A discussion of the acceptable risk problem

Jørn Vatn
Department of Production and Quality Engineering
Norwegian University of Science and Technology
N–7491 Trondheim, Norway

Abstract

The petroleum activities on the Norwegian Continental Shelf are subject to regulations issued by the Norwegian Petroleum Directorate. One important issue in these regulations is the use of acceptance criteria, and this paper discusses some philosophical aspects of acceptance criteria for risk, and the role of statistical decision theory within safety management. Statistical decision theory has been applied in several studies within the nuclear industry, but has not been fully adopted within the petroleum activity. The discussion concludes by listing important measures to manage the acceptable risk problem.

1 INTRODUCTION

Almost every activity involves risk. The goal is to eliminate the risk, but it would be naive to believe that this may be completely achieved, hence we will always face the acceptable risk problem. The issue of acceptable risk has been discussed extensively in the literature, and a very good starting point is given in the book ‘‘Acceptable risk’’ by Fischhoff *et al.* [1] They claim that risk is never acceptable *unconditionally*. Risk is only acceptable if some benefit can compensate for the risk, or putting it another way: it is the *decision* yielding risk which is acceptable, not the risk in itself. Therefore they argue that the acceptable risk problem is a decision problem, i.e. a choice between alternatives. In this paper I discuss how a decision problem approach may be used in accordance with the regulations issued by the Norwegian Petroleum Directorate (NPD) [2].

A conceptual model combining the key issues is shown in Fig. 1. In the area of unacceptable risk, it is compulsory to implement means to reduce the risk to the ALARP [3, 4] region (‘‘As Low as Reasonably Practicable’’), indicated with the boldface arrow. In the ALARP region additional efforts may be used to reduce the risk further. How far to reduce the risk should be the result of a tradeoff analysis.

This paper discusses how so-called ‘‘acceptance criteria’’ may define the region of unacceptable risk and how decision analysis may be used to carry out the tradeoff analysis in the ALARP region. The issues of defining acceptance criteria and the tradeoff in the ALARP region are both *normative* issues. It is well accepted to be formal on the acceptance criteria issue, but introducing formalism on the tradeoff issue is more controversial. I will claim that up to now, the focus has been too much on the acceptance criteria issue, and too little on the explicit tradeoff analysis in the ALARP region.

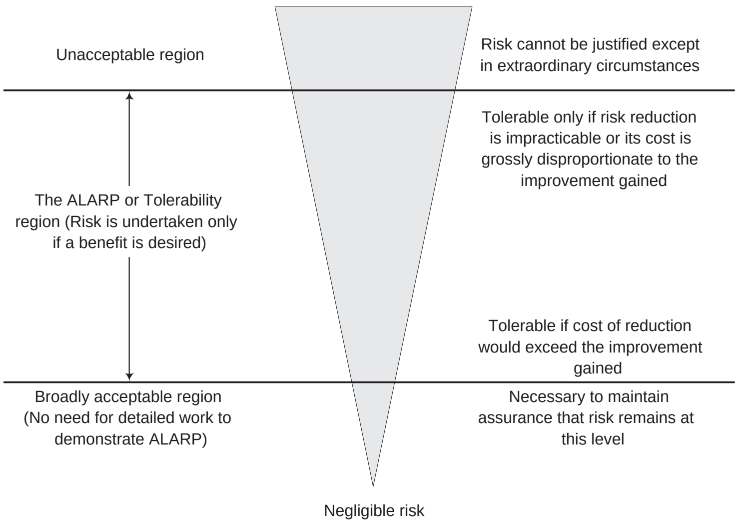

Figure 1: Levels of risk and ALARP

There are few controversies about the qualitative principles here, but the applicability of formal analyzes is often criticized, and it should be noted that international standards often take a heuristic approach [5].

Despite the fact that the theory of decision analysis has been known for several decades, successful implementations are rare. Decision problems have many aspects, and only *normative approaches* that attempt to guide the *decision maker* in making rational decisions are dealt with here. Among the normative approaches mainly *formal* analyzes are discussed.

There are many interests involved when discussing acceptable risk problems: the interests of the society, the interests of the employees, and the interests of the management in an oil company. The purpose of this paper is to focus on how management may make use of formal decision analysis. In the following therefore, the ‘‘decision maker’’ refers to decision makers in management.

1.1 Formal analysis

Formal analyzes attempt to solve decision problems by using formally defined principles of rationality. Formal analyzes are appropriate both with respect to fulfilling the acceptance criteria, and the ALARP principle. Fischhoff *et al.* [1] distinguish between *cost–benefit analysis* and *decision analysis*. In practical use these two methods might be very similar, but the perspectives are quite different. Both methods attempt to evaluate and compare the advantages and disadvantages of actions. Several steps are required:

1. *Establish preferences*. The main issue is to state what matters, and how. This involves definition of objectives and (preference) attributes and how their interrelations. Examples of attributes are numbers of persons injured, amount of gas flared, amount of deferred production etc. It is a challenge to identify the appropriate set of attributes to use, especially when knowledge regarding the long-term effects of an activity is insufficient. Cost–benefit analyzes evaluate all the attributes in *monetary* units, whereas decision analyzes evaluate the attributes in a subjective *utility* unit.
2. *Identifying actions*. The aim here is to identify actions that may have a positive effect on one

or more attributes, but in some situations one is forced to choose between actions that all have negative effects. The actions are evaluated later in the analysis.

