RISK AND SAFETY IN TECHNOLOGY

Sven Ove Hansson

1 INTRODUCTION

Risk is ubiquitous in technology, and safety has been a central concern of engineering as long as there have been engineers. In codes of engineering ethics, the engineer's responsibility for the safety of workers and the public is strongly emphasized.

This chapter begins with a section on the definition of key terms in technological risk and safety. After that follow two sections that describe the two major, complementary approaches to technological risk and safety: safety engineering (Section 3) and risk analysis (Section 4). The final Section 5 is devoted to ethical analysis of risk.

2 DEFINING THE KEY TERMS

Technological risk and safety is an area in which the terminology is far from well-established. The definition of key terms often differs between different branches and traditions of engineering. These differences depend largely on lack of communication between different expert communities, but there is also a normative or ideological element in the terminological confusion. Different uses of "risk" and "safety" can express different priorities concerning what hazards should be subject to preventive or mitigating measures.

In this section, six key terms will be discussed: risk, uncertainty, hazard, safety, security, and the precautionary principle.

2.1 Risk

The word "risk" has several clearly distinguishable meanings, both in technology and other social areas. This can be illustrated with statements about the risks associated with nuclear energy. First: "A reactor melt-down is the most serious risk that affects nuclear energy." Here, we use "risk" in the following sense:

1. risk = an *unwanted event* that may or may not occur.

Next, consider the following statement: "Hidden cracks in the tubing is one of the major risks in a nuclear power station." Here we use "risk" in another sense:

Handbook of the Philosophy of Science. Volume 9: Philosophy of Technology and Engineering Sciences.
Volume editor: Anthonie Meijers. General editors: Dov M. Gabbay, Paul Thagard and John Woods.

2. risk = the *cause* of an unwanted event that may or may not occur.

We can quantify the risk of a major nuclear accident for instance in the following way: "The risk of a melt-down during this reactor's life-time is less than one in 10,000." This is yet another concept of risk:

3. risk = the *probability* of an unwanted event that may or may not occur.[1]

In risk analysis, nuclear energy is often compared to other energy sources in terms of the statistically expected number of victims. We may say, for instance: "The total risk from nuclear energy is smaller than that from coal energy." Then the word is used in the following sense:

4. risk = the statistical expectation value of unwanted events that may or may not occur.

Expectation value means probability-weighted value. Hence, if 200 deep-sea divers perform an operation in which the individual risk of death is 0.1 % for each individual, then the expected number of fatalities from this operation is 200 × 0.1 % = 0.2. The risk of fatalities in this operation can then be said to be 0.2. Expectation values have the important property of being additive. Suppose that a certain operation is associated with a 1 % probability of an accident that will kill five persons, and also with a 2 % probability of another type of accident that will kill one person. Then the total expectation value is 0.01 × 5 + 0.02 × 1 = 0.07 deaths. In similar fashion, the expected number of deaths from a nuclear power plant is equal to the sum of the expectation values for each of the various types of accidents that can occur in the plant.

In decision theory, an essential distinction is drawn between decisions "under risk" and "under uncertainty". The difference is that in the former case, but not the latter, probabilities are assumed to be known. Hence we may say: "The probabilities of failure of different types of electric switches for the control room are so well known that a decision which of them to install can be classified as a decision under risk". This corresponds to the following definition:

5. risk = the fact that a decision is made under conditions of *known probabilities* ("decision under risk")

With this we have identified five common meanings of "risk". Many attempts have been made to put an end to the ambiguity of "risk" through a stipulative definition, but unfortunately these attempts have gone in different directions. In a joint book from 1981, several of the leading researchers in the field wrote that

[1] The probabilities referred to in risk analysis are usually taken to be objective probabilities (frequencies or tendencies in the real world). When the term "subjective probability" is used, it usually refers to subjective estimates of objective probabilities. (In decision theory, "subjective probabilities" can also refer to degrees of belief that are not necessarily the outcome of an attempt to estimate an objective probability.)

"[w]hat distinguishes an acceptable-risk problem from other decision problems is that at least one alternative option includes a threat to life or health among its consequences. We shall define *risk* as the existence of such threats" [Fischhoff *et al.*, 1981, p. 2]. This is close to our definition (1). In 1983, a Royal Society working group defined risk as "the probability that a particular adverse event occurs during a stated period of time, or results from a particular challenge", i.e. our definition (3) [Royal Society, 1983, p. 22]. In the same vein, the US National Research Council [1983] defined risk assessment as an assessment of the "probability that an adverse effect may occur as a result of some human activity."

At present, by far the most common technical definition of risk is our (4), namely risk as statistical expectation value. This is in fact the newest of the five meanings of risk referred to above. Although expectation values have been calculated since the 17^{th} century, the use of the term "risk" in this sense has much more recent origin. It was introduced into technological risk analysis in the influential Reactor Safety Study (WASH-1400, the Rasmussen report) from 1975 [Rechard, 1999, p. 776]. Today it is the most common technical meaning of the term "risk". The International Organization for Standardization [2002] defines risk somewhat vaguely as "the combination of the probability of an event and its consequences", where "combination" is often (but need not be) interpreted as multiplication. All the major variants of technological risk analysis are based on the identification of risk with expectation value. In addition, this concept of risk is used in related fields such as chemical risk assessment. In cost-benefit analysis, the risks that enter the analysis are expectation values. In studies of risk perception, the "subjective risk" reported by the subjects is compared to the "objective risk", which is identified with the expectation value [Hansson, 1993].

The identification of risk with expectation value requires that the severity of outcomes can be measured in numerical terms. Ideally such a measure should refer to over-all utility, in which case risk analysis becomes a branch of expected utility theory. In risk-benefit analysis, monetary values are used as proxies for utilities. In many engineering applications, the number of deaths is used, somewhat simplistically, as a measure of the severity of outcomes.

Since "risk" has been widely used in various senses for more than 300 years, it should be no surprise that attempts to reserve it for a technical concept that was introduced 30 years ago give rise to communication problems. It is advisable to be respectful to common usage. When there is a risk of misunderstanding, it is preferable to employ a technical term such as "expectation value" for the technical concept, rather than trying to eliminate the established colloquial uses of "risk".[2]

2.2 Uncertainty

Not all dangers come with probabilities assigned to them. In decision theory, the terms "risk" and "uncertainty" are used to distinguish between those that do and

[2]In the 1983 Royal Society report, the term "detriment" was proposed to denote the product of risk and harm, but this proposal never caught on.

those that do not. In one of the most influential textbooks in decision theory, the terms are defined as follows:

> We shall say that we are in the realm of decision making under:
>
> (a) *Certainty* if each action is known to lead invariably to a specific outcome (the words prospect, stimulus, alternative, etc., are also used).
>
> (b) *Risk* if each action leads to one of a set of possible specific outcomes, each outcome occurring with a known probability. The probabilities are assumed to be known to the decision maker. For example, an action might lead to this risky outcome: a reward of $10 if a 'fair' coin comes up heads, and a loss of $5 if it comes up tails. Of course, certainty is a degenerate case of risk where the probabilities are 0 and 1.
>
> (c) *Uncertainty* if either action or both has as its consequence a set of possible specific outcomes, but where the probabilities of these outcomes are completely unknown or are not even meaningful. [Luce and Raiffa, 1957, p. 13]

This usage of the key terms "risk" and "uncertainty" differs distinctly from everyday usage. In everyday conversations, we would not hesitate to call a danger a risk although there are no means to determine its probability. By uncertainty we would mean a state of mind rather than the absence of information. It is not uncommon for this difference between technical and non-technical usage to give rise to confusion.

The gambler's decisions at the roulette table are clear examples of decisions under risk, i.e. decisions with known probabilities. Given that the wheel is fair, the probabilities of various outcomes — gains and losses — are easily calculable, and thus knowable, although the gambler may not take them into account. For a clear example of a decision under uncertainty, consider instead the decision of an explorer to enter a distant part of the jungle, previously untrod by human foot. There are tigers and poisonous snakes in the jungle, but no estimates better than guesses can be given of the probability of being attacked by them. Such attacks are known dangers with unknown probabilities. In addition, the jungle may contain a large number of other species — from microorganisms to mammals — some of which may be dangerous although they are completely unknown. Not only their probabilities but also their very existence is unknown.[3]

In real life we are seldom in a situation like that at the roulette table, when all probabilities are known with certainty (or at least beyond reasonable doubt). Most of the time, we have to deal with technological dangers without knowing their probabilities, and often we cannot even foresee what dangers we will have to deal with. This is true not least in the development of new technologies. The social and environmental effects of a new technology can seldom be fully grasped

[3]Unknown dangers are not included in Luce and Raiffa's definition of uncertainty, as quoted above, but in subsequent discussions the concept of uncertainty has been extended to include them.

beforehand, and there is often considerable uncertainty with respect to the dangers that it may give rise to [Porter, 1991; Rosenberg, 1995].

There is, however, a tendency in risk studies to proceed as if decisions on technologies were made under conditions analogous to gambling at the roulette table. This has been called the *tuxedo syndrome*, and it has been argued that the metaphor of entering an unexplored jungle better describes the plight of engineers and others responsible for new technology [Hansson, 2008]. It is important to observe that even in cases when the plausibility of dangers can be meaningfully summarized in terms of probabilities, there may yet remain significant uncertainties about the accuracy of these estimates.

There is often a drift in the sense of the word "risk", in the following sense: A discussion or an analysis begins with a general phrase such as "risks in the building industry" or "risks in modern energy production". This includes both dangers for which probability estimates are already available and dangers for which it is doubtful whether meaningful probability estimates can at all be obtained. As the discussion or analysis goes more into technical detail, the term "risk" is narrowed down to refer only to that which can be quantified in probabilistic terms. In the course of this narrowing-down, the analyst often loses sight of uncertainties that should have been taken into account in an accurate analysis of the dangers at hand. To avoid the neglect of such dangers, it is important to treat uncertainties explicitly, in particular if a narrow concept of risk is used in the technical analysis.

