let us take a macro perspective – looking at a large number of such activities, for example, the whole oil industry. Risk is now linked to the occurrence of any major accident in the industry; where and how the event occurs is not the issue. Again, a risk assessment is conducted. It is concluded that there is a relatively high probability that such an accident could occur. Consequently, one cannot say that it is a black swan if such an event actually occurs (see Figs. 3 and 4). From a macro perspective, a realistic analysis would state that we must expect that a major accident will occur somewhere in the next ten years. However, there is no law that says that it will actually happen. We are not subject to fate or destiny. Each unit (organisation, company, installation) works hard to prevent such an accident from actually occurring. It is believed that with systematic safety work this goal can be achieved. Accordingly, any such serious accident normally comes as a surprise (as discussed by, e.g. [42]); it is a black swan for those involved in the operation and management of the activity. Hence, one must be careful in describing the perspective when discussing whether an event is a black swan.

Paté-Cornell [31] discusses the concept of black swans and relates it to the “perfect storm” metaphor. This storm resulted from the combination of a storm that started over the United States, a cold front coming from the north, and the tail of a tropical storm originating in the south. All three meteorological features were known before and occur regularly, but the combination is very rare. The crew of a fishing boat decided to take the risk and face the storm, but they had not foreseen its strength. The storm strikes the boat, it capsizes and sinks; nobody survives [31].

This extreme storm is now used as a metaphor for a rare event that may occur, where we understand the relevant phenomena. The experts can calculate the probabilities of such events and the associated risks with a high degree of precision. They can make accurate predictions of what will happen, stating that in one in ten such situations the waves will be like this, and in one in 100 such cases the waves will become so big and so on. When we build oil and gas installations offshore we take into account such events. We set requirements for the installation's strength to enable it to withstand extreme waves, but there is always a limit. We must accept that there may be a wave that is so large that the installation will not tolerate it, but such an event should have a very small probability.

The situation has similarities to that found in other areas such as health and traffic. We know quite precisely in many cases what proportion of the population will contract certain diseases next year and how many people will die in traffic. Actions can be taken to reduce risks, and we can measure changes over time. When one looks at the number of fatalities from the 1970 s to the present day, the figures show a steady decrease, despite the fact that traffic has increased. The risk management works.

The metaphor of the “perfect storm” is thus about events where science in traditional form prevails, where we have precise probabilities and relevant statistics, and where we can make accurate predictions about the future. The black swan type of event c) seems to be covered by the rare event of the perfect storm form. There is, however, an important difference. In relation to perfect storms, the variation in the phenomena is known; we face risk problems where the uncertainties are small; the knowledge base is strong and accurate predictions can be made. As the knowledge base is so strong, black swans can for all practical reasons be ignored. The probabilities are frequentist probabilities, characterising the variation in the phenomena, and they are known to a degree that is viewed as certainty.

For the black swans of type c), we are in a situation where we cannot make this type of accurate prediction. The variation in the phenomena cannot be described with this type of precision. We need to rely on judgements (subjective judgements), where probability refers to the knowledge-based (judgemental, subjective) assignment of uncertainties and degrees of belief. When stating that an event is judged to have negligible probability and not believed to occur, it is with reference to such a perspective. Clearly, in such cases we may experience surprises compared to the judgements made.

Think again about the oil and gas installation. The management of the installation may ignore the possibility of a specific event occurring, arguing in this way. It is not a perfect storm type of event as it cannot be predicted with accuracy. Taking the macro perspective for the industry as discussed above, we are closer to the perfect storm situation. If we consider the industry as a whole, it does not make sense to talk about black swans as the probability of occurrence is rather high. But let us make a thought-construction; we assume that the frequentist probability of the occurrence of such an event is rather low. It is a rare event. Would it then be a black swan? No, is the answer, as the variation is still known, the phenomena studied are well understood. Further reflections may however challenge this view.

If we have a situation with perfect information about the variation of the phenomena – we know the frequentist probability distribution (we are in the perfect storm situation); one can argue that the occurrence of a low frequentist probability event should not come as a surprise. It is rare, but it is known with certainty that the event will occur sooner or later. Hence, it is not a black swan (type c). However, one can also argue differently. Given the knowledge about the variation in the phenomena, it is considered so unlikely that the event will occur the next year, say, that it is not believed to occur. Hence, it can be viewed as a black swan of type (c) if it in fact does occur. Again we see that whether the event is a black swan or not is in the eyes of the beholder.

In practice, we cannot fully understand the variation. If we go back some years, we would not think about terrorism events as a contributing factor to the variation, and hence a black swan may occur even if the phenomenon is considered well understood. This type of black swans is, however, not of type c), but of type a) or b). The discussion here relates to the distinction between common-cause variation and special-cause variation in the quality discourse [12,16,36,37]. The common-cause variation captures “normal” system variation, whereas the special causes are linked to the unusual variation and the surprises, the black swans [7].

We may also talk about “near-black swans”, meaning surprises relative to one's knowledge/beliefs, but where the event did not result in extreme consequences; the barriers worked and avoided the extreme outcomes. A black swan can occur as a result of a set of events and conditions, and a subset of these may generate a near-black swan.