3. *''Description of the world''*. How do the various actions affect the attributes? Probabilistic models are usually necessary.
4. *Optimizing*. The goal is to identify the actions giving the best result. For cost–benefit analyzes this means minimizing expected costs or maximizing expected profit. For decision analyzes the goal is to maximize expected utility. The optimization would be carried out subject to fulfilling the acceptance criteria.

In the Nordic Benchmark Study [6] formal decision analysis and utility theory has been used in risk decision making. The study concludes that one of the main benefits of such an approach is that facts and decision options of the case will be listed for an open discussion. Other examples of using utility theory in decision making are given in Refs ,,, and .

1.2 Probability

The term probability is frequently used in everyday language, but the exact meaning of the word is hard to capture. A variety of definitions have been proposed and pros and cons may be found for most of them. In the literature this topic has been frequently discussed [12, 13, 14, 15, 16]. The important message here is that one chooses a definition that is strictly compatible with the intended use of the word 'probability' in acceptance risk problem. For example in the Norwegian Standard NS5814: Requirements for risk analyses [17] it is stated that ''Risk shall be defined as a list of consequences and their *probability*''. The standard does not define probability. This is problematic because it leaves a lot of open questions, especially when these probabilities shall be compared with the acceptance criteria. The discussion very often polarizes, the Bayesians on one side and the frequentists on the other. However, the important thing is what you are doing and why you are doing it, it is irrelevant whether you are a Bayesian or not.

To approach some philosophical aspects of probability I find the ideas of Immanuel Kant (1724–1804) useful. Kant uses the phrases *'Das Ding an sich'* and *'Das Ding für mich'* [18] to differentiate between how the world really is (an sich) and my perception of the world (für mich). One perspective is therefore to claim that randomness does not exist *an sich* in the world, but is a consequence of the *für mich* insufficient understanding of the world. It might be argued that the advances in quantum mechanics, especially the ''Heisenberg uncertainty principle'' [19] has called for a probabilistic view of the world. However, for me it is not obvious that the world by *itself* is uncertain, but rather our *perception* of the world.

The subjective (or Bayesian) interpretation of probability is often criticized. From a scientific point of view it is hard to accept that probability does not exist in the world (an sich) but only in someone's mind (für mich). To place probability outside the heads of the individuals, I find a *''für uns''* perception of probability fruitful. In the Bayesian approach to expert judgment the idea is to incorporate evidence into ''our body of knowledge and beliefs'' [20]. The term 'our' could here be interpreted as those persons who accept the assumption made explicit in the analysis, and not one specific person. A ''für uns'' interpretation will be used in this manner in the following.

1.3 Definition of risk

Risk can be defined in many ways, and the definition to choose depends on the circumstances. Some preliminary definitions are needed before risk can be defined:

Accidental event (AE) An uncontrolled event that may lead to loss of human life, personal injury, damage to the environment and loss of assets and financial interests.[2] The word ‘hazard’[3] is also used in this context. In the NPD regulations [2] an accidental event represents the *initiating event*, and does not represent the actual damage that might follow.

Consequence (C) A possible result of an accidental event [17]. There may be several consequences of an accidental event, e.g., loss of human life, environmental pollution etc. Consequences may be described in terms of occurrence of events, or the magnitude of measurable quantities.

Focused undesired event (FUE) A specified event in the consequence spectrum of an accidental event. The course of events after the initiating event is often considered as an infinite sequence, and the FUEs are appropriate events representing actual damage that we are focusing on. Examples of focused undesired events are ‘‘release of more than x tons of oil’’, ‘‘one or more fatalities of an accident’’ and ‘‘material damage cost higher than \$ y’’. It is believed that it is easier to define acceptance criteria separate for each dimension, hence each FUE should also as far as possible represent only one dimension.

Below two principles for defining risk are given:

1. Risk is defined relative to *accidental events*. The accidental events are ‘‘pinch–points’’ [21] in the risk analysis process. The risk associated with an accidental event is a combination of the probability, or frequency of the accidental event and the magnitude of the consequences of the occurrence. The consequences represent different dimensions, e.g. loss of life, environment pollution etc., hence risk involves a consequence spectrum and is therefore multidimensional. To define risk, first let $\Pr(AE_i)$ denote the probability that accidental event number i occurs in a specified period of time. (If AE_i may occur more than once in the actual period of time, the occurrence *frequency* of AE_i would be more informative. Such an approach is not pursued in this paper.) Given that AE_i has occurred the consequences are uncertain, and thus described by a joint probability density function $\pi(C_1, C_2, \ldots \mid AE_i)$. The risk of accidental event i is then defined by

$$R_i = \Pr(AE_i) \otimes [\pi(C_1, C_2, \ldots \mid AE_i) \qquad (1)$$

   where $\otimes$ is a multiplicative operator. In practical applications it is common do derive the marginal probability density functions, $\pi_j(C_j \mid AE_i)$, for each consequences yielding:

$$R_i = \Pr(AE_i) \otimes [\pi_1(C_1 \mid AE_i), \pi_2(C_2 \mid AE_i), \ldots]$$

   To describe the entire risk, the risk associated with each accidental event should be summed up for all accidental events. Quantifying the risk into one single number is not recommended because this will require a weighting procedure which necessarily will involve value statements. Value statements should not be integrated into the definition of risk, but used when defining acceptance criteria.