2.3 Hazard

In many engineering contexts, a distinction is made between a risk and a hazard. A hazard can be defined as a potential risk. Hence, consider two persons who set off the same type of firework. One of them is an expert pyrotechnician, the other a drunken teenager who never lit fireworks before. The hazard is the same in both cases, namely that the device may explode close to the person who handles it. The risk (in sense 3 or 4 as defined in Section 2.1) is much smaller in the former case, provided that the pyrotechnician takes the standard precautions.

Hazard is mostly treated as a non-quantitative concept. Attempts have been made to quantify it. In such quantifications, the focus is on consequences, not on probabilities [Rahman *et al.*, 2005].

The limit between a hazard and a non-hazard is vague. If an airplane is just about to crash, and this airplane is in the air very close to you, we would say that you are exposed to a hazard. However, if the plane is flying normally at high altitude, we would not think of it as a hazard as it is passing by. More generally speaking, when the risk associated with a hazard is considered negligible, we tend not to call it a hazard at all.

2.4 Safety

The concept of safety is sometimes used in an absolute, sometimes in a relative sense. In order to illustrate the meaning of absolute safety, suppose that you buy a jacket that is promised to be of fire-proof fabric. Later, it actually catches fire. Then you might argue, if you apply the absolute notion of safety, that you were not safe from fire in the first place. If the producer of the jacket tries to argue that you in fact *were* safe since the fabric was highly unlikely to catch fire, you would say that he was simply wrong. In some contexts, therefore, "I am safe against the unwanted event X" is taken to mean that there is no risk at all that X will happen.

Technical safety has often been defined as absolute safety. For example, in research on aviation safety it has been claimed that "Safety is by definition the absence of accidents" [Tench, 1985]. However, in practice absolute safety is seldom achievable. For most purposes it is therefore not a very useful concept. Indeed, the US Supreme court has supported a non-absolute interpretation, stating that "safe is not the equivalent of 'risk free' " [Miller, 1988, p. 54]. With this interpretation, a statement such as "this building is fire-safe" can be read as a short form of the more precise statement: "The safety of this building with regard to fire is as high as can be expected in terms of reasonable costs of preventive actions." In this vein, the American department of defence has stated that safety is "the conservation of human life and its effectiveness, and the prevention of damage to items, consistent with mission requirements" [Miller, 1988, p. 54].

Usage of the term "safe" (and derivatives such as "safety") in technical applications, e.g. in aviation safety, highway safety etc, vacillates between the absolute concept ("safety means no harm"), and a relative concept that only requires the risk reduction that is considered feasible and reasonable. It is not possible to eliminate either of these usages, but it is possible to keep track of them and avoid confusing them with each other.

Safety is usually taken to be the inverse of risk: when the risk is high, then safety is low, and conversely. This may seem self-evident, but the relationship between the two concepts is complicated by the fact that, as we saw in Subsection 2.1, the concept of risk is in itself far from clear. It has been argued that if risk is taken in the technical sense as statistical expectation value (expected harm), then safety cannot be the antinomy of risk, since other factors such as uncertainty have to be taken into account when assessing safety [Möller *et al.*, 2006]. With a broader definition of risk, an antonymic relationship between the two concepts may be more plausible.

2.5 Security

Safety and security are closely related concepts. (Some languages have the same word for these two English terms, such as the German "Sicherheit".) With safety we usually mean protection against unintentional threats, and with security protection against intentional threats. Typical security issues are protection of a coun-

try against war, protection of individuals against violence, protection of property against theft, malicious or wanton destruction and economic crime, and protection of computers against unauthorized intrusion.

Many of the measures that are taken to improve safety also increase security, and vice versa. Hence, the same sprinkler system can extinguish a fire caused by arson or by an accident. Security checks that prevent unauthorized persons from entering a building with dangerous machinery contribute to preventing both accidents and sabotage, etc. However, safety and security have traditionally been treated as separate issues, and they have been delegated to different professions. In many organizations that have to deal with both types of issues, surprisingly little has been done to coordinate them.

2.6 The precautionary principle

The precautionary principle is frequently invoked in debates on environmental issues. Strictly speaking, it is a principle for decision-making under scientific uncertainty that has been codified in a number of international treaties.[4]

Formulations of the precautionary principle can be divided into two major groups: *argumentative* and *prescriptive* versions of the principle. An argumentative version of the precautionary principle is found in Principle 15 of the Rio Declaration [UNCED 1993]. It requires that "lack of full scientific certainty shall not be used as a reason for postponing cost-effective measures to prevent environmental degradation". Prescriptive versions of the precautionary principle prescribe actions. Perhaps the most famous of these is the so-called Wingspread Statement: "When an activity raises threats to the environment or human health, precautionary measures should be taken, even if some cause-and-effect relationships are not fully established scientifically" [Raffensperger and Tickner, 1999, pp. 354-355].

Most prescriptive versions of the precautionary principle share four common components [Sandin, 1999]. Consider the following possible formulation of the precautionary principle:

> It is mandatory to limit, regulate, or prevent potentially dangerous technologies even before scientific proof is established.

We find four different components in this formulation, namely

1. the *threat component*, expressed in the phrase "potentially dangerous technologies";
2. the *uncertainty component*, expressed in the phrase "even before scientific proof is established";
3. the action component, expressed in the phrase "to limit, regulate, or prevent";

[4]The phrase "precautionary principle" is often used less specifically to denote cautious decision-making in general. It is better to use other terms for this, such as "cautious decision-making".

4. the prescription component, expressed in the phrase "is mandatory".

The first two of these can be summarized as the *trigger* of the precautionary principle, whereas the last two constitute the *precautionary response* [Ahteensuu, 2008].

The uncertainty dimension ensures that action is triggered even in the absence of full scientific evidence. It is the most characteristic part of the principle. It is what distinguishes the precautionary principle from other principles or argumentation forms for the protection of health and the environment.

In summary, the precautionary principle proclaims that policy decisions in environmental decisions can legitimately be based on scientific evidence of a danger that is not strong enough to constitute full scientific proof that the danger exists [Sandin *et al.*, 2004]. However, it is not unproblematic to describe this as a special principle for environmental policies, as some sort of extra cautiousness that is presumed not to apply in other decisions. From a decision-theoretical point of view, allowing decisions to be influenced by uncertain information is not a special principle that needs to be specially defended. To the contrary, doing so is nothing else than ordinary practical rationality, as it is applied in most other contexts [Hansson, 2006a]. If there are strong scientific indications that a volcano may erupt in the next few days, decision-makers will expectedly evacuate its surroundings as soon as possible, rather than waiting for full scientific evidence that the eruption will take place. Furthermore, as we will see in the next section, traditional safety engineering is largely based on cautious thought patterns that are similar to the precautionary principle but of much older origin.

3 SAFETY ENGINEERING

Engineering is primarily a practice. For as long as humanity has used technological artefacts, we have taken measures to protect humans against various risks associated with these artefacts. Since the 19th century, many engineers have specialized in worker's safety and other safety-related tasks. With the development of technological science, the ideas behind safety engineering have been subject to academic treatments. There are now many ways to systematize the practices of safety engineering, but none of them has gained general acceptance. A major reason for this is that the discussion on safety engineering is fragmented between different areas of technology. In this section, three major principles of safety engineering will be discussed, and it will be argued that there is an underlying principle that unites them all.

3.1 Three principles of safety engineering

The following principles are in general use in many fields of engineering.

1. *Inherently safe design.* A recommended first step in safety engineering is to minimize the inherent dangers in the process as far as possible. This means

that potential hazards are excluded rather than just enclosed or otherwise coped with. Hence, dangerous substances or reactions are replaced by less dangerous ones, and this is preferred to using the dangerous substances in an encapsulated process. Fireproof materials are used instead of inflammable ones, and this is considered superior to using flammable materials but keeping temperatures low. For similar reasons, performing a reaction at low temperature and pressure is considered superior to performing it at high temperature and pressure in a vessel constructed for these conditions.

2. *Safety factors.* Constructions should be strong enough to resist loads and disturbances exceeding those that are intended. A common way to obtain such safety reserves is to employ explicitly chosen, numerical safety factors. Hence, if a safety factor of 2 is employed when building a bridge, then the bridge is calculated to resist twice the maximal load to which it will in practice be exposed.

3. *Multiple independent safety barriers.* Safety barriers are arranged in chains. The aim is to make each barrier independent of its predecessors so that if the first fails, then the second is still intact, etc. Typically the first barriers are measures to prevent an accident, after which follow barriers that limit the consequences of an accident, and finally rescue services as the last resort. One of the major lessons from the Titanic disaster is that an improvement of the early barriers (in this case: a hull divided into watertight compartments) is no excuse for reducing the later barriers (in this case: lifeboats).

Several *caveats* have to be appended to this list of safety engineering practices: The terminology is not universally accepted, and some of these principles are also known under other names. The three principles are partly over-lapping. Furthermore, safety engineering includes many more principles and practices than the three mentioned above: negative feedback mechanisms, education of operators, maintenance of equipment and installations, incidence reporting etc.

In the following subsections, each of the three principles will be treated in somewhat more detail.

3.2 Inherent safety

Inherent safety, also called primary prevention, consists in the elimination of a hazard. It is contrasted with secondary prevention that consists in reducing the risk associated with a hazard. This can be done by reducing either the probability or the consequences of an adverse event such as an accident in which the hazard is realized. For a simple example, consider a process in which inflammable materials are used. Inherent safety would consist in replacing them by non-inflammable materials. Secondary prevention would consist in removing or isolating sources of ignition and/or installing fire-extinguishing equipment. As this example shows, secondary prevention usually involves added-on safety equipment.