## 3. Review and discussion of some basic approaches for managing risk and black swans

The introductory section pointed to some of the main approaches to treating risk and surprises. The easiest case to handle is risk problems where the uncertainties are small – the knowledge base is strong and accurate predictions can be conducted; that is, the situation is characterised by perfect storms. For this type of problem, standard risk analysis using statistical methods, as described for example by Paté-Cornell [31], can be used to make rational decisions, as seen for instance in the traffic and health areas. However, few problems in real life can be classified as perfect storms; surprises and black swans may occur, and the issue is how to confront this type of risk.

It is obviously not straightforward to assess and manage the black swan type of risks, and different approaches are recommended. Perhaps the most commonly referred to is the use of precursors of such events [31,19,43,23]. Through a mix of alertness, quick detection, and early response, the black swan event can be avoided.

But how can we avoid missing or ignoring early signals and precursors of serious events, or, on the opposite side, exaggerating them? It is common practice to refer to false negatives (no indication of a risk situation when one is actually present) and false positives (erroneous signals indicating some risk situation is present when it is not), but how can we make judgements about these “errors” when we do not know the outcomes of the events or situations under observation before they occur? It is easy to identify (and claim) that we missed a risk event or situation with hindsight a posteriori, when the accident, disaster or crisis has occurred, but how can we know in advance that we are missing, ignoring or exaggerating signals or precursors, given that we are typically exposed to a large number of threats/hazards? The reference for our evaluation of the signals and precursors cannot be the unknown consequences or outcomes of events yet to occur. The only possible way out seems to be to rely on the results of risk and uncertainty assessments, where the warning system itself can be viewed as a form of risk assessment. However, risk assessment has its limitations as a tool for this purpose and, in cases when the knowledge base is not strong, we need to base the judgements on hypotheses and assumptions, and we may act too slowly (or too quickly). An example is the AIDS epidemic, which was detected in the United States by the Center for Disease Control in 1981, and given that it had probably been spreading for decades, the response was rather slow [31].

This leads us to the use of two other approaches for dealing with black swan type of events: adaptive risk analysis and robust analysis. Adaptive analysis is based on the acknowledgement that one best decision cannot be made but rather a set of alternatives should be dynamically tracked to gain information and knowledge about the effects of different courses of action. On an overarching level, the basic process is straightforward: one chooses an action based on broad considerations of risk and other aspects, monitors the effect, and adjusts the action based on the monitored results [26]. A central idea is that because uncertainty is pervasive, one optimal management choice is not achievable; rather, we have a range of often competing decision alternatives, which we dynamically track to gain information and knowledge about the system, and about the effects of different courses of action. In this way we may also avoid black swan type of events.

In non-trivial systems, surprises will occur and then we need to be able to absorb and analyse relevant information and take adequate actions. Adaptive risk analysis can be a useful tool in this respect. So is abductive thinking:

> You observe a fact … In order to explain and understand this, you cast about in your mind for some glimmering theory, explanation, flash and so forth. The process of abduction takes place between the result (observed fact) and the rule (explanation), and concludes with the positioning of a hopefully satisfactory hypothesis. ([18], p. 183)

As presented by Pettersen [32], abduction can be seen as the process of noticing an anomaly and getting an explanatory hunch [14]. By means of abduction, a new idea (or hypothesis) is brought up from the region where “all things swim”. This process can be shown as a three-step process:

1. A surprising fact is noticed.
2. An aesthetic (unfettered) exploration of qualities and relationships is made.
3. Abductive reasoning is applied to make a guess that could explain the surprising fact (Chiasson, 2001) [14].

In a process plant, abduction could mean to notice that the pressure is increasing, explore why, and provide a hypothesis to explain it. Testing can then be carried out to prove or disprove the hypothesis. The approach is in line with fundamental ideas of the quality discourse, which highlights that knowledge is built on theory ([25]; see also [12]). As formulated by [16], rational prediction and analysis requires theory and builds knowledge through systematic revision and extension of theory based on a comparison of prediction with observation. Without theory, experience has no meaning, and without theory there is no learning.

Bayesian decision analysis provides a strong theoretical framework for choosing optimal decisions in the case of information in the form of signals and warning, but, in many cases, it is difficult to use in practice. Instead, based on a crude assessment of risk and other relevant concerns, we may search for procedures that prescribe what to do for given signal/warning levels. The idea is simply to make such an assessment for different signal/warning levels, and establish some adequate decision rules for how to act in the different cases. This would give a level of preparedness in the case of specific signals/warnings but would not necessarily provide much support in the case of surprising events. It would be impossible to prescribe what to do in all cases; hence, the approach needs to be supplemented with other methods.

Following Cox [15], we can write the robustness problem in general terms as (C,P,u,a), where C are the consequences of the actions, P the probabilities of C given the actions, u the utility function of C, and a the actions. However, it may be difficult to assign some of these values, for example, P, when the uncertainties are large. A robust approach is then required. The key is to make decisions that are good for a set of values of, for example, C and P, and in this way the approach can also withstand some types of surprises. However, the set-up used for robustness analysis often excludes the possibility of many forms of black swans, as the framework reflects the current knowledge and beliefs. The protective measures could, for example, be based on a probability model reflecting variation due to a set of key risk sources but fail to include an important one. Robust analysis is not easily conducted in practice, as for example, discussed by Aven [6]. There are many ways of looking at robustness, and it is difficult to find arguments for why some are better than others.