2. The Norwegian standard, NS5814 [17] states that ‘‘Risk shall be defined as a list of consequences and their probability’’. Here risk is defined *independent* of the actual accidental events leading to the consequences of concern. In this case risk is defined by $[(p_1, C_1), (p_2, C_2), \ldots]$, where $p_j = Pr(C_j)$ is the occurrence probability of consequence C_j. In this second definition it is natural to think of the consequences as events, whereas in the first definition the consequences typically represent random variables.

The first principle is most appropriate when conducting a risk analysis because the definition comprises the basic elements needed in the analysis. The second definition is most appropriate when acceptance criteria are to be defined, and in tradeoff analyzes. Since the FUEs represents the consequences we are focusing on, I will claim that one should restrict oneself to only consider the focused undesired events when acceptance criteria are defined. Then the risk spectrum is defined by $[(p_1, \mathrm{FUE}_1), (p_2, \mathrm{FUE}_2), \ldots]$, where $p_j = \Pr(\mathrm{FUE}_j)$ is the occurrence probability of the focused undesired event j.

There are situations where the type and magnitude of the consequences are beyond our comprehension. This means that we are not able to state relevant FUEs, or give reasonable probability statements about the FUEs. The definition of risk under such circumstances is far from trivial, see Ref. for a discussion.

1.4 Modeling aspects

In this section it is argued that there are three stages of modeling (abstraction of the world – playing around with the model – prediction of the world). The nature of these stages are necessarily not the same, hence the methodology to apply and especially the interpretation of probability differ. The first stage is a mapping from the real world into low level models, e.g. component life time distributions and on demand probabilities. The second stage is the *intermediate* modeling, where the goal is to represent our understanding of the world into a mathematical model on a higher level, e.g. a fault tree. The input to such models is *parameters* on component level, and the output is a probability statement on system level. Usually it is appropriate to first carry out the intermediate modeling for *fixed* values of the parameters, and thereafter to derive an uncertainty distribution on the top level probability statement from the parameter uncertainty and from any model uncertainties. The third type is the *ultimate* modeling, where the goal is to make statements about the real world.

It should be emphasized that an uncertainty distribution on the TOP–event *probability* can be maintained on the intermediate level in the same manner as can be done for e.g. ‘‘$\Pr(X > 7)$’’ in a binomial model when the input parameter p is uncertain. However, when making ultimate statements about *real world* or *observable* events, the uncertainty should be integrated out to yield one probability number. Such a number is not clouded by uncertainties and therefore suited for decision making in accordance with the thoughts of Aven and Pörn [16]. Even if the integration averages out assumptions, this information will not be lost if the assumptions are stated as a part of the analysis.

2 GENERAL APPROACH

In this section the general setting of an acceptable risk problem is discussed, introducing the concept of acceptance criteria and the ALARP principle.

2.1 Use of acceptance criteria and the ALARP principle

In the Norwegian sector of the North Sea, all operators have to define risk acceptance criteria according to requirements by the NPD [2]. ‘‘Acceptable risk’’ problems are solved by combining the acceptance criteria with the so–called ALARP principle [23, 24]. It is compulsory for the companies to define values for *unacceptable* probabilities (p_i^A) of certain (focused) undesired events (FUE_i), corresponding to the upper horizontal line in Fig. 1. The situation is considered unacceptable if a Quantitative Risk Assessment (QRA) reveals a higher probability $p_i = Pr(\text{FUE}_i)$ than the limit p_i^A, and risk reduction means are compulsory. In this presentation only quantitative acceptance criteria are discussed although the NPD regulations give an opening to use qualitative acceptance criteria when appropriate.

The region where the probability is lower than the limit is denoted the ‘‘ALARP’’ region ($p_i < p_i^A$), see Fig. 1. In this region the company shall make effort to reduce risk further. In the ALARP region *cost–benefit analysis* or *decision analysis* methods are appropriate, i.e. a tradeoff analysis may determine the desired level of p_i as long as $p_i < p_i^A$.

It should be noted that the term ‘ALARP’ is not explicitly used in the requirements issued by NPD [2] where the more vague expression ‘‘the level on risk in the activities must at all times be kept as low as possible’’ is used, but there are reasons to believe that the ALARP principle will be used more explicitly in future revisions. The British Safety Case Regulations [4] uses the expression ‘‘…and measures taken to reduce risks to persons to as low as reasonably practicable’’, and a very explicit use of the ALARP principle is found in the Guidelines on Risk Issues [3].