Traditionally, four types of safety measures are recommended in inherently safer design of plants:

- minimize (intensify): use smaller quantities of hazardous materials
- substitute: replace a hazardous material by a less hazardous one
- attenuate (moderate): use the hazardous material in a less hazardous form
- simplify: avoid unnecessary complexity in facilities and processes, in order to make operating errors less likely [Khan and Abbasi, 1998; Bollinger *et al.*, 1996].

It is important not to overstate the case of inherent safety. Many safety problems cannot be solved with inherent safety. Often, socially desirable means cannot be achieved without hazardous materials or processes. In such cases, reliance must be put on well-developed secondary prevention. However, there are also many cases with a choice between eliminating and managing the hazard. Proponents of inherent safety have shown that other things being equal, elimination is the better option. The major reason for this is that, as long as the hazard still exists, it can be realized by some unanticipated triggering event. Even with the best of control measures, some unforeseen chain of events can give rise to an accident. Even the best added-on safety technology can fail, or be destroyed in the course of an accident.

An additional argument for inherent safety is its usefulness in dealing with security threats. Add-on safety measures can often easily be deactivated by those who want to do so. When terrorists enter a chemical plant with the intent to blow it up, it does not matter much that all ignition sources have been removed from the vicinity of explosive materials (although this may perhaps have solved the safety problem). The perpetrators will bring their own ignition source. In contrast, most measures that make a plant inherently safer will also contribute to diverting terrorist threats. If the explosive substance has been replaced by a non-explosive one, or the inventories of explosive and inflammable substances have been drastically reduced, then the plant will be much less attractive to terrorists and will therefore also be a less likely target of attack.

Inherent safety has a particularly important role in the chemical industry. Most of the development of techniques for inherent safety has taken place within or in cooperation with chemical companies. The other major industry where inherent safety is often discussed is the nuclear industry. Inherent safety is often referred to in efforts to construct new, safer types of reactors. A reactor will be inherently safer than those currently in use if, even in the case of failure of all active cooling systems and complete loss of coolant, the fuel element temperatures will not be exceed the limits at which most radioactive fission products remain confined within the fuel elements [Brinkmann *et al.*, 2006].

3.3 Safety factors

Probably, humans have made use of safety reserves since the origin of our species. We have added extra strength to our houses, tools, and other constructions in order to be on the safe side. However, the use of numerical factors for dimensioning safety reserves seems to be of relatively recent origin, probably the latter half of the 19$^{\text{th}}$ century. The earliest usage of the term recorded in the Oxford English Dictionary is from W.J.M. Rankine's book *A Manual of Applied Mechanics* from 1858. In the 1860s, the German railroad engineer A. Wohler recommended a factor of 2 for tension [Randall, 1976]. The use of safety factors is now since long well established in structural mechanics and in its many applications in different engineering disciplines. Elaborate systems of safety factors have been developed, and specified in norms and standards [Clausen *et al.*, 2006].

A safety factor is typically intended to protect against a particular integrity-threatening mechanism, and different safety factors can be used against different such mechanisms. Hence one safety factor may be required for resistance to plastic deformation and another for fatigue resistance. A safety factor is most commonly expressed as the ratio between a measure of the maximal load not leading to the specified type of failure and a corresponding measure of the maximal load that is expected to be applied. In some cases it may instead be expressed as the ratio between the estimated design life and the actual service life.

In some applications safety margins are used instead of safety factors. A safety margin differs from a safety factor in being additive rather than multiplicative. In order to keep airplanes sufficiently apart in the air a safety margin in the form of a minimal distance is used. Safety margins are also used in structural engineering, for instance in geotechnical calculations of embankment reliability [Duncan, 2000].

According to standard accounts of structural mechanics, safety factors are intended to compensate for five major categories of sources of failure [Knoll, 1976; Moses, 1997]:

1. higher loads than those foreseen,
2. worse properties of the material than foreseen,
3. imperfect theory of the failure mechanism in question,
4. possibly unknown failure mechanisms, and
5. human error (e.g. in design).

The first two of these refer to the variability of loads and material properties. Such variabilities can often be expressed in terms of probability distributions. However, when it comes to the extreme ends of the distributions, lack of statistical information can make precise probabilistic analysis impossible. Let us consider the variability of the properties of materials. Experimental data on material properties are often insufficient for making a distinction between e.g. gamma and lognormal distributions, a problem called *distribution arbitrariness* [Ditlevsen, 1994]. This

has little effect on the central part of these distributions, but in the distribution tails the differences can become very large. This is a major reason why safety factors are often used as design guidance instead of probabilities, although the purpose is to protect against failure types that one would, theoretically, prefer to analyze in probabilistic terms.

> Theoretically, design by using structural system reliability is much more reasonable than that based on the safety factor. However, because of the lack of statistical data from the strength of materials used and the applied loads, design concepts based on the safety factor will still dominate for a period. [Zhu, 1993]

The last three of the five items on the list of what safety factors should protect against all refer essentially to errors in our theory and in our application of it. They are therefore clear examples of uncertainties that are not easily amenable to probabilistic treatment. In other words: The eventuality of errors in our calculations or their underpinnings is an important reason to apply safety factors. This is an uncertainty that is not reducible to probabilities that we can determine and introduce into our calculations. It is for instance difficult to see how a calculation could be accurately adjusted to compensate self-referentially for the possibility that it may itself be wrong. However, these difficulties do not make these sources of failures less important. Safety factors are used to deal both with those failures that can be accounted for in probabilistic terms and those that cannot.

3.4 Independent safety barriers

The use of multiple safety barriers is based on the simple principle that even if one measure that we take to avert a danger should fail, there should be some other measure in place that averts it.

The archetype of multiple safety barriers is an ancient fortress. If the enemy manages to pass the first wall, there are additional layers that protect the defending forces. Some engineering safety barriers follow the same principle of concentric physical barriers. Interesting examples of this can be found in nuclear waste management. The waste will be put in a copper canister that is constructed to resist the foreseeable challenges. The canister is surrounded by a layer of bentonite clay that protects the canister against small movements in the rock and "acts as a filter in the unlikely event that any radionuclides should escape from a canister".[5] This whole construction is placed in deep rock, in a geological formation that has been selected to minimize transportation to the surface of any possible leakage of radionuclides. The whole system of barriers is constructed to have a high degree of redundancy, so that if one the barriers fails the remaining ones will suffice.

With the usual standards of probabilistic risk analysis, the whole series of barriers around the waste would not be necessary. Nevertheless, sensible reasons can be given for this approach, namely reasons that refer to uncertainty. Perhaps the

[5] http://www.skb.se/templates/SKBPage_ 8762.aspx.

copper canister will fail for some unknown reason not included in the calculations. Then, hopefully, the radionuclides will stay in the bentonite, etc. In this particular case, redundancy can also be seen as a means to meet public scepticism and opposition (although it is not self-evident that redundant safety barriers will make the public feel safer).

The notion of multiple safety barriers can also refer to safety barriers that are not placed concentrically like the defence walls of a fortress, but are arranged consecutively in a functional sense. The essential feature is that the second barrier is put to work when the first one fails, etc. Consider for instance the protection of workers against a dangerous gas such as hydrogen sulphide that can leak from a chemical process. An adequate protection against this danger can be constructed as a series of barriers. The first barrier consists in constructing the whole plant in a way that excludes uncontrolled leakage as far as possible. The second barrier is careful maintenance, including regular checking of vulnerable details such as valves. The third barrier is a warning system combined with routines for evacuation of the premises in the case of a leakage. The fourth barrier is efficient and well-trained rescue services.

The basic idea behind multiple barriers is that even if the first barrier is well-constructed, it may fail, perhaps for some unforeseen reason, and that the second barrier should then provide protection. For another illustration of this principle, we can consider what is possibly the most well-known example of technological failure in modern history, the Titanic that sank with 1500 persons in April 1912. It was built with a double-bottomed hull that was divided into sixteen compartments, constructed to be watertight. Four of these could be filled with water without danger. Therefore, the ship was believed to be unsinkable, and consequently it was equipped with lifeboats only for about half of the persons onboard.

We now know that the Titanic was far from unsinkable. But let us consider a hypothetical scenario. Suppose that tomorrow a ship-builder comes up with a convincing plan for an unsinkable boat. Calculations show that the probability of the ship sinking is incredibly low and that the expected cost per life saved by the life-boats is above 1000 million dollars, a sum that can evidently be more efficiently used to save lives elsewhere.

How should the naval engineer respond to this proposal? Should she accept the verdict of the probability calculations and the economic analysis, and exclude lifeboats from the design? There are good reasons why a responsible engineer should not act in this way: The calculations may possibly be wrong, and if they are, then the outcome may be disastrous. Therefore, the additional safety barrier in the form of lifeboats (and evacuation routines and all the rest) should not be excluded. Although the calculations indicate that such measures are inefficient, these calculations are not certain enough to justify such a decision.

The major problem in the construction of safety barriers is how to make them as independent of each other as possible. If two or more barriers are sensitive to the same type of impact, then one and the same destructive force can get rid of all of them in one swoop. Hence, any number of concentric walls around a fortified

city could not protect the inhabitants against starvation when they run out of provisions. Similarly, three consecutive safety valves on the same tube may all be destroyed in a fire, or they may all be incapacitated due to the same mistake by the maintenance department. A major geologic event or an attacking enemy can destroy all the barriers of a subterranean nuclear waste repository at the same time, etc. Therefore, it is essential, when constructing a system of safety barriers, to make the barriers as independent as possible. Often, more safety is obtained with fewer but independent barriers than with many that are sensitive to the same sources of destruction.

3.5 The common trait

The principles of safety engineering discussed in this section have one important trait in common: they all aim at protecting us not only against risks (in the technical sense) but also against hazards that cannot be assigned meaningful probability estimates, such as the possibility that some unforeseen event triggers a hazard that is seemingly under control. Although explicit discussions of risk and uncertainty are new to engineering, the insights encoded in safety engineering are much older than that – indeed older than probability theory.