It has proven helpful to distinguish management strategies for handling the risk agent (such as a chemical or a technology) from those needed for the risk-absorbing system (such as a building, an organism, or an ecosystem) [33]. With respect to the risk agent, risk assessments can be useful, but equally important are cautionary and precautionary strategies (see Section 4.3), which include principles such as containment, substitution, safety factors, ALARP (as low as reasonably practicable), redundancy and diversity in designing safety devices, and best available technology. For the risk-absorbing systems, robustness and resilience are two main categories of strategies or principles when studying risk problems characterised by moderate or large uncertainties. Measures to improve robustness include [33,9] inserting conservatisms or safety factors as an assurance against variation, introducing redundant and diverse safety systems to meet multiple stress situations, avoiding high vulnerabilities, establishing building codes and zoning laws to protect against specific hazards. See Joshi and Lambert [21] for an example of such a “robust management strategy” using diversification of engineering infrastructure investments.

It is an objective to make the risk-absorbing systems resilient so they can withstand surprises. Resilience is a protective strategy against unforeseen or unthinkable events, and key instruments for it include “the strengthening of the immune system, diversification of the means for approaching identical or similar ends, design of systems with flexible response options and the improvement of conditions for emergency management and system adaptation” ([9], p. 129).

Resilience engineering has become an important field for the understanding and management of safety in socio-technical systems; see, for example, Hollnagel et al. [19]. In order to be resilient, a system or an organisation must have the following four qualities: the ability to

I) respond to regular and irregular threats in a robust, yet flexible manner,
II) to monitor what is going on, including its own performance,
III) to anticipate risk events and opportunities and
IV) to learn from experience [19].

This approach also highlights signals and precursors of serious events. It is a common feature of most approaches that intend to meet surprises. If we look at basic insights from organisational theory and learning, we see that both this feature and resilience are main building blocks. A good example is the concept of collective mindfulness, linked to high reliability organisations (HROs), with its five principles: preoccupation with failure, reluctance to simplify, sensitivity to operations, commitment to resilience, and deference to expertise. There is a vast amount of literature (see, e.g. [20,24,43,44]) providing arguments for organisations to organise their efforts in line with these principles to obtain high performance (high reliability) and effectively manage risks, the unforeseen and potential surprises.

In addition, the quality discourse, with its link to “common-cause variation” and “special-cause variation”, and the continuous focus on learning and improvements, should be mentioned as an approach for dealing with surprises and black swans, as noted in the previous section. The focus on improvements leads us to the concept of antifragility [41]. According to Taleb, the antifragile is seen as a blueprint for living in a “black swan world”, the key being to love randomness, variation and uncertainty to some degree, and thus also errors. We all know that to be in top shape and improve our bodies and minds we need some stressors, so do other activities and systems [8].

## 4. Confronting the three types of black swans

Meeting black swans is obviously difficult as they extend beyond current thinking, but as we have seen there exist a number of different approaches, strategies and measures that can be used to confront such events. In this section, we seek further insights by specifically addressing the three types of black swans a)–c) described in Section 2, with emphasis on the b) and c) categories. When writing “black swans”, we tacitly also think about “near-black swans”.

### *4.1. Black swans of the unknown unknown type*

Unknown unknowns are events that were completely unknown to the scientific environment, and it is of course difficult to be prepared for such events. Focusing on resilience and signals & warnings provides useful general means, in addition to general scientific work generating knowledge about the relevant phenomena. Increased knowledge reduces in general the probability of a black swan of this type. Hence, testing and research are generic measures to meet this black swan type of risk.

Take the swine flu in 2009 as an example. It was caused by a type A influenza (H1N1) virus. The World Health Organization declared that the flu had developed into a full-scale world epidemic, and a vaccine was quickly developed. In some countries (Sweden, Finland, Norway and Iceland), the authorities explicitly set the goal of vaccinating the whole population. The illness turned out to be quite mild, but it had some severe side effects that were previously unknown; see, e.g. Munsterhjelm-Ahumada [29]. These side effects came as surprises; they were black swans of the type unknown unknowns.

The vaccination was carried out because the authorities believed that the flu itself would cause serious illness and problems, at a much higher level than the side effects. Normally there is time for fairly thorough testing of the vaccine to control the risk related to side effects, but in 2009, this was not the case. The uncertainties were large.

Obviously, there could be unknown side effects in the case of vaccination, and analyses and judgements need to be conducted to characterise the risks. The problem here was that the decision concerning vaccination had to be taken so quickly. It was impossible to avoid a weak knowledge base. There was no time for thorough testing and research, and adaptive management. The authorities also had to balance the need for faithful risk characterisations and the desire to get the population vaccinated. In the Nordic countries mentioned above, the authorities initiated public relations campaigns, which can be described as “moral persuasion”. Solidarity became the slogan: Be vaccinated to protect your fellow citizens [29].