2.2 How to treat probability in this context?

The numbers $p_i = \text{Pr}(\text{FUE}_i)$ represent the degree of belief about whether FUE_i will occur or not. In this context it does not make sense to put uncertainty on the value of p_i, what is uncertain is whether FUE_i will occur or not as described by the single ‘‘number’’ p_i (see a thorough discussion in Ref. .) During the QRA process several models and sets of input parameter values may be considered. These are denoted *models of the world* [12]. By putting weights to these models of the world, we may deduce a probability distribution, $\pi_i(\cdot)$ for the ‘‘parameter’’ p_i. This is our *intermediate* model of the world. To make an ultimate statement about the world in terms of a probability statement, the *für uns* probability for FUE_i is

$$p_i = \int_0^1 p \cdot \pi_i(p)dp \qquad (2)$$

A challenging question is thus whether an action is considered unacceptable if a significant mass of the distribution $\pi_i(\cdot)$ is above p_i^A. This is in fact an irrelevant issue, as long as if $p_i = \int_0^1 p \cdot \pi_i(p)dp < p_i^A$, then the *für uns* probability for FUE_i is sufficiently low to accept the action.

2.3 To which undesired events should acceptance criteria be defined?

There may be many (focused) undesired events that are considered so serious that one would like to define acceptance criteria for them. The idea in the NPD regulations is that the operator should decide which event acceptance criteria should be defined for, hence the operator himself has to take the responsibility. It may seem strange to develop acceptance criteria for *random variables*.

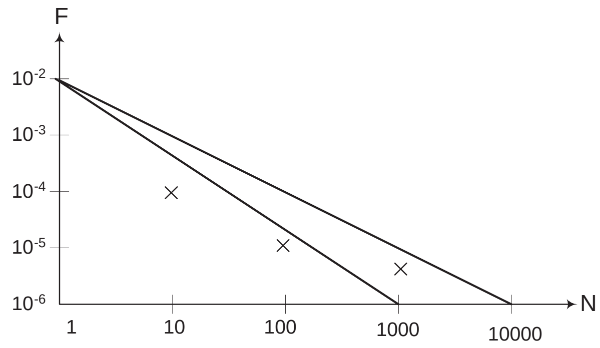

Figure 2: Frequency (F) of N or more fatalities per year

One way would be to define acceptance criteria for the expected value. However, it will usually be the extreme values that we care about. When the outcome is a random variable, it would be more natural to define acceptance criteria in terms of percentiles in the distribution.

It is obvious that sub–optimization is likely if many acceptance criteria are defined, see Ref. chapter 7 for a thorough discussion. Aven advocates that acceptance criteria with respect to loss of life should only be defined for *individual* risk, and the probability for so–called *gross accidents*. In special cases we might also define acceptance criteria to events that do not contribute much to the total risk, but that are considered ‘‘worse’’ than the actual number of fatalities, e.g. ‘‘3 persons trapped in a blocked escape route’’. Reasonable examples of acceptance criteria are therefore:

1. No person should have a higher probability than 10^{-x} of being killed during one year. It has been argued[26] to take $x = 3$. The argument is that this is the order of risk most people are exposed to, hence such a risk can not be totally unacceptable.
2. The probability of a gross accident (five or more persons killed) should be less than 10^{-y} per year.
3. The probability that all escape routes from central positions on the platform are blocked within one hour if an accident occurs, should be less than 10^{-z}.
4. The probability that ‘‘a man over board’’ cannot be salvaged by a supply boat should be less than 10^{-w}.

Only threat to life has been discussed here, but it is also relevant to define acceptance criteria for injuries, environmental risk, material losses etc. Aven [25] argues that acceptance criteria should not be defined for production regularity as they would unnecessarily narrow the set of possible solutions. It would not be possible to give general rules for which undesired events acceptance criteria should be defined. It is, however, my opinion that acceptance criteria often are defined rather arbitrarily and do not reflect real preferences, and it would often be advantageous to redefine some of these acceptance criteria to explicitly stated tradeoffs.

2.4 F-N curves

F-N curves [27] are often used to demonstrate the relation between small accidents and larger accidents with lower frequency, see Figure 2. The two straight lines in Figure 2 represent two sets of ‘‘acceptance criteria’’. The area above each line defines the region of unacceptable risk.

The lower line represents ‘‘risk aversion’’, showing an added aversion to large fatality events. However, the term ‘risk aversion’ as used here is not exactly the same as risk aversion used in the von Neumann & Morgenstern [28] sense, see Section 2.6.

To avoid mixing with the von Neumann & Morgenstern [28] definition of risk aversion, the term ‘large accident aversion’ may be used.

The lines in Figure 2 represent two different sets of normative statements, whereas the crosses represent the identified risk associated with a specific installation (descriptive part).

2.5 Decision (tradeoff) analysis

The main purpose of the decision analysis is 1) to define tradeoffs between safety, environmental protection, production regularity and costs, 2) identify and evaluate actions, and 3) choose those actions giving the best result. The acceptance criteria defined earlier will put constraints on this optimization model. Decision analysis may thus be applicable both with respect to fulfilling the acceptance criteria, and the tradeoff analysis in the ALARP region. The area of negligible risk in Fig. 1 will not be considered explicitly; it is not considered problematic if a cost effective means move us to the area of negligible risk. Whereas the acceptance criteria are defined relative to *events*, it would be natural to use *random variables* as the basis for the modeling in the ALARP region.