4 RISK ANALYSIS

In the late 1960s, rapidly growing public opposition to new technologies gave rise to a new market for applied science: a market for expertise on risks and on the public's attitudes to risks. The demand came mostly from companies and institutions associated with the technologies that had been subject to public opposition. The supply was met by professionals and academics with training in the natural, behavioural, and social sciences. Most of their undertakings focused on chemicals and on nuclear technology, the same risk factors that public opposition had targeted on. The new field was institutionalized as the discipline of risk analysis, with professional societies, research institutes, and journals of its own.

4.1 The subdisciplines of risk analysis

The short history of risk analysis can be summarized in terms of five major approaches to risk that made their appearance consecutively, and now all coexist [Otway, 1987]. The first of the five approaches was *acceptable risk*. Many of the earliest studies in the field aimed at determining a level of "risk", i.e., of the statistically expected number of fatalities, that is accepted or that should be accepted. A common procedure was to compare new technological risks to risks that are accepted in everyday life.

The next step was *risk-benefit analysis*, in which both the risks and the benefits of a technology were quantified so that they could be compared. The standard method is to assign a monetary value to all relevant outcomes (including the loss of

human lives) in order to make risks and benefits computationally comparable. This approach was largely the result of economists entering the stage of risk science.

The third step resulted in part from the growing involvement of psychologists in risk analysis. This step consisted in studies of *risk perception* that were much in vogue in the early 1980's. The ordering of risks obtained from questionnaires was said to measure "subjective risk", and was compared to the expected number of deaths that was called "objective risk". The difference was commonly conceived as a sign of irrationality or misperception.

The next approach was *risk communication* that aims at providing lay people with information that helps them to see risks in a certain way. Typically, risk communication is considered to be successful if it has made people adjust their "subjective risk" to fit in with the "objective risk".

The fifth approach was studies of *trust* that emanated from the difficulties that both public authorities and companies have encountered when trying to change the public's opinion on risk through various measures of risk communication. The problem, seen in their perspective, seems to be that the public does not have trust in the sources of information. How such trust can be achieved is currently a major theme at conferences on risk.

All the major variants of risk analysis are associated with the same formal model of risk, namely the expectation value definition of risk that was introduced above in Subsection 2.1. In other words, the common procedure is to multiply "the probability of a risk with its severity, to call that the expectation value, and to use this expectation value to compare risks" [Bondi, 1985, p. 9]. The following is a typical example of the jargon:

> The worst reactor-meltdown accident normally considered, which causes 50 000 deaths and has a probability of 10^{-8}/reactor-year, contributes only about two per cent of the average health effects of reactor accidents. [Cohen, 1985, p. 1]

In what follows, we will consider three central methodological issues in risk analysis (Subsections 4.2–4.4) and three additional such issues in risk-benefit analysis (Subsections 4.5–4.7). The final Subsection 4.8 is devoted to a comparative discussion of risk analysis and safety engineering.

4.2 Fault tree methodology

When there is statistically sufficient experience of an event-type, such as a machine failure, then we can determine its probability by collecting and analysing that experience. Hence, if we want to know the probability that the airbag in a certain make of car fails to release in a collision, we should collect statistics from the accidents in which such cars were involved.

For new and untested technologies this method is not available. Accident statistics is not at hand to determine the probability of airbag failure in a new car model that is just to be introduced. If the construction is essentially unchanged since

previous models, then we may rely on statistics from these earlier models, but this is not advisable if there have been significant changes.

Even after many years' experience of a technology there may be insufficient data to determine the frequencies of unusual types of accidents or failures. As one example of this, there have (fortunately) been too few severe accidents in nuclear reactors to make it possible to estimate their probabilities. In particular, most of the reactor types in use have never been involved in any serious accident. It is therefore not possible to determine the risk (probability) of a severe accident in a specified type of reactor.

One common way to evade these difficulties is to calculate the probability of major failures by means of a careful investigation of the various chains of events that may lead to such failures. By combining the probabilities of the subevents in such a chain, a total probability of a serious accident can be calculated. Such calculations were in vogue in the 1970s and 1980s. Today there is a growing scepticism against them, due to several difficult problems with this methodology. One such problem is that accidents can happen in more ways than we can think of beforehand. There is no method by which we can identify all chains of events that may lead to a major accident in a nuclear reactor, or any other complex technological system.

Another problem with this methodology is that the probability of a chain of events can be very difficult to determine even if we know the probabilities of each individual event. Suppose for instance that an accident will happen if two safety valves both fail. Furthermore suppose that we have experience showing that the probability is 1 in 500 that a valve of this construction will fail during a period of one year. It does *not* follow from this that the probability that both will fail in that period is $1/500 \times 1/500$, i.e. 1/250,000. The reason for this is that failures in the two valves are not independent events.

In spite of these difficulties, the construction and analysis of such event chains (often called fault-trees) is not a useless exercise. To the contrary, it can be an efficient way to identify weaknesses in a complex technological system. It is important, though, to keep in mind that an exhaustive list of negative events cannot be obtained, and that therefore total risk levels cannot be determined in this way.

4.3 Indetectable effects

Much of modern science is devoted to the study of composite systems: ecosystems, the human body, the world economy, etc. Each of these contains so many components and potential interactions that it is in practice unpredictable. Some of these systems are unpredictable not only in practice but also in principle, due to chaotic phenomena. In addition, science is always subject to another type of uncertainty, namely that of unknown factors. Only seldom do we have good reasons to believe that our scientific models are complete in the sense that no important components or interactions have been left out.

In some cases, knowledge about complex systems can be gained through systematized experience. This applies for instance to the effects of therapeutic agents on the human body. Due to the complexity of the body, it is in practice impossible to theoretically predict the effects of a new drug. Instead, after preliminary trials have been completed, medicines are tried out experimentally on groups of patients. Based on statistics from such studies (clinical trials) the effects of medical drugs can be ascertained with reasonable accuracy.

However, there are many cases in which this type of "statistical bypass" to knowledge about complex systems is not available. We do not have access to one hundred earths on which we can experiment to determine a tolerable level of greenhouse gas emissions. Furthermore, even in the cases when statistical information is available, it does not always reduce uncertainties as efficiently as one might hope.

To see this, let us consider the example of health effects of chemical substances. To what extent is it possible to determine the presence or absence of such effects through direct studies of exposed humans? Unfortunately, the answer to that question is rather disconcerting.

To simplify the discussion, let us focus on lifetime risks of lethal effects of toxic substances. To begin with, suppose that 1000 persons are all subject to a chemical exposure that gives rise to hepatic angiosarcoma (a rare cancer of the liver) among 0.5 % of the exposed. Among unexposed individuals, the frequency of this disease is very close to zero. If a proper investigation is made, chances are very high that the overrepresentation of this disease among the exposed population will be discovered.

Next, suppose that another group of 1000 persons are subject to an industrial exposure that increases the incidence of lung cancer from 10.0 to 10.5 %. The expected number of additional cancer cases is the same as in the previous case. However, as can be shown with probability calculus, the difference between 10.0 and 10.5 % is in this case indistinguishable from random variations. Hence, the effects of this substance cannot be detected in a study of the exposed population [Hansson, 1999].

As a rough rule of thumb, epidemiological studies cannot reliably detect excess relative risks if they are about 10 % or smaller. For the more common types of lethal diseases, such as coronary disease and lung cancer, lifetime risks are of the order of magnitude of about 10 %. Therefore, even in the most sensitive studies, an increase in lifetime risk of the size 10^{-2} (10 % of 10 %) or smaller may be indetectable (i.e. indistinguishable from random variations). In animal experiments we have similar experimental problems, and in addition problems of extrapolation from one species to another.

There is no objective answer to the question how small health effects should be of concern to us. However, many attempts have been made to set a limit of concern, expressed either as "acceptable risk" or "de minimis risk". Most people seem to agree that if a human population is exposed to a risk factor that will, statistically, kill one person out of 10^9, then that risk will not be an issue of high priority. Arguably, it is no big problem that our risk assessment methods are

insufficient to discover risks of that order of magnitude. On the other hand, most of us would consider it a serious problem if a risk factor that kills one person out of 100 or 1000 cannot be detected. The most common proposals for limits of concern for lethal risks are 1 in 100,000 and 1 in 1,000,000. It is difficult to find proposals above 1 in 10,000. These values are of course not objective or scientific limits; they belong to the ethical realm. However, it is important to note the presence of what has been called an ethical gap, a gap between those risk levels that are scientifically detectable and those that are commonly regarded to be ethically acceptable or at least of minor concern [Hansson, 2002]. This gap has the breadth of 2–4 orders of magnitude. Hence, even if no adverse effects have been found in exposed populations, there may still be effects at risk levels that are at least 100 to 1000 times higher than commonly proposed levels of concern or acceptability.

4.4 The unreliability of probability estimates

When probabilities cannot be estimated from empirically known frequencies, the standard method is to instead use experts' estimates of probabilities. The reliability of risk analysis will then depend on the assumption that there are no or only small systematic differences between objective probabilities and experts' estimates of these probabilities.

However, this assumption is not correct. Significant differences between such objective frequencies and expert's estimates of them are well known from experimental psychology, where they are described as lack of *calibration*. Probability estimates are (well) calibrated if "over the long run, for all propositions assigned a given probability, the proportion that is true equals the probability assigned" [Lichtenstein *et al.*, 1982, pp. 306-307]. Thus, half of the statements that a well-calibrated subject assigns probability 0.5 are true, as are 90 per cent of those that she assigns probability 0.9, etc.