Faithful risk characterisations addressing possible unknown side effects (black swans) were not very well highlighted; the black swan risk was not really an issue. One can speculate whether it was a deliberate policy. For sure, the decision was a difficult one for the authorities because of the time pressure; they had to balance difficult judgements about the development of the flu, efficiency of the vaccination, risk and uncertainty issues, as well as ethical aspects.

On this basis, it is not surprising that so many people decided to take the vaccine. The decision became quite easy, to follow the advice from the authorities.

From the individual person's point of view, one may argue that the black swan risk should have been reported more faithfully. To make an adequate decision, one has to be risk-informed. It is, however, not straightforward how such information should have been best communicated. The knowledge base is weak, and it is impossible to express meaningful numbers characterising risk. Instead, we have to rely on more general qualitative statements. Here is a suggestion for how the risk could have been described and communicated:

> This vaccine could have unknown side effects. There are uncertainties. We think it is unlikely that severe side effects

will occur, but the knowledge base is rather weak and we cannot exclude the possibility.

It is not enough to limit the statements to probability characterisations. Expressing that it is unlikely that severe side effects will occur without also referring to the knowledge base and the potential for surprises would mislead the receiver. The knowledge supporting the probability is as important as the probability itself.

### 4.2. Black swans of the unknown known type

A black swan of this type is an event that is not on the list of those identified by the relevant risk assessment, but it is not an unknown type of event. Its possible occurrence is known by other persons, groups, or communities. We can formalise it in this way: A' are those events that we have identified in the risk assessment, and A is the occurrence of the actual event, which is a type of event known by others than those involved in the risk assessment. The event is a black swan (or a near black swan) of this type (unknown known) if A is not covered by A'.

To meet this black swan type of risk, we need

– Improved risk assessment to identify these events
– Improved communication to transfer knowledge to relevant persons

Many types of traditional risk assessment methods address the issue of what can happen, for example, HAZOP, HazId, FMEA, fault tree and event tree analysis [45,28]. These methods have been shown to work in practice, but we always need to look for potential improvements, as the identification of the event is an extremely important part of risk assessment and management – if the event is left out, the risk management will also easily leave it out. One non-traditional method worth mentioning here is anticipatory failure determination (AFD) [22]. It is a hazard/threat identification and analysis approach, which utilises I-TRIZ, a form of the Russian-developed Theory of Inventive Problem Solving. Traditional failure analysis addresses the question, “How did this failure happen?” or “How can this failure happen?”. The AFD-TRIZ goes one step further and questions “If I wanted to create this particular failure, how could I do it?” The power of the technique comes from the process of deliberately “inventing” failure events and scenarios [27]. Scenario analysis [13] could also be a useful tool to generate events and scenarios. Here, there is no search for completeness and descriptions of the uncertainties and risks, as in traditional risk analysis. The anticipatory backwards scenarios are of particular importance – starting from a future imagined event/state of the total system, the question is asked: what is needed for this to occur?

There is potential for further development of analysis methods to identify events/scenarios along the lines of AFD, using different types of creative thinking; there are many others than TRIZ, as discussed by, for example, Sternberg [39]. Such developments should have a particular focus on aspects that block knowledge transfer and cause a poor or inadequate understanding of the studied system. As an example, say that the actual system has some special features, which give it a special sensitivity (vulnerability) with respect to some specific operational conditions. We can think of an operation on a person who has a special sensitivity with respect to some substances. Then it is essential that the scenario/event identification is based on an acknowledgment of this sensitivity [34].

Such special features can be seen as a special case of a more general problem: the implications of the analysts basing their judgements on more or less clear assumptions, hypotheses and explanations. A key task of the event/scenario identification is to challenge these. A method often used for this purpose is *red teaming*, which serves as a devil's advocate, offering alternative interpretations and challenging established thinking [27]. For example, “businesses use red teams to simulate the competition; government organisations use red teams as ‘hackers’ to test the security of information stored on computers or transmitted through networks; the military uses red teams to address and anticipate enemy courses of action” ([1], p. 136). Red teaming challenges assumptions, generalisations, pictures, or images that influence how we understand the world and how we take action, that is, our mental models [35].

In the following, an example of how a red-team analysis can be used to improve a standard event/scenario identification is outlined:

### 4.3. Adjusted event/scenario process using a red team

The assessment process has three main stages, involving two analyst teams, referred to as teams I and II. In *Stage 1*, analyst team I performs a standard event/scenario identification (let us denote the events/scenarios by $A_1$'). This stage includes a self-evaluation of the analysis, where the focus is on the mental models (assumptions etc.) that could restrict the event/scenario's space.

In *Stage 2*, analyst team II challenges team I and their mental models, acting as a red team (the devil's advocate), and, for example,

- argues against the mental models used by team I
- searches for unknown knowns.

In the final *Stage 3*, the two analyst teams are to provide a joint list $A_2$'. In practice, the process needs to incorporate aspects of probability and uncertainties – this will be explained in Section 4.3.