To carry out the analysis a set of attributes (random variables) are defined for safety, environmental protection, production regularity and costs. Principles for this procedure are given by e.g. Keeney & Raiffa [29]. Further discussion and examples are given in e.g. Refs. ,, and . Usually several attributes are defined within each of the four categories above. For the safety aspects a starting point is to introduce the following attributes

- x_1 = Number of fatalities per year
- x_2 = Number of injuries per year
- x_3 = Number of gross accidents per year

Using these three attributes will imply that risk–reducing measures will be implemented where it is cheapest to reduce risk. It may be argued that one should take more effort to reduce risk for people having the highest risk, even if this is more expensive. To account for this, additional attributes are needed, e.g. x_4 = number of fatalities per year for the most risk exposed personnel category.

Note that variable x_1 above does not discriminate between the situation of y one–fatality accidents and one accident with y fatalities. Variable x_3 was therefore introduced to make special attention to the gross accidents, cf. the F-N curves in Section 2.4.

When the attributes are defined they are combined into a utility function. The multi-attribute utility function combines the various performance measures in a way that accounts for value tradeoffs and attitudes toward risk. The principal issues of constructing the utility function are given in Ref. . However, this is not an easy task. The important ‘‘value of life’’ issue is discussed in Section 2.7. Other issues also arise, e.g. what is the total (negative) utility of hydrocarbon release; it is insufficient only to include ‘‘clean up cost’’. Penalties from the regulators, weakened opinion of the company, idealism of the management etc. should also be considered.

When the utility function has been identified, the various actions are evaluated. A variety of tools for safety and reliability modeling is available. In addition to the safety and reliability modeling, cost models are also required to optimize the expected utility under the constraints defined by the acceptance criteria. A high skill of modeling knowledge is required to carry out the optimization. As pointed out by Ref. , there is generally a lack of this knowledge, thus there are still many mountains to climb.

In this presentation the *utility* function has accounted both for value tradeoff and attitudes toward risk. This approach is essential when the values of the attributes may not be determined in advance for each action. The situation simplifies if the effect of each action is known in advance, and a *value* function is sufficient, i.e. accounting only for value tradeoff [29].

There are many objections against the ALARP principle as well as against decision analysis. One objection deals with the time aspect; consider an old installation that either is going to be scrapped, or sold within a period of two years. From the point of view of the operator, implementing safety measures would not seem cost effective due to the short time horizon. If safety measures are not implemented this seems unfair to the employees on that installation. In this situation decision analysis will prove to be a better tool than cost–benefit analysis. In cost–benefit analysis the cost of losing a human life is usually given explicitly, and independent of the circumstances. In this example this means that it would be more efficient for the operator to spend his money elsewhere if the return is higher in terms of reduced risk. Within the framework of decision analysis, the object function is a subjective utility function, and the decision maker is free to introduce those attributes necessary to describe his preferences. In the above example he may introduce an attribute similar to x_4 above, where x_4 has a higher ‘‘weight’’ in the utility function. This means that when used properly, decision analysis is an efficient tool for maintaining any preference setting. It should also be noted that tradeoff analysis in the ALARP region can never override fulfillment of the acceptance criteria.

2.6 Attributes should measure quantities in the ‘‘real world’’

The attributes form the basis for the utility function which basically is the degree of satisfaction of a specific outcome in the ‘‘real world’’. The attributes should therefore also describe phenomena in the ‘‘real world’’. Examples of attributes were given in Section 2.5. Since the future value of the attributes are unknown, they are treated as random variables. Many attributes represent the *number* of rare events where $p = \Pr(X = 1)$ is the probability that the event occurs exactly *once* in a specific time interval of concern. Despite the fact that p does not represent a ‘‘real world’’ quantity, there might be situations where it is convenient to use p as an attribute. However, as shown in the next paragraph, *risk aversion* in the von Neumann & Morgenstern [28] sense is undefined for p as an attribute.

Here risk aversion means that the decision maker prefers the expected outcome $E_\xi(X)$ rather than achieving a random variable drawn from the distribution ξ, where X is an outcome in the ‘‘real world’’. In mathematical terms this means that the utility function is *concave*. The probability density functions of X are also drawn for the two actions A and B. The expected value is higher for action B compared to action A, but the uncertainty is also greater. If it is important to avoid a very low value, action A is preferred because it is unlikely to achieve low values for this action.

The following example illustrates what happens if p is used as an attribute. The *probability* that an oil platform is able to withstand the centennial wave is denoted p. Care must here be taken regarding what is meant by probability. A ‘‘frequency’’ interpretation of p is denoted p_F. The ‘‘true’’ value of p is then unknown, and described by a probability distribution $\pi(\cdot)$ as in Section 2.2. Further let $p_U = \int_0^1 p\pi(p)dp$ represent the ultimate, or *für uns* probability for withstanding the centennial wave. Since p_U is not a random variable risk aversion cannot be defined for p_U. However, p_F is a random variable for which risk aversion may be defined.