Experimental studies indicate that there are only a few types of predictions that experts perform in a well-calibrated manner. Professional weather forecasters and horse-race bookmakers make well-calibrated probability estimates in their respective fields of expertise [Murphy *et al.*, 1984; Hoerl and Fallin, 1974]. In contrast, most other types of prediction that have been studied are subject to substantial overconfidence. Physicians assign too high probability values to the correctness of their diagnoses [Christensen-Szalanski and Bushyhead, 1981]. Geotechnical engineers were overconfident in their estimates of the strength of a clay foundation, etc. [Hynes and Vanmarcke, 1976].

We do not know how well calibrated the experts' estimates of probabilities are that are used in risk analysis. To the extent that they are badly calibrated, the outcome of risk analysis is correspondingly inaccurate.

4.5 Problems of incommensurability

Most of the philosophical discussion about risk-benefit analysis has been concerned with the difficulties involved in assigning an economic value to that which we conceive as invaluable, such as a human life or an animal species. Clearly, human lives do not have a monetary price in the common sense of the word. A risk-benefit analyst who assigns a monetary value to the loss of a human life does not thereby imply that someone can buy another person, or the right to kill her, for that price. In any sensible interpretation of a risk-benefit analysis, the values of lives are for calculation purposes only.

The underlying motivation for values of life is that they can be used as a means to reduce multi-dimensional decision problems to uni-dimensional ones. The common way to do this, technically, is to convert all dimensions of a decision to monetary values, even those that are incommensurable with money. The essential problem — or perhaps even dilemma — is that in order to achieve this we need to comparatively evaluate entities that we conceive as incomparable.

Defenders of risk-benefit analysis tend to emphasize that comparisons between lives and money are not unique to risk-benefit analysis. They are in fact unavoidable components of many of the decisions that we have to make in different social sectors. We could, for instance, always spend more money than we do on traffic safety, taking the resources from activities that do not save lives. Our decision not to spend more than we do contains an implicit value of life; we do not pay to save more lives than we do since it would be too costly. The problem does not come with risk-benefit analysis, it is only more clearly exhibited when a risk-benefit analysis is performed in order to guide the decision.

On the other hand, money has connotations not shared by non-monetary units that can sometimes be used for the same or similar purposes, such as QALYs (quality-adjusted life years). The use of money instead of some other unit may therefore send a message that can be conceived as desecrating the value of life.

One of the most common methods used to derive calculation values for non-market goods is contingent valuation (willingness to pay studies, WTP). This means that the values are based on people's answers to questions of the type "How much would you be prepared to pay for saving the giant panda from extinction?" The presumption is that the sum of everyone's answers to that question determines the value that the non-extinction of the giant panda should be assigned in an economic analysis. It turns out, however, that our answers to such questions do not give good indications of our priorities. Hence, Beattie and coworkers [1998] found that many respondents tend to report an amount that would not seriously disturb their normal expenditure and savings patterns, typically a sum in the range £50–200 per annum. Respondents were also insensitive to the magnitude of the risk reduction. No other, more reliable method seems to be available for eliciting calculation values from questionnaire respondents.

Some methods used in risk-benefit analysis, including contingent valuation, tend to give more influence to affluent people since they can pay more than others to

have it their way [Copp, 1987]. Although this is a common feature in risk-benefit analysis, it is methodologically easy to avoid for instance by relating (actual or hypothetical) payments by an individual to that individual's income.

4.6 Transferability across contexts

In risk-benefit analyses, cost estimates are regularly transferred across contexts. This applies, in particular, to estimates of life values. Two examples can be given to illustrate this practice. The first example is a risk-benefit analysis of mammography that was performed in 1992 by the American FDA. The analysis made use of values of life to determine, in monetary terms, the economic benefits from saving a life with mammography. The life values were derived from estimates of how much more male workers are paid when working in occupations with a high risk of fatal accidents [Heinzerling, 2000, pp. 205–206]. The second example is a risk-benefit analysis performed in 2000 by the American EPA in order to determine a new standard for arsenic in drinking water. Here again, values of life were taken from studies of how much compensation male workers receive for risks of fatal accidents [Heinzerling, 2002, p. 2312].

In both these cases, it would have been possible to use life values derived from the very context of the risk-benefit analysis in question. Women could have been asked how much they are prepared to pay for mammography, given realistic assumptions about the risk reduction it gives rise to. Their willingness to pay for reduced risks could then be used in a risk-benefit analysis for mammography. Although the use of such values would not have been unproblematic, it would at least have been much closer to the relevant context than the life value that was actually used. Similarly, people could have been asked how much they were prepared to pay for reduced levels of arsenic in drinking water, given realistic assumptions about the health effects of such a reduction. Alternatively, willingness to pay for healthy water (in the form of mineral water) could have been obtained from actual markets. Instead, life values were transferred from another context, namely that of wage compensation for occupational health risks.

The assumed transferability across contexts that is illustrated in these examples is in fact an essential condition without which current methods of cost-benefit analyses cannot be justified. If all values used in a risk-benefit analysis had to be derived from the precise context of the particular analysis, then the practice of risk-benefit analysis would come close to that of performing opinion polls on the topic to be analyzed. Once transferability across contexts is given up, we seem to enter a slippery slope in which the characteristic features of risk-benefit analysis as we know them today would be lost. In order to avoid this, and defend transferability across contexts, one would have to claim that it is *better* to use values from a certain context (such as wages that compensate for workplace risks) in a risk-benefit analysis concerning another context (such as mammography), than to use values derived in the context of the analysis in question. No such argument seems to be available. In particular, it has not been shown that our

decisions on what employment offers to seek and accept are better informed than most other decisions that we make (such as decisions on mammography and on contaminants in drinking-water).

4.7 Interpersonal aggregation

In a risk-benefit analysis, all risks and all benefits are combined in one and the same balance. This means that a disadvantage affecting one person can be fully compensated for by an advantage of the same size that affects some other person. In other words, *interpersonal compensability* of advantages and disadvantages is assumed [Hansson, 2004].[6] This is an assumption that risk-benefit analysis shares with utilitarianism. We can express it as a weighing principle, as follows:

> *The collectivist weighing principle:*
>
> An option is acceptable to the extent that the sum of all individual risks that it gives rise to is outweighed by the sum of all individual benefits that it gives rise to.

This is not the only way in which risks can be weighed against benefits. Another possibility is to perform the weighing individually for each affected person, and require a positive balance for each person:

> *The individualist weighing principle:*
>
> An option is acceptable to the extent that the risks affecting each individual are outweighed by benefits for that same individual.

Individualist weighing has a strong tradition in social practices that have their origin in the physician–patient relationship. For an example, consider a physician who selects patients for a clinical trial with a new, experimental treatment. Such a treatment involves risks and benefits that have to be weighed against each other. If the physician based this decision on a conventional risk-benefit analysis, then she would include a patient in the study if the risk to this patient is outweighed by the total social benefit. The total social benefit includes the expected gains from the study for future patients. With such a criterion, a patient can be included in the trial even if the risks by far exceed the expected gains to her personally. This, of course, is not how such decisions are made. Instead, they are made in accordance with the individualist weighing principle. A patient is not offered to participate in a clinical trial unless it is believed that the risks to which she will be exposed are outweighed by the expected advantages for her of the experimental treatment [Hansson, 2006b].

[6]Interpersonal compensability should not be conflated with the related but distinct issue of interpersonal comparability. Even if a benefit is greater than a harm, it need not cancel out the harm. Interpersonal comparability does not imply interpersonal compensability, but they are nevertheless closely related since the former is a necessary prerequisite for making the latter operative.

For another example, consider recommendations by health authorities on fish consumption. Although fish is generally speaking healthy food, contaminants in fish caught in certain waters give reason to recommend limits in fish consumption. Such recommendations are based on the positive and negative health effects on the individual (and in the case of pregnant or breast-feeding women, on corresponding effects on the child) [Knuth *et al.*, 2003]. It would be regarded as inappropriate to base such recommendations on a full risk-benefit analysis that included other factors, such as the effects of diminished fish consumption on employment in the fishing industry or on regional economics.

Hence we perform risk-benefit analyses, with collectivist risk-weighing, when deciding on road projects and other engineering projects, but use other types of calculations, based on individual risk-weighing, when deciding on clinical trials and dietary advice. In some policy areas we have a tradition of sacrificing individual interests for the sake of collective goals, whereas individual interests have a much stronger protection in other areas. It is a problem for risk-benefit analysis to motivate why we should employ total (collective) aggregation instead of alternative methods that protect individuals against sacrifice of their interests for collective goals.

4.8 Risk analysis versus safety engineering

Probabilistic risk analysis and cost-benefit analysis have sometimes been seen as competitors of traditional forms of safety engineering. This is a too narrow view of the matter. Instead, it should be recognized that neither of these methods can in practice tell the full truth about risk and safety. It is more constructive to see them as complementary. Probabilistic risk analysis is often an indispensable tool for priority-setting and for the effect evaluation of safety measures. On the other hand, some of the uncertainties that safety engineering deals with successfully tend to be neglected in probabilistic calculations. Methodological pluralism, rather than monopoly for one single methodology, is to be recommended.

Currently there is a trend in several fields of engineering towards increased use of probabilistic risk analysis. This trend will strengthen safety engineering, provided that it leads to a broadening of the knowledge base and not to exclusion of the wide range of dangers — from one's own miscalculations to terrorist attacks — for which no meaningful probability estimates can be obtained.

5 THE ETHICS OF RISK

Throughout the history of moral philosophy, moral theorizing has for the most part referred to a deterministic world in which the morally relevant properties of human actions are both well-determined and knowable. In recent years, moral philosophers have in most cases left it to decision theorists to analyse the complexities that the indeterminism of real life gives rise to. Mainstream ethical (and metaethical) theories still focus on deterministic problems; in fact they lack the

means to deal with problems involving risk and uncertainty [Hansson, 2003]. In this section we will investigate how moral theory can be extended so that it can support decisions on technological risks.