Returning to the example of a system with special features, a key purpose of the red team in *Stage 2* is to challenge the mental models of analyst team I – which could mean that the team believes that the system has no special features. By critically questioning the validity of assumptions and pictures/images of the system being studied, such features can be revealed and the list of event/scenarios is likely to be adjusted.

As some people – but not those involved in the risk assessment and the related management process – possess the knowledge, the analysis aims to achieve knowledge building and transfer of knowledge. To obtain such results, communication is essential. It could be communication between groups of people from different organisational units and between individuals. We may think of the Deepwater Horizon accident, where a worker did not alert others on the rig as pressure increased on the drilling pipe, a sign of a possible “kick” [17]. A kick is an entry of gas/fluid into the wellbore, which can set off a blowout. The worker had the knowledge (information), but this knowledge was not communicated to the right people.

### 4.4. Black swans of the probability type judged negligible

This third type of black swan events is those that are on the list of known events in the risk analysis but whose probability of occurrence is judged negligible, and thus are not believed to occur. Yet such events do occur, as was discussed in Section 2. How should we deal with this type of events? Should we just accept the risks – we have accepted that there is a small probability of an extreme event occurring – meaning that the event could occur?

We remember the discussion in Section 2 about this issue, relating to perfect storms. For these black swans of type c), we are in a situation where we cannot make accurate predictions as in the case of perfect storms. The variation in the phenomena cannot be described with this type of precision. The knowledge base is weak, and the probabilities are subjective (judgemental, knowledge-based), more or less strongly founded. There is no clear correspondence between the probability assignments and the actual occurrence of the events. Hence, it is appropriate to scrutinise both the judgements about acceptable risk and negligible probability, and the background knowledge that supports these judgements.

Such a scrutiny needs to be based on the acknowledgement that

i) acceptable risk should not be determined by judgement about probability alone
ii) events may occur even if very low probabilities are assigned
iii) cautionary and precautionary principles constitute essential pillars of the risk management linked to such events (black swans).

The risk analyst may derive a set of probabilities for specific events to occur and combine them with different loss categories, but these numbers must be seen in relation to the strength of knowledge that supports the probabilities. We may have two situations, giving the same probabilities: one where the assignment is supported by a strong evidence base, and the other, which relies on very poor background knowledge [10]. In engineering contexts, common practice is built on probabilistic criteria (like a $1 \cdot 10^{-4}$ probability limit) to determine what is an acceptable design [11]. Such an approach cannot in general be justified as it ignores the degree of knowledge that supports the probability assignments. The assignments can be based on many critical assumptions, and these assumptions could conceal important aspects of risk and uncertainty [4]. As discussed in Section 4.2, we may for instance assume that the present system is a standard one, but it may turn out to have special features, for example, being extremely sensitive to some specific hazards.

For a general discussion of such criteria, see Aven and Vinnem [11]. These authors state that such criteria must be used with care as they can easily lead to the wrong focus, meeting the criteria instead of finding the overall best arrangements and measures. However, for the practical execution of risk management activities, it is not difficult to see that some type of criteria may be useful in simplifying the decision-making process. To be able to meet the above critique related to the strength of knowledge supporting the probability assignments, an adjusted procedure has been suggested if such criteria are to be used [4] (see also Table 1):

1. If risk is found acceptable according to probability with large margins, the risk is judged as acceptable unless the strength of knowledge is weak (in this case the probability-based approach should not be given much weight).
2. If risk is found acceptable according to probability, and the strength of knowledge is strong, the risk is judged as acceptable.
3. If risk is found acceptable according to probability with moderate or small margins, and the strength of knowledge is not strong, the risk is judged as unacceptable and measures are required to reduce risk.
4. If risk is found unacceptable according to probability, the risk is judged as unacceptable and measures are required to reduce risk.

The approach relies on cautionary thinking. It generates a process that looks for measures to reduce risk and avoid the event occurring – despite the fact that the judged probability is very low.

The cautionary principle states that, in the face of uncertainty, *caution* should be a ruling principle, for example, by not starting an activity or by implementing measures to reduce risks and uncertainties [9]. The level of caution adopted has, of course, to be balanced against other concerns, for example, costs, but to be cautious goes beyond balancing the expected benefit of risk reductions expressed in monetary terms against expected costs. The precautionary principle may be considered a special case of the cautionary principle, in that it applies in the face of *scientific uncertainties* [3].

Think about the vaccine example of Section 4.1. Here, we faced scientific uncertainties concerning the side effects, and if a person did not take the vaccine he or she could refer to the precautionary principle. Health experts could state that the probability of side effects is low and hence acceptable, but the strength of knowledge supporting this type of statement is poor, and there is a need for considerations that give due weight to the uncertainties as discussed above.

Many safety and security measures are justified by reference to the cautionary principle as referred to in Section 3. We implement robust design solutions to be able to meet deviations from normal conditions, we implement emergency preparedness measures even if the probability of their use is very small and so on (Aven et al., 2007) [11]. We see beyond the probabilities as we know that surprises can occur relative to our judgements. This is to be cautious.