Now, assume that design A represents known technology with relatively low uncertainty compared to the new design B, i.e. $Var_{\pi_A}(p_F) < Var_{\pi_B}(p_F)$. In this case risk aversion means that design A might be preferred even if $E_{\pi_A}(p_F) > E_{\pi_B}(p_F)$. Since $E_\pi(p_F) = p_U$, the result of such an approach is that the design with the *highest* (für uns) probability is preferred. This is unreasonable, and risk aversion therefore does not make sense for probabilities as attributes.

2.7 The value of life

A key issue in formal analysis is to place a value on the loss of human life. There is no universal agreement on how to value lives. I will argue that there are at least two different perspectives; that of the individual and that of the decision maker. Obviously there may be conflicting interests between the individual and the decision maker. Conditional influence diagrams [31] may be used to shed light on this problem. The problem may be reduced if the decision maker takes the individual’s perspective as a baseline for his decision; but what this really means is far from obvious.

A common approach in cost–benefit analysis is to say that the value of a persons’s life equals the amount of money one would need to invest to earn the income that he or she would have earned. This approach may be criticized from several points of view [1, 32] and will not be pursued further here.

Howard *et al.* [33] argues that the appropriate concern is one’s value to *oneself*, i.e. an individual perspective. He found that it is reasonable to put an infinite value on one’s life when the chances of dying is large, e.g. 80%. When the risk is low, e.g. in the order of 0.1% a ‘‘small–risk value of life’’ is in the range of $1 million to $4 million. This approach is also denoted ‘‘The changes-in-risk-of-death approach’’ [32]; a given change in risk of death, ΔP, is valued in terms of what individuals are willing to pay, ΔW, to achieve this risk reduction. The ‘‘small–risk value of life’’ is given by $\Delta W/\Delta P$. Due to the non–linear slope of the indifference curve [29] it would not be meaningful for the *individual* to extrapolate this ‘‘small–risk value of life’’ to a ‘‘high–risk value of life’’. At least two arguments may be raised against a utility approach for the individual here; (1) the risk attribute used in such an approach is essentially a probability that is not a ‘‘real world’’ attribute, and risk aversion is not adequate as concluded in Section 2.6, (2) the outcome ’’Loss of one’s life’’ has an infinite negative utility which contradicts with the basic assumptions in the axiomatic utility theory. A philosophical argument may be used in favor of a utility approach: The individual as a decision maker does not consider the risk as a *probability* but rather as *fear*, and fear *can* be claimed to be a ‘‘real world’’ attribute for which risk aversion may be assessed. A further discussion here is left to psychologists. My conclusion on this issue is that for the individual it would be strange to define risk aversion towards risk, hence a simple *value* function is adequate, and the ratio $\Delta W/\Delta P$ is sufficient to determine the value function. For this value function, ‘‘probability’’ is used as an ‘‘attribute’’.

From the decision maker’s perspective a utility approach is still adequate, even if some problems are encountered. Basically the decision maker has to (1) ensure that the acceptance criteria are met with respect to the individual’s risk and (2) decide the risk level by a tradeoff analysis. One approach is to use the individual’s ‘‘small–risk value of life’’, $\Delta W/\Delta P$ as weight in the overall utility (tradeoff) function. In such an approach the ‘‘small–risk value of life’’ is usually referred to as the value of a *statistical life*. It is an open question whether the individual's and the decision maker's perspective here really coincide; for the individual it does not make sense to extrapolate $\Delta W/\Delta P$ outside the ''small–risk'' area, but one may claim that the decision maker does this.

The discussion so far has not focused on the numbers, but rather the principles. The utility function represents preferences of the decision maker, who may take the approach above; the tradeoff should reflect the tradeoff of the individuals. This implies that he needs to do some analysis to reveal the individual's preferences. Regarding the numbers, a survey [10] of values being used in various projects, shows a range from $1 million to $10 million. This survey is a part of a methodology description where ''the LIPS value of a public life'' is set to $5.5 million [10]. In a study of managing nuclear waste a value of $4 million is used [9]. Practice among oil companies in the North Sea is to use values up to $10 million.

2.8 Taking it further

Formal analysis methods have been available for several decades, and the very well founded approach to acceptable risk problems suggested by Fischhoff *et al.* [1] has been available since 1981. Despite this, the ideas are far from implemented in the offshore industry. It is believed that the theory and the ideas have a future and hope that this paper will accelerate the process. A major problem with decision analysis as well as operations research and other cost benefit analyses is that they are so complex that the message is incomprehensible to most people. The challenge is therefore to simplify as much as possible without losing the important aspects.

As argued by Fischhoff *et al.* [1] decision theory is preferred relative to cost–benefit analysis because they do not explicitly state the ''value of life''. The message is as stated in Ref. : ''The value tradeoff K_8 is 4. This means that for the purposes of comparing the sites, the policy judgment was made that the impact of one additional statistical public fatality due to a transportation accident was just as undesirable as an additional $4 million in construction costs.''