5.1 The causal dilution problem

How can we generalize ethical theories so that they can be effectively applied to problems involving risk and uncertainty? The problem of how to perform this generalization can be specified in terms of *the causal dilution problem*.

> *The causal dilution problem (general version):*
>
> Given the moral appraisals that a moral theory T makes of value-carriers with well-determined properties, what moral appraisals does (a generalized version of) T make of value-carriers whose properties are not well-determined beforehand?

The term "moral appraisal" covers a wide range of assignments of moral status, such as declarations that something is forbidden, permitted, morally required, good, bad, better than something else to which it is compared, etc. The term "value-carriers" refers to all entities that can be assigned (moral) value, including in particular human actions and the outcomes of human actions.

Under conditions of risk, we can restate the causal dilution problem as follows:

> *The causal dilution problem (probabilistic version):*
>
> Given the moral appraisals that a moral theory T makes of value-carriers with well-determined properties, what moral appraisals does (a generalized version of) T make of probabilistic mixtures of such value-carriers?

5.2 Actualism

We will begin with utilitarianism, the moral theory that has most often been applied to problems of risk. One fairly obvious approach to the causal dilution problem for utilitarianism is the following [Carlson, 1995]

> *Actualism*
>
> The utility of a (probabilistic) mixture of potential outcomes is equal to the utility of the outcome that actually materializes.

To exemplify the actualist approach, consider an engineer's decision whether or not to reinforce a bridge before it is being used for a single, very heavy transport. There is a 50 % risk that the bridge will collapse if it is not reinforced. Suppose that she decides not to reinforce the bridge and that everything goes well; the bridge is not damaged. According to the actualist approach, what she did was right. This is, of course, contrary to common moral intuitions.

The actualist solution requires that we use moral terms such as "right" and "wrong" in a way that differs radically from ordinary usage. If we accept the actualist usage, then it will in most cases be impossible to know what is right or wrong (or permitted, morally required, good, best, etc.) to do. In this way, action-guidance is expelled from moral discourse. However, action-guidance is largely what we need ethics for. Therefore, this is an unusually unhelpful approach. If we follow it, then action-guidance will have to be reintroduced in some other way.

5.3 Expected utility

The standard decision-theoretical solution to the causal dilution problem for utilitarianism is the maximization of expected utility. To maximize expected utility means to choose among a set of alternatives one of those that have the highest expected, i.e. probability-weighted utility:

> *Expected utility:*
>
> The utility of a probabilistic mixture of potential outcomes is equal to the probability-weighted average of the utilities of these outcomes.

The strongest argument in favour of maximizing objectivist expected utility is that this is a fairly safe method to maximize the outcome in the long run. Suppose, for instance, that the expected number of deaths in traffic accidents in a region will be 300 per year if safety belts are compulsory and 400 per year if they are optional. Then, if these calculations are correct, about 100 more persons per year will actually be killed in the latter case than in the former. We know, when choosing one of these options, whether it will lead to fewer or more deaths than the other option. If we aim at reducing the number of traffic casualties, then this can, due to the law of large numbers, safely be achieved by maximizing the expected utility (i.e., minimizing the expected number of deaths).

The validity of this argument depends on the large number of road accidents that levels out random effects in the long run. Therefore, the argument is not valid for case-by-case decisions on unique or very rare events. Suppose, for instance, that we have a choice between a probability of 0.001 of an event that will kill 50 persons and a 0.1 probability of an event that will kill one person. Here, random effects will not be levelled out as in the traffic belt case. In other words, we do not know, when choosing one of the options, whether or not it will lead to fewer deaths than the other option. In such a case, taken in isolation, there is no compelling reason to maximize expected utility.

Nevertheless, a decision in this case to prefer the first of the two options (with the lower number of expected deaths) may very well be *based on* a reasonable application of expected utility theory, namely if the decision is included in a sufficiently large group of decisions for which a metadecision has been made to maximize expected utility. As an example, a case can be made that a criterion for the regulation of safety equipment in motorcars should be one of maximizing expected utility (minimizing expected damage). The consistent application of this criterion

in all the different specific regulatory decisions should minimize the damage caused by technical failures of motor vehicles.

The larger the group of decisions that are covered by such a rule, the more efficient is the levelling-out effect. In other words, the larger the group of decisions, the larger catastrophic consequences can be levelled out. However, there is both a practical and an absolute limit to this effect. The *practical* limit is that decisions have to be made in manageable pieces. If too many issues are lumped together, then the problems of information processing may lead to losses that outweigh any gains that might have been hoped for. Obviously, decisions can be partitioned into manageable bundles in many different ways, and how this is done may have a strong influence on decision outcomes. As an example, the protection of workers against radiation may not be given the same priority if it is grouped together with other issues of radiation as if it is included among other issues of work environment.

The *absolute* limit to the levelling-out effect is that some extreme outcomes, such as a nuclear war or a major ecological threat to human life, cannot be levelled out even in the hypothetical limiting case in which all human decision-making aims at maximizing expected utility. Perhaps the best example of this is the Pentagon's use of secret utility assignments to accidental nuclear strike and to failure to respond to a nuclear attack, as a basis for the construction of command and control devices [Paté-Cornell and Neu, 1985].

Even when the levelling-out argument for expected utility maximization is valid, compliance with this principle is not required by rationality. In particular, it is quite possible for a rational agent to refrain from minimizing total damage in order to avoid imposing high-probability risks on individuals.

To see this, let us suppose that we have to choose, in an acute situation, between two ways to repair a serious gas leakage in the machine-room of a chemical factory. One of the options is to send in the repairman immediately. (There is only one person at hand who is competent to do the job.) He will then run a risk of 0.9 to die due to an explosion of the gas immediately after he has performed the necessary technical operations. The other option is to immediately let out gas into the environment. In that case, the repairman will run no particular risk, but each of 10,000 persons in the immediate vicinity of the plant runs a risk of 0.001 to be killed by the toxic effects of the gas. The maxim of maximizing expected utility requires that we send in the repairman to die. This is also a fairly safe way to minimize the number of actual deaths. However, it is not clear that it is the only possible response that is rational. A rational decision-maker may refrain from maximizing expected utility (minimizing expected damage) in order to avoid what would be unfair to a single individual and infringe her rights.

5.4 Deontological and rights-based theories

The causal dilution problem for rights-based theories was formulated (in its probabilistic version) by Robert Nozick: "Imposing how slight a probability of a harm that violates someone's rights also violates his rights?" [Nozick, 1974, p. 7]. In

somewhat more general language we can restate it, and its deontological counterpart, as follows:

> *The causal dilution problem for deontological/rights-based moral theories (general version):*
>
> Given the duties/rights that a moral theory T assigns with respect to actions with well-determined properties, what duties/rights does (a generalized version of) T assign with respect to actions whose properties are not well-determined beforehand?

> *The causal dilution problem for deontological/rights-based moral theories (probabilistic version):*
>
> Given the duties/rights that a moral theory T assigns with respect to actions with well-determined properties, what duties/rights does (a generalized version of) T assign with respect to probabilistic mixtures of such actions?

An extension of a deontological theory to indeterministic cases can be obtained by just prescribing that a prohibition to bring about a certain outcome implies a prohibition to cause an increase in the risk of that outcome (even if the increase is very small). Similarly, for a rights-based theory, it could be claimed that if I have a right that you do not bring about a certain outcome, then I also have a right that you do not perform any action that has a non-zero risk of bringing about that outcome. Unfortunately, such a strict extension of rights and prohibitions is socially untenable. Your right not to be killed by me certainly implies a prohibition for me to perform certain acts that involve a risk of killing you, but it cannot prohibit all such acts. Such a strict interpretation would make human society impossible. I am allowed to drive a car in the town where you live, although this increases the risk of being killed by me.

Hence, rights and prohibitions have to be defeasible so that they can be cancelled when probabilities are small. The most obvious way to achieve this is to assign to each right (prohibition) a probability limit. Below that limit, the right (prohibition) is cancelled. However, as Nozick observed, such a solution is not credible since probability limits "cannot be utilized by a tradition which holds that stealing a penny or a pin or anything from someone violates his rights. That tradition does *not* select a threshold measure of harm as a lower limit, in the case of harms certain to occur" [Nozick, 1974, p. 75].

Clearly, a moral theory need not treat a slight probability of a sizable harm in the same way that it treats a slight harm. The analogy is nevertheless relevant. The same basic property of traditional rights theories, namely the uncompromising way in which they protect against disadvantages for one person inflicted by another, prevents them from drawing a principled line either between harms or between probabilities in terms of their acceptability or negligibility. In particular, since no rights-based method for the determination of such probability limits seems to be

available, they would have to be external to the rights-based theory. Exactly the same problem obtains for deontological theories.

Probability limits do not solve the causal dilution problem for these types of theories. No other solution of the causal dilution problem for these theories seems to be available.

5.5 Contract theories

Contract theories may perhaps appear somewhat more promising. The criterion that they offer for the deterministic case, namely consent among all those involved, can also be applied to risky options. Can we then solve the causal dilution problem for contract theories by saying that risk impositions should be accepted to the degree that they are supported by a consensus?

Unfortunately, this solution is far from unproblematic. Consent, as conceived in contract theories, is either actual or hypothetical. Actual consent does not seem to be a realistic criterion in a complex society in which everyone performs actions with marginal but additive effects on many people's lives. According to the criterion of actual consent, you have a veto against me or anyone else who wants to drive a car in the town where you live. Similarly, I have a veto against your use of coal to heat your house, since the emissions contribute to health risks that affect me. In this way we can all block each other, creating a society of stalemates. When all options in a decision are associated with risk, and all parties claim their rights to keep clear of the risks that others want to impose on them, the criterion of actual consent does not seem to be of much help.