Consider the Fukushima Daiichi nuclear disaster in Japan in March 2011. Here, the probability that such an event would occur was considered negligible. The judgement was based on many considerations and assumptions as discussed by Paté-Cornell [31]. These considerations and assumptions can obviously be questioned as earthquakes from the ninth and seventeenth centuries causing tsunamis reaching heights far beyond the design criterion of the plant were not accounted for in the design of the nuclear reactors.

It is not obvious that better risk assessment would have led to vital changes in the Fukushima case, but it could have. Several weaknesses in the Fukushima assessments are pointed to by Paté-Cornell [31], and there is also potential for improvements of the risk assessments in general, as mentioned in Section 4.2, by giving further attention to the knowledge and surprise dimensions. We can illustrate this by returning to the adjusted event/scenario process using the red team approach described in Section 4.2. This analysis can be modified to capture a more general risk assessment process. The method follows the same steps as the one described in Section 4.2, but its focus is not only the events and scenarios A', but also the consequences of these events/scenarios, C', as well as descriptions/measurements of related uncertainties

**Table 1**
Adjusted procedure for use of risk acceptance criteria in view of considerations of the strength of knowledge.

| | | | | |
|---|---|---|---|---|
| **Probability-based justification** | *Above limits* | Unacceptable risk | Unacceptable risk | Unacceptable risk |
| | *Small margin below* | Unacceptable risk | Unacceptable risk | Acceptable risk |
| | *Large margins* | Further considerations needed | Acceptable risk | Acceptable risk |
| | | *Poor* | *Medium* | *Strong* |
| | **Strength of knowledge** | | | |

(typically using probability) and judgements of the strength of knowledge supporting these descriptions/measurements. The assumptions have a key role in the evaluations, and “judgements of the assumption deviation risk”, as discussed by Aven [4], is an example of a tool that can be used to study the importance of these assumptions. Under *Stage 3*, analyst team II challenges team I and their assumptions, acting as a red team, and argues for the occurrence of events with assigned negligible probabilities.

Cautionary thinking also applies to the use of cost–benefit analyses, where the purpose is to determine whether a measure should be implemented – balancing different concerns. This tool is based on expected values and hence important aspects of variation, uncertainties and risk are not reflected [11]. A safety measure may not be justified by reference to a cost–benefit analysis, but cautionary thinking, highlighting the additional uncertainties and risk, could justify it. Such cautionary thinking is reflected in the adjusted implementation scheme for the ALARP principle presented in Aven [4].

## 5. Conclusions

To confront possible black swans, we need to balance risk-based approaches, cautionary/precautionary (robustness, resilience, adaptive) and discourse-based approaches. This is the general answer and is fundamental for risk management, with and without special focus on surprises and black swans. Only in cases where the knowledge is very strong and the uncertainties small, can the risk-based approach be used alone. In most situations, all three strategies are required. The challenge is to find an adequate balance between these approaches and strategies, often between the first two. When the stakes are high and the uncertainties large, we obviously need to highlight robust and resilient solutions and arrangements to be prepared in case some extreme unforeseen events should occur. Potential surprises and black swans call for robustness and resilience, and antifragility, as discussed in Section 3.

We need to pay more attention to certain aspects compared to typical current practices, and the article has pointed to and highlighted some measures, linked to the way we should think in relation to these issues. The goal has been to help analysts and decision makers in two main ways:

1. It provides appropriate concepts and a platform for a deeper understanding of what the risk associated with surprises and black swans is all about.
2. It provides analysis and management principles that can prevent, or at least reduce, the probability of black swan type of events (which have negative consequences), and in addition stimulate and lay a basis for the development of appropriate specific methods that can give such an effect (reduce the risk related to black swan type of events).

In risk analysis, events and scenarios are identified, for example, hydrocarbon leaks, and barrier systems are in place should such events in fact occur. Many types of events happen in the course of a year, but they have not serious implications as the barriers work as intended. This is also the case for near misses – but the margins are in some cases small. Minor changes could have resulted in a disaster. When a major accident occurs, it is often because there are several “surprising events”. To address these issues, it is important to have an understanding of the various concepts and how they relate. The article contributes to this understanding (point 1 above).

In addition, the article provides specific help on how we should proceed in order to meet the potential black swan risks. Knowledge and uncertainty are key concepts. Black swans are surprises in relation to someone's knowledge and beliefs. In the September 11 example, some people had the knowledge, others did not. In the Fukushima example, it was the judgements and probabilities that were essential, but they are based on data, information and arguments/opinions, so here too the issue is knowledge. We must think beyond current practice and theory. We need new principles and methods. The article aims at contributing to such developments by outlining some ideas and laying a foundation for further research in this topic.

These ideas and this foundation allow for and encourage considerations and reinterpretations of the way risk is assessed at different stages of an activity; these are essential features of a management regime supporting continuous improvement. Current risk perspectives based on probability are considered to be less adequate for this purpose, as the frameworks presume some stronger level of stability in the processes analysed.