To ease the communication between the decision maker, the involved parties and the analyst the use of *influence diagrams* [34, 8, 7] has been introduced in recent years to visualize important aspects with the decision problems. One advantage of decision analysis is it's ability to separate the facts, i.e. what we believe, and the preferences during the analysis. The influence diagrams are suitable tools for maintaining this segregation during the process.

To make decision analysis available as an efficient tool, simplification will usually be required. With respect to establishing the multi-attribute utility function this means that we should assume an additive utility function for the next ten years as a baseline, as it has been up to now [11, 9, 7]. A first approach to model risk aversion is to use the following form of the utility function $u(x) = -e^{-cx}$, where c is a positive constant. If the (uni-dimensional) pay–off X can be treated as a normal distributed random quantity with mean μ and variance σ^2 it may be shown that the expected utility is given by:

$$E(u(X)) = -e^{-c\mu+c^2\sigma^2/2} = -e^{-c\mu}e^{c^2\sigma^2/2} \quad (3)$$

The score on the expected utility increases for increased expectation, but a penalty is achieved for actions with associated large uncertainty. This is in accordance with the intuitive understanding of risk aversion.

It is beyond the scope of this paper to give a complete guideline for solving the ''acceptable risk'' problem, but the following steps should be considered:

1. Start with standard boundary definition and assumption statements.
2. Introduce influence diagrams as a communicating tool during the entire process.
3. Identify conditions for which acceptance criteria are adequate and assign values to the acceptance criteria. Usually these criteria are related to the *probabilities* of undesired *events*. The acceptance criteria are considered as constraints in the optimization.
4. Translate and add regulations and laws issued by the authorities to the list of constraints.
5. Identify attributes within the classes safety, environmental protection, production regularity and costs. The attributes are *random variables* in the real world.
6. Put relative weights on the attributes within each class, thus forming *sub*–utility functions. In the first place an additive utility function may be used even if not verified.
7. If risk aversion is relevant for some attributes, fit a simple risk averse utility function like $u(x) = -e^{-cx}$.
8. Anchor each class to a common utility unit by tradeoff analysis. Then add the sub–utility functions into one overall utility function.
9. Identify actions having a positive effect on one or more attributes.
10. Model the effect of actions on the various attributes by appropriate tools, e.g. fault tree.
11. Find the best set of actions, i.e. maximize expected utility under given constraints to achieve the best set of actions.

Note that the normative aspects are addressed in several steps. Acceptance criteria are defined first. These are normative both with respect to which event to include, and the unacceptable probabilities. The set of attributes is chosen next, and a utility function is assigned to them. This is also normative both with respect to which variables to consider, and the tradeoffs between them. In step 10 modeling knowledge and tools are required. Generally there is a lack of these as pointed out by Dekker [30], who has set up a general framework to familiarize appropriate models.[35] Vaurio[36] has also proposed a rather generalized model for modeling effect of maintenance together with fault tree analysis. Such a modeling framework has also been developed by Vatn *et al.* [7].

3 SUMMARY AND CONCLUSIONS

In this paper the view is that risk is never acceptable unconditionally. It is only *actions* that are acceptable if some benefit can be achieved. Therefore I agree with Fischhoff *et al.* [1] who state that the acceptable risk problem should be considered as a decision analysis problem. A framework has been sketched, where decision analysis may be used in accordance with regulations issued by NPD.[2]

In the NPD regulations focus is currently more on the acceptance criteria than on the tradeoff analysis in the ALARP region. In this paper I have claimed that to take full advantage of decision theory, the focus should be moved towards the tradeoff analysis. This means that one should be restrictive with the number of acceptance criteria to introduce, and rather express the normative issues in terms of value tradeoffs appropriate for an overall optimization.

Pragmatic solutions should not be rejected to achieve progress in implementing decision analysis as a decision tool. Decision analysis has a sound axiomatic basis and is therefore a preferred framework. However, one should realize that decision analysis does not make things happen by themselves, thus formal analysis is a required, but not a sufficient, tool for treating the acceptable risk problem.

4 ACKNOWLEDGMENT

This paper is written as a part of my Ph.D at the Norwegian University of Science and Technology. In particular I want to thank my supervisor, Professor Marvin Rausand for useful suggestions. Valuable comments from the anonymous referees are also acknowledged.

References

[1] B. Fischhoff, S. Lichtenstein, P. Slovic, S.L. Derby, and R.L. Keeney. *Acceptable Risk*. Cambridge University Press, New York, 1981.

[2] NPD. *Regulations concerning implementation and use of risk analyses in the petroleum activities*. Norwegian Petroleum Directorate, P.O.Box 600, N–4001 Stavanger, Norway, 1991.

[3] The Engineering Council, 10 Maltravers Street, London WC2R3ER, England. *Guidelines on Risk Issues*, 1993.

[4] HSE Information Centre, Broad Lane, Sheffield S3 7HQ, England. *A guide to the Offshore Installations (Safety Case) Regulations 1992*, 1992.

[5] IEC 61508. *Functional safety: safety-related systems*. International Electrotechnical Commission, Geneva, 1994.

[6] J. Holmberg, U. Pulkkinen, K. Pörn, and K. Shen. Risk Decision Making in Operational Safety Management – Experience from the Nordic Benchmark Study. *Risk Analysis*, 14(6):983–991, 1994.