We are left then with hypothetical consent. However, as the debate following Rawls's *Theory of* Justice has shown, there is no single decision-rule for risk and uncertainty that all participants in a hypothetical initial situation can be supposed to adhere to [Hare, 1973; Harsanyi, 1975]. It remains to show that a viable consensus on risk-impositions can be reached among participants who apply different decision-rules in situations of risk and uncertainty.[7] Apparently, this has not been done, and hence, contract theory does not either have a solution to the causal dilution problem.

5.6 Restating the problem

The difficulties that we encounter when trying to solve the causal dilution problem within the frameworks of common types of moral theories are indications of a deeper problem. The attempted solutions reviewed above are all based on an implicit derivation principle: It is assumed that given moral appraisals of actions with deterministic outcomes, we can derive moral appraisals of actions whose outcomes are probabilistic mixtures of such deterministic outcomes. In other words,

[7] If a unanimous decision is reached due to the fact that everybody applies the same decision-rule, then the problem has not been solved primarily by contract theory but by the underlying theory for individual decision-making.

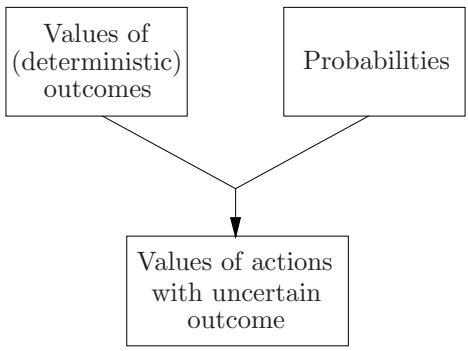

Figure 1. The standard view of how values of indeterministic options can be determined.

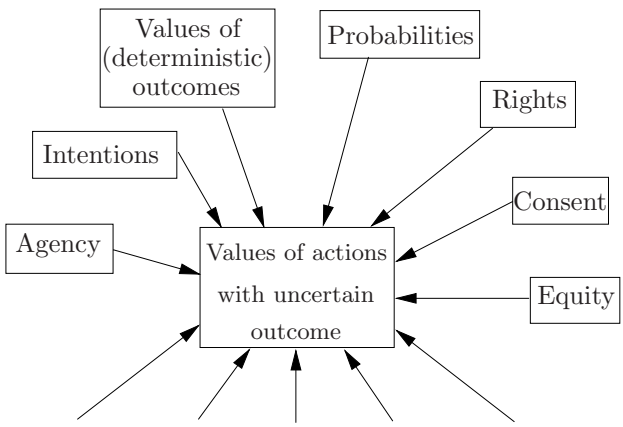

Figure 2. A less incomplete picture of the influences on the values of indeterministic options.

it is assumed that probabilities and (deterministic) utilities are all the information that we need (Figure 1). However, this picture is much too simplified. The morally relevant aspects of situations of risk and uncertainty go far beyond the impersonal, free-floating sets of consequences that decision theory operates on. Risks are inextricably connected with interpersonal relationships. They do not just "exist"; they are taken, run, or imposed [Thomson, 1985]. To take just one example, it makes a moral difference if it is one's own life or that of somebody else that one risks in order to earn a fortune for oneself. Therefore, person-related aspects such as agency, intentionality, consent etc. will have to be taken seriously in any reasonably accurate account of real-life indeterminism (Figure 2).

A moral analysis of risk that includes considerations of agency and responsibility will be an analysis more in terms of the verb (to) "risk" than of the noun (a) "risk". Major policy debates on risks have in part been clashes between the "noun" and the "verb" approach to risk. Proponents of nuclear energy emphasize how small *the risks* are, whereas opponents question the very act of *risking* improbable but potentially calamitous accidents.

Based on this analysis, the causal dilution problem can be replaced by an *defeasance problem* that better reflects the moral issues of risk impositions:

> *The defeasance problem:*
>
> It is a prima facie moral right not to be exposed to risk of negative impact, such as damage to one's health or one's property, through the actions of others. What are the conditions under which this right is defeated, so that someone is allowed to expose other persons to risk?

5.7 Solving the defeasance problem

In social practice, the prima facie moral right not to be exposed to risk has to be defeated quite often. Social life would be impossible if we were not allowed to expose each other to certain risks. It is important to observe that a right can be meaningful and socially important even if it can often be defeated. That it is a right means that it prevails whenever there are no overweighing reasons not to realize it.

To make this more concrete: as car-drivers we put each other's lives at risk. However, if we are all allowed to drive a car, exposing each other to certain risks, then we can all lead more mobile lives, and this will on balance be to the benefit of all of us (or so we may assume). The same principle can be applied to exchanges of different types of risks and benefits, as long as these exchanges are mutually beneficial. However, there is (or should be) a limit: No single person should be exposed to risks to an extent or in ways that are to the benefit only of others, not herself. We cannot require that every single risk-exposure be to the risk-exposed person's benefit, but we can demand that the totality of risk-exposures be arranged so that everyone gains, and no one is exploited. This will lead us to the following solution to the defeasance problem:

(E) Nobody should be exposed to a risk unless it is part of an equitable social system for risk-taking that works to her advantage.

This rule needs, of course, to be specified in several respects, both for theoretical purposes and to make it useful in concrete applications. It should be compared to the dominating approach in risk analysis that can be summarized as follows:

(RA) A risk imposition is acceptable if the total benefits that it gives rise to outweigh the total risks, measured as the probability-weighted disutility of outcomes.

By choosing a rule such as (E), rather than (RA), we change the agenda for discussions on risk. We choose to treat each risk-exposed person as a sovereign individual who has a right to a fair treatment, rather than as a carrier of utilities and disutilities that would have the same worth if they were carried by someone else. In order to argue according to (RA) that it is acceptable to impose a risk on a particular person, one has to give sufficient reasons for accepting the risk as such, as an impersonal entity. According to (E), one instead has to give sufficient reasons for accepting that this particular person is exposed to the risk.

5.8 Hypothetical retrospection

With further developments, the approach introduced in the previous subsection can help us to deal with the distributive issues in risk. However, it does not help us to deal with the equally fundamental issue of which risks we should accept. This is a matter that transcends the limit between morality and rational self-interest.

Of course the standard answer to this question is that we should apply expected utility theory. However, that theory has several weaknesses, some of which have been highlighted above. Another important weakness is its instability against the actual occurrence of a serious negative event that was included in the calculation. This can be seen by studying the post-accident argumentation after almost any accident. If the expected utility argumentation were followed to the end, then many accidents would be defended as consequences of a maximization of expected utility that is, *in toto*, beneficial. However, this type of reasoning is very rarely heard in practice. Seldom do we hear a company that was responsible for a deadly accident justify the loss of lives by saying that it was the result of a decision which, in terms of its total effects, produced far more good than harm. Instead, two other types of reactions are common. One of these is to regret one's shortcomings and agree that one should have done more to prevent the accident. The other is to claim that somebody else was responsible for the accident.

It should also be noted that accident investigation boards are instructed to answer the questions "What happened? Why did it happen? How can a similar event be avoided?", not the question "Was the accident defensible in an expected utility calculation?". Once a serious accident has happened, the application of expected utility maximization appears much less satisfactory than what it did

before the accident. In this pragmatic sense, expected utility maximization is not a stable strategy.

A framework for argumentation is needed that increases our ability to come up with risk decisions that we are capable of defending even if things do not go our way. Such a framework can be obtained by a systematizing a common type of arguments in everyday discussions about future possibilities, namely arguments that refer to how one might in the future come to evaluate the possible actions under consideration. These arguments are often stated in terms of predicted regret: "Do not do that. You may come to regret it." This type of argument can be systematized into a procedure in which future developments are systematically identified, and decision alternatives are evaluated under each of these possible developments. Such *hypothetical retrospection* can be used as a means to achieve more well-considered social decisions in issues of risk. However, it cannot be adequately accounted for in terms of regret-avoidance. Psychologically, regret is often unavoidable for the simple reason that it may arise in response to information that was not available at the time of decision. Therefore, regret-avoidance has to be replaced by more carefully carved-out methods and criteria for hypothetical retrospection [Hansson, 2007].

For a simple example, consider a factory owner who has decided to install an expensive fire alarm system in a building that is used only temporarily. When the building is taken out of use, the fire alarm has yet never been activated. The owner may nevertheless consider the decision to install it to have been right, since at the time of decision other possible developments had to be considered in which the alarm would have been life-saving. This argument can be used, not only in actual retrospection, but also, in essentially the same way, in hypothetical retrospection before the decision. Alternatively, suppose that there is a fire in the building. The owner may then regret that he did not install a much more expensive but highly efficient sprinkler system. In spite of this regret he may consider the decision to have been correct since when he made it, he had to consider the alternative, much more probable development in which there was no fire, but the cost of the sprinklers would have made other investments impossible. Of course, this argument can be used in hypothetical retrospection just like the previous one. In this way, when we perform hypothetical retrospection from the perspective of a particular branch of future development, we can refer to each of the alternative branches and use it to develop either counterarguments or supportive arguments. In short, in each branch we can refer to all the others.

Hypothetical retrospection can be developed into a precise procedure for collective deliberation on future risks and uncertainties [Godman and Hansson, 2009]. However, it can also be simplified to a risk manager's version: "Make a decision that you can defend also if an accident happens." In both cases, hypothetical retrospection aims at ensuring that whatever happens, the decision one makes will be morally acceptable (permissible) from the perspective of actual retrospection. Just as we can improve our moral decisions by considering them from the

perspective of other concerned individuals, we can also improve them by considering alternative future perspectives.

BIBLIOGRAPHY

[Ahteensuu, 2008] M. Ahteensuu. *In Dubio Pro Natura?* PhD thesis in philosophy, University of Turku 2008.

[Beattie *et al.*, 1998] J. Beattie, J. Covey, P. Dolan, L. Hopkins, M. Jones-Lee, G. Loomes, N. Pidgeon, A. Robinson, and A. Spencer. On the Contingent Valuation of Safety and the Safety of Contingent Valuation: Part 1–*Caveat Investigator*. *Journal of Risk and Uncertainty* 17:5-25, 1998.