Various types of analyses, including robustness analyses, can provide insights and decision support, and, in many cases, the use of different types of such analyses may be useful to inform the decision maker, but we have to acknowledge that there is often considerable arbitrariness in the choices made by the analysts and all the tools used have strong limitations. As a consequence, care should be shown in making too strong conclusions based on the results of such analyses. There is always a need for a managerial review and judgement, which places the results of the formal analyses in a broader context where the limitations and boundaries of the analyses are taken into account before a decision is made. A decision mechanically determined by the analytical approach can seldom be justified.

The risk-based approaches incorporate risk assessments but need to be extended and have a broader scope than the standard probabilistic analysis commonly seen in textbooks and practice today. The current risk conceptualisation and treatment frameworks should be extended to include the black swan risk, and a new generation of risk assessment and decision support methods needs to be developed, which places more emphasis on this risk. Some ideas have been outlined in Section 4. For the type c) black swans, where the occurrence is not foreseen because of low assigned probability, a proper understanding of the risk and probability concepts is essential. Low probability events may occur, and we need to scrutinise both the judgements made and their basis. Improved tools are required for this purpose.

To obtain such improvements, we need a platform that incorporates adequate concepts, assessments and management principles and methods. This is a research issue, and to this end we find contributions in, for example, Aven [2] and Aven and Krohn [8].

In Aven and Krohn [8], new conceptual risk frameworks highlighting knowledge and uncertainties are combined with specific insights capturing variation, knowledge and uncertainties as mentioned. The main input comes from the collective mindfulness concept linked to HROs, the quality discourse with its focus on variation, system thinking and continuous improvements, as well as the concept of antifragility [41].

There is a common belief among many engineers and managers that to manage an activity, and avoid accidents and perform operations as planned, it is sufficient to develop procedures and ensure compliance with these. Such a compliance perspective fails, however, in practice for non-trivial activities, as a perfect system cannot be developed; surprises always occur. The system understanding is too static, and improvements and excellence are not sufficiently stimulated. We have to acknowledge that to obtain excellence and avoid accidents we need to acknowledge the performance, risk and knowledge “dynamics”. We need to see beyond compliance. For many types of systems, the signals and warnings are of a form that requires judgements and actions that need considerations beyond specified procedures.

This means that a traditional engineering risk analysis perspective cannot be used to meet the black swan type of risk. We need to incorporate the dynamic aspects, which are better covered by principles such as those of the collective mindfulness concept and by resilience engineering, as well as tools of adaptive analysis and similar approaches. Through the study of the three types of black swans, we have seen the importance of analysis to gain knowledge, but also the need for recognition of what is outside the analysis sphere. With the perspectives on risk adopted in this article, there is an acknowledgement of uncertainty and strength of knowledge as key pillars of the risk descriptions, and the need for a proper understanding of what constitutes assessment and what belongs to the management sphere, is essential. There will always be a need for managerial review and judgement that sees beyond the assessment part, as thoroughly discussed by Aven [6]. The cautionary and precautionary principles need to be seen as rational elements of any approach to meet the black swan risks, as they provide guidance on how to deal with the leap from assessment to decision-making.

## Acknowledgements

The author is grateful to two anonymous reviewers for their useful comments and suggestions to the original version of this article. The work has been partly funded by the Norwegian Research Council – as a part of the Petromaks 2 programme. The support is gratefully acknowledged.