[7] J. Vatn, P. Hokstad, and L. Bodsberg. An overall model for maintenance optimization. *Reliability Engineering and System Safety*, 51:241–257, 1996.

[8] J. Vatn. Maintenance Optimization from a Decision Theoretical Point of View. In *Proceedings, ESREL'95*, pages 273–285, London, 1995. Chameleon Press Limited.

[9] R.L. Keeney and D. von Winterfeldt. Managing Nuclear Waste from Power Plants. *Risk Analysis*, 14(1):107–130, 1994.

[10] LIPS. Laboratory integration prioritization system, lips (draft). Technical report, The LIPS Committee for Utilization and Technical Evalutation, February 1994.

[11] M.W. Merkhofer and R. L. Keeney. A Multiattribute Utility Analysis of Alternative Sites for the Disposal of Nuclear Waste. *Risk Analysis*, 7:173–194, 1987.

[12] G. Apostolakis. The concept of probability in safety assessment of technological systems. *Science*, 250:1359–64, 1990.

[13] J. K. Vaurio. On the meaning of probability and frequency. *Reliability Engineering and System Safety*, 28:121–130, 1990.

[14] S. R. Watson. The meaning of probability in probabilistic safety analysis. *Reliability Engineering and System Safety*, 45:261–269, 1994.

[15] T. W. Yellman and T. M. Murray. Comment on 'The meaning of probability in probabilistic safety analysis'. *Reliability Engineering and System Safety*, 49:201–205, 1995.

[16] T. Aven and K. Pörn. How should we express and interpret the results of quantitative risk analyses? *Reliability Engineering and System Safety*, xx:yy–zz, 1996.

[17] Norwegian association of standardisation, Po.Bo 7020, 0306 Oslo, Norway. *Norwegian Standard 5814. Risk Analysis Requirements*, 1991.

[18] I. Kant. *Immanual Kant's Critique of pure reason*. St. Martin's Press, 1963. Translated by Norman Kemp Smith.

[19] E. Kreyszig. *Introductory functional analysis with applications*. John Wiley & Sons, 1978. pp 580–581.

[20] G. Apostolakis. Expert Judgment in Probabilistic Safety Assessment. In C.A. Clarotti and D.V. Lindley, editors, *Accelerated Life Testing and Experts' Opinions in Reliability*. North–Holland, 1988. Proceedings of the international school of physics 'Enrico Fermi'.

[21] S. Kaplan. *Risk Assessment and Risk Management - Basic Concepts and Terminology*. Hemisphere Publ. Corp., Boston, Massachusetts, USA, 1991. In Risk Management: Expanding Horizons in Nuclear Power and Other industries, pp. 11–28.

[22] J. Elster. *Studies in Rationality and Social Change - Explaining Technical Change*. Cambridge University Press/Universitetsforlaget, Cambridge/Oslo, 1983.

[23] T. Aven. On safety management in the petroleum activities on the Norwegian Continental Shelf. *Reliability Engineering and System Safety*, 45:285–291, 1994.

[24] U. Kjellén and S. Sklet. Integrating analyses of the risk of occupational accidents into the design process. Part I: A review of types of acceptance criteria and risk analysis methods. *Safety Science*, 18:215–227, 1995.

[25] T. Aven. *Reliability and Risk Analysis*. Elsevier Science Publishers, London, 1992.

[26] F. Warner. Risk Assessment - A Study Group Report. Technical report, The Royal Society, 6 Carlton House Terrace, London, SW1Y 5AG, 1983.

[27] G. Ballard. Guest editoral: Societal risk–progress since Farmer. *Reliability Engeineering and System Safety*, 39:123–127, 1993.

[28] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behaviour*. Princeton University Press, Princeton, N.J, 2nd edition, 1947.

[29] R. L. Keeney and H. Raiffa. *Decisions with Multiple Objectives: Preference and Value Tradeoffs*. John Wiley & Sons, New York, 1976.

[30] R. Dekker. On the use of operations research models for maintenance decision making. *Microelectronics and Reliability*, 35(9-10):1321–1331, 1995.

[31] Y. Hong and G.E. Apostolakis. Conditional influence diagrams in risk management. *Risk Analysis*, 13(625–636), 1993.

[32] K. Needleman. Methods of valuing life. In *Technological Risk*. University of Waterloo, 1982.

[33] R. A. Howard, J. E. Matheson, and K. L. Miller. The value of life and nuclear design. In *Proceedings of the Topical Meeting on Probabilistic Analysis of Nuclear Reactor Safety*, 1978. 2, IV.2-1-IV.2-9.

[34] R. D. Shachter. Evaluating influence diagrams. *Operations Research*, 34(6):871–882, 1986.

[35] R. Dekker. A General Framework for Optimisation, Priority Setting, Planning and Combining Maintenance Activities. Technical Report 9270/A, Econometric Inst. Erasmus Univ., Rotterdam, 1992.

[36] J. K. Vaurio. Optimization of test and maintenance intervals based on risk and cost. *Reliability Engineering and System Safety*, 49:23–36, 1995.