[Bollinger *et al.*, 1996] R. E. Bollinger, D.G. Clark, A.M. Dowell III, R.M. Ewbank, D.C. Hendershot, W.K. Lutz, S.I. Meszaros, D.E. Park, and E.D. Wixom. *Inherently Safer Chemical Processes – A Life Cycle Approach*. Center for Chemical Process Safety of the American Institute of Chemical Engineers, New York, 1996.

[Bondi, 1985] H. Bondi. Risk in perspective. In *Risk,* M.G. Cooper, ed., pp. 8-17, 1985.

[Brinkmann *et al.*, 2006] G. Brinkmann. J. Pirson, S. Ehster, M.T. Dominguez, L. Mansani, I. Coe, R.. Moormann, and W. Van der Mheen. Important viewpoints proposed for a safety approach of HTGR reactors in Europe. Final results of the EC-funded HTR-L project. *Nuclear Engineering and Design* 236:463-474, 2006.

[Carlson, 1995] E. Carlson. *Consequentialism Reconsidered*. Kluwer, 1995.

[Christensen-Szalanski and Bushyhead, 1981] J. J. J. Christensen-Szalanski and J.B. Bushyhead. Physicians' use of probabilistic information in a real clinical setting. *Journal of Experimental Psychology: Human Perception and Performance* 7:928-935, 1981.

[Clausen *et al.*, 2006] J. Clausen, S.O. Hansson, and F. Nilsson. Generalizing the Safety Factor Approach. *Reliability Engineering and System Safety* 91:964-973, 2006.

[Cohen, 1985] B. L. Cohen. Criteria for Technology Acceptability, *Risk Analysis* 5:1-3, 1985.

[Copp, 1987] D. Copp. The Justice and Rationale of Cost-Benefit Analysis. *Theory and Decision* 23:65-87, 1987.

[Ditlevsen, 1994] O. Ditlevsen. Distribution arbitrariness in structural reliability. In *Proc. of ICOSSAR'93: Structural Safety & Reliability*. G. Schuëller, M. Shinozuka, and J. Yao, eds., pp. 1241-1247, 1994.

[Duncan, 2000] J. M. Duncan. Factors of safety and reliability in geotechnical engineering. *Journal of Geotechnical and Geoenvironmental Engineering* 126:307-316, 2000.

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

[Godman and Hansson, 2009] M. Godman and S.O. Hansson. European Public Advice on Nanobiotechnology – Four Convergence Seminars. *Nanoethics*, 2009.

[Hansson, 1993] S. O. Hansson. The false promises of risk analysis. *Ratio* 6:16-26, 1993.

[Hansson, 1999] S. O. Hansson. The Moral Significance of Indetectable Effects. *Risk* 10:101-108, 1999.

[Hansson, 2002] S. O. Hansson. Uncertainties in the knowledge society. *Social Science Journal* 171:39-46, 2002.

[Hansson, 2003] S. O. Hansson. Ethical criteria of risk acceptance. *Erkenntnis* 59:291-309, 2003.

[Hansson, 2004] S. O. Hansson. Weighing Risks and Benefits. *Topoi* 23:145-152, 2004.

[Hansson, 2006a] S. O. Hansson. Economic (ir)rationality in risk analysis. *Economics and Philosophy*, 22:231-241, 2006.

[Hansson, 2006b] S. O. Hansson. Uncertainty and the Ethics of Clinical Trials. *Theoretical Medicine and Bioethics* 27: 149–167, 2006.

[Hansson, 2007] S. O. Hansson. Hypothetical retrospection. *Ethical Theory and Moral Practice* 10:145-157, 2007.

[Hansson, 2008] S. O. Hansson. From the Casino to the Jungle. *Synthese*, in press, 2008.

[Hare, 1973] R. M. Hare. Rawls's Theory of Justice. *American Philosophical Quarterly* 23:144-155 and 241-252, 1973.

[Harsanyi, 1975] J. C. Harsanyi. Can the maximin principle serve as a basis for morality – Critique of Rawls, J theory. *American Political Science Review* 69(2): 594-606, 1975.

[Heinzerling, 2000] L. Heinzerling. The rights of statistical people. *Harvard Environmental Law Review* 24: 189-207, 2000.

[Heinzerling, 2002] L. Heinzerling. Markets for Arsenic. *Georgetown Law Journal* 90:2311-2339, 2002.

[Hoerl and Fallin, 1974] A. E. Hoerl. and H. K. Fallin. Reliability of Subjective Evaluations in a High Incentive Situation. *Journal of the Royal Statistical Society* 137, Part 2, 227-230, 1974.

[Hynes and Vanmarcke, 1976] M. Hynes and E. Vanmarcke, Reliability of embankment performance predictions. *Proceedings of the ASCE Engineering Mechanics Division, Specialty Conference*, Waterloo, Ontario, Canada, University of Waterloo Press, 1976.

[International Organization for Standardization, 2002] International Organization for Standardization. *Risk Management – Vocabulary – Guidelines for use in standards*, ISO/IEC Guide 73:2002, 2002.

[Khan and Abbasi, 1998] F. I. Khan and S.A. Abbasi. Inherently safer design based on rapid risk analysis. *Journal of Loss Prevention in the Process Industries* 11:361-372, 1998.

[Knoll, 1976] F. Knoll. Commentary on the basic philosophy and recent development of safety margins. *Canadian Journal of Civil Engineering* 3:409-416, 1976.

[Knuth *et al.*, 2003] B. A. Knuth, N.A. Connelly, J. Sheeshka, and J. Patterson. Weighing Health Benefit and Health Risk Information when Consuming Sport-Caught Fish. *Risk Analysis* 23:1185-1197, 2003.

[Lichtenstein *et al.*, 1982] S. Lichtenstein *et al.* Calibration of probabilities: The state of the art to 1980. In *Judgment Under Uncertainty, Heuristics and Biases*. Kahneman *et al.*, eds., pp. 306-334, 1982.

[Luce and Raiffa, 1957] R. D. Luce and H. Raiffa. *Games and Decisions. Introduction and critical survey*. Wiley, 1957.

[Miller, 1988] C. O. Miller. System Safety. In *Human Factors in Aviation*. E. L. Wiener and D. C. Nagel, eds., pp. 53-80, Academic Press, 1988.

[Möller *et al.*, 2006] N. Möller, S.O. Hansson and M. Peterson. Safety is More Than the Antonym of Risk. *Journal of Applied Philosophy* 23(4):419-432, 2006.

[Moses, 1997] F. Moses. Problems and prospects of reliability-based optimisation. *Engineering Structures* 19:293-301, 1997.

[Murphy and Winkler, 1984] A. H. Murphy and R.L. Winkler. Probability forecasting in meteorology. *Journal of the American Statistical Association* 79:489-500, 1984.

[National Research Council, 1983] National Research Council (NRC) *Risk Assessment in the federal government: Managing the process*. National Academies Press, 1983.

[Nizick, 1974] R. Nozick, R. *Anarchy, State, and Utopia*. Basic Books 1974.

[Otway, 1987] H. Otway. Experts, Risk Communication, and Democracy. *Risk Analysis* 7:125-129, 1987.

[Paté-Cornell and Neu, 1985] M. E. Paté-Cornell and J.E. Neu. Warning Systems and Defense Policy: A Reliability Model for the Command and Control of U.S. Nuclear Forces. *Risk Analysis* 5:121-138, 1985.

[Porter *et al.*, 1991] A. L. Porter *et al. Forecasting and Management of Technology*. John Wiley & Sons, 1991.

[Raffensperger and Tickner, 1999] C. Raffensperger and J. Tickner, eds. *Protecting Public Health and the Environment: Implementing the Precautionary Principle*. Island Press, 1999.

[Rahman *et al.*, 2005] M. Rahman, A.-M. Heikkilä, and M. Hurme. Comparison of inherent safety indices in process concept evaluation. *Journal of Loss Prevention in the Process Industries* 18:327-334, 2005.

[Randall, 1976] F. A. Randall. The safety factor of structures in history. *Professional Safety*, January 1976:12-28.

[Rechard, 1999] R. P. Rechard. Historical Relationship Between Performance Assessment for Radioactive Waste Disposal and Other Types of Risk Assessment. *Risk Analysis* 19(5):763-807, 1999.

[Rosenberg, 1995] N. Rosenberg. Why Technology Forecasts Often Fail. *Futurist*, July-August 1995: 16-21.

[Royal Society, 1983] Royal Society. *Risk Assessment. Report of a Royal Society Study Group*. Royal Society, 1983.

[Sandin, 1999] P. Sandin. Dimensions of the Precautionary Principle. *Human and Ecological Risk Assessment* 5: 889-907, 1999.

[Sandin *et al.*, 2004] P. Sandin, B.-E. Bengtsson, Å. Bergman, I. Brandt, L. Dencker, P. Eriksson, L. Förlin, P. Larsson, A. Oskarsson, C. Rudén, A. Södergren, P. Woin, and S.O. Hansson. Precautionary Defaults – A New Strategy for Chemical Risk Management. *Human and Ecological Risk Assessment* 10(1):1-18, 2004.

[Tench, 1985] W. Tench. *Safety is No Accident*. Collins, 1985.

[Thompson, 1985] J. Thomson. Imposing Risk, in *To Breathe Freely*. M. Gibson, ed., pp. 124-140. Rowman & Allanheld, 1985.

[UNCED, 1993] UNCED. *The Earth Summit: The United Nations Conference on Environment and Development*, Rio De Janeiro 1992. Introduction and commentary by S.P. Johnson. Graham & Trotman, 1993.

[Zhu, 1993] T. L. Zhu. A reliability-based safety factor for aircraft composite structures. *Computers & Structures* 48:745-748, 1993.