## References

[1] Ambrose F, Ahern B. Unconventional red teaming, anticipating rare events: can acts of terror, use of weapons of mass destruction or other high profile acts be anticipated? <http://redteamjournal.com/papers/U_White_Paper-Anticipating_Rare_Events_Nov2008rev.pdf>; 2008 [accessed 15.01.14].
[2] Aven T. Quantitative risk assessment. The scientific platform. Cambridge: Cambridge University Press; 2011.
[3] Aven T. On different types of uncertainties in the context of the precautionary principle. Risk Anal 2011;31(10):1515–25 (With discussion 1538-1542).
[4] Aven T. Practical implications of the new risk perspectives. Reliab Eng Syst Safe 2013;115:136–45.
[5] Aven T. On the meaning of the black swan concept in a risk context. Safe Sci 2013;57:44–51.
[6] Aven T. On how to deal with deep uncertainties in a risk assessment and management context. Risk Anal 2013;33(12):2082–91.
[7] Aven T. On the meaning of the special-cause variation concept used in the quality discourse-and its link to unforeseen and surprising events in risk management. Reliab Eng Syst Safe 2014;126:81–6.
[8] Aven T, Krohn BS. A new perspective on how to understand, assess and manage risk and the unforeseen. Reliab Eng Syst Safe 2014;121:1–10.
[9] Aven T, Renn O. Risk management and risk governance. London: Springer Verlag; 2010.
[10] Aven T, Zio E. Some considerations on the treatment of uncertainties in risk assessment for practical decision-making. Reliab Eng Syst Safe 2011;96:64–74.
[11] Aven T, Vinnem JE. Risk management. Berlin: Springer Verlag; 2007.
[12] Bergman B. Conceptualistic pragmatism: a framework for Bayesian analysis? IIE Trans 2009;41:86–93.
[13] Chermack TJ. Scenario planning in organizations. San Francisco: BK Publishers; 2011.
[14] Chiasson P. Abduction as an aspect of retroduction, in the digital encyclopedia of Charles S. Peirce, 〈http://www.digitalpeirce.fee.unicamp.br/home.htm〉; 2001 [accessed 21.10.14].
[15] Cox LAT. Confronting deep uncertainties in risk analysis. Risk Anal 2012;32:1607–29.
[16] Deming WE. The New Economics. 2nd ed. Cambridge, MA: MIT CAES; 2000.
[17] Financial Post. Deepwater rig worker weeps as he admits he overlooked warning of blast that set off America's worst environmental disaster, 〈http://business.financialpost.com/2013/03/14/halliburton-worker-weeps-as-he-admits-he-overlooked-warning-of-blast-that-set-off-americas-biggest-oil-spill-in-gulf/?__lsa=42e0-28bb〉; 2013 [accessed 15.01.14].
[18] Harrowitz N. The body of the detective model: Charles S. Peirce and Edgar Allan Poe. In: Eco U, Sebeok TA, editors. The sign of three: Dupin, Holmes, Peirce. Indiana: Indiana University Press; 1983.
[19] Hollnagel E, Woods D, Leveson N. Resilience engineering: concepts and precepts. United Kingdom: Ashgate; 2006.
[20] Hopkins A. Issues in safety science. Safe Sci 2014;67:6–14.
[21] Joshi NN, Lambert JH. Diversification of engineering infrastructure investments for emergent and unknown non-systematic risks. J Risk Res 2011;14(4):1466–4461.
[22] Kaplan S, Visnepolschi S, Zlotin B, Zusman A. New tools for failure and risk analysis: anticipatory failure determination (AFD) and the theory of scenario structuring. Southfield, MI: Ideation International Inc.; 1999.
[23] Lambert JH, Schulte BL, Joshi NN. Multiple criteria intelligence tracking for detecting extremes from sequences of risk incidents. J Ind Manage Optim 2008;4(3):511–33.
[24] Le Coze J-C. Outlines of a sensitising model for industrial safety assessment. Safe Sci 2013;51:187–201.
[25] Lewis CI. Mind and the world order: outline of a theory of knowledge. New York: Dover Publications; 1929.
[26] Linkov I, Satterstrom F, Kiker G, Batchelor C, Bridges T, Ferguson E. From comparative risk assessment to multi-criteria decision analysis and adaptive management: Recent developments and applications. Environ Int 2006;32:1072–93.
[27] Masys AJ. Black swans to grey swans: revealing the uncertainty. Disaster Prev Manage 2012;21(3):320–35.
[28] Meyer T Reniers G. (2013) Engineering risk management. De Gruyter Graduate: Berlin.
[29] Munsterhjelm-Ahumada K. Health authorities now admit severe side effects of vaccination swine flu, pandemrix and narcolepsy, orthomolecular medicine news service, March 20, 〈http://orthomolecular.org/resources/omns/v08n10.shtml〉; 2012 [accessed 11.01.14].
[30] NC. (2011) National commission on the BP deepwater horizon oil spill and offshore drilling: deepwater. The Gulf oil disaster and the future of offshore drilling. Report to the President 11.01.11.
[31] Paté-Cornell ME. On black swans and perfect storms: risk analysis and management when statistics are not enough. Risk Anal 2012;32(11):1823–33.
[32] Pettersen KA. Acknowledging the role of abductive thinking: a way out of proceduralization for safety management and oversight? In: Bieder C, Bourrier M, editors. Trapping safety into rules: how desirable or avoidable is proceduralization? Surrey, UK: Ashgate; 2013. p. 107–17.
[33] Renn O. Risk governance. Coping with uncertainty in a complex world. London: Earthscan; 2008.
[34] Røed W, Flage R, Aven T. (2014) A new method for hazard identification based on a specific feature characterisation. Paper to be submitted for possible publication.
[35] Senge P. The fifth discipline: the art and practice of the learning organization. New York: Doubleday Currency; 1990.
[36] Shewhart WA. Economic control of quality of manufactured product. New York: Van Nostrand; 1931.
[37] Shewhart WA. Statistical method from the viewpoint of quality control. Washington, DC: Dover Publications; 1939.
[38] Smithells RW, Newman CGH. Recognition of thalidomide defects. J Med Genet 1992;29:217–23.
[39] Sternberg RJ. Handbook of creative thinking. Cambridge: Cambridge University Press; 1999.
[40] Taleb NN. The black swan: the impact of the highly improbable. London: Penguin; 2007.
[41] Taleb NN. Anti fragile. London: Penguin; 2012.
[42] Turner B, Pidgeon N. Man-made disasters. 2nd ed. London: Butterworth-Heinemann; 1997.
[43] Weick K, Sutcliffe K. Managing the unexpected: resilient performance in an age of uncertainty. Canada: Jossey Bass; 2007.
[44] Weick KE, Sutcliffe KM, Obstfeld D. Organizing for high reliability: processes of collective mindfulness. Res Organ Behav 1999;2:13–81.
[45] Zio E. An introduction to the basics of reliability and risk analysis. Singapore: World Scientific Publishing; 2007.