2 A Panorama of the Philosophy of Risk

Sven Ove Hansson
Royal Institute of Technology, Stockholm, Sweden

S. Roeser, R. Hillerbrand, P. Sandin, M. Peterson (eds.), *Handbook of Risk Theory*,
DOI 10.1007/978-94-007-1433-5_2, © Springer Science+Business Media B.V. 2012

Abstract: The role of philosophy in the development of the risk sciences has been rather limited. This is unfortunate since there are many problems in the analysis and management of risk that philosophers can contribute to solving. Several of the central terms, including "risk" itself, are still in need of terminological clarification. Much of the argumentation in risk issues is unclear and in need of argumentation analysis. There is also still a need to uncover implicit or "hidden" values in allegedly value-free risk assessments. Eight philosophical perspectives in risk theory are outlined: From the viewpoint of *epistemology*, risk issues have brought forth problems of trust in expertise and division of epistemological labor. In *decision theory*, the decision-maker's degree of control over risks is often problematic and difficult to model. In the *philosophy of probability*, posterior revisions of risk estimates (in so-called hindsight bias) pose a challenge to the standard model of probabilistic reasoning. In the *philosophy of science*, issues of risk give us reason to investigate what influence the practical uses of knowledge can legitimately have on the scientific process. In the *philosophy of technology*, the nature of safety engineering principles and their relationship to risk assessment need to be investigated. In *ethics*, the most pressing problem is how standard ethical theories can be extended or adjusted to cope with the ethics of risk taking. In the *philosophy of economics*, the comparison and aggregation of risks falling to different persons give rise to new foundational problems for the theory of welfare. In *political philosophy*, issues such as trust and consent that have been discussed in connection with risk give us reason to reconsider central issues in the theory of democracy.

Introduction

Philosophy is often seen as an unworldly discipline, dealing with abstract and contrived issues that have very little connection with real life. Concededly, philosophy has a long tradition of unabashedly delving into intellectual problems that have no immediate application. In this it does not differ from most other academic disciplines. But philosophy also has another side. It has a strong tradition, going back at least to Socrates and Aristotle, of probing into issues that societies and individuals need to understand better in order to solve practical problems. And just as in other disciplines, some of the progress made in studies driven by pure intellectual curiosity has turned out to provide us with indispensible tools for investigations aimed at solving practical problems. Examples of this can be found in the philosophy of risk as well as other areas of applied philosophy.

In what follows, a brief historical background (section ➲ Historical Background) will be followed by a presentation of the major types of contributions that philosophy can make to risk research (section ➲ What Philosophy Can Contribute) and an overview over eight philosophical perspectives on risk (section ➲ Philosophical Perspectives in Risk Theory). Finally, some topics for further research will be summarized (section ➲ Further Research).

Historical Background

Modern risk research originated in studies from the 1960s and 1970s that had a strong focus on chemical risks and the risks associated with nuclear energy. From its beginnings, risk research drew on competence in areas such as toxicology, epidemiology, radiation biology, and nuclear engineering. Today, many if not most scientific disciplines provide risk analysts with specialized knowledge needed in the study of one or other type of risk – medical specialties are needed in the study of risks from diseases, engineering specialties in studies of technological failures, etc. In addition, several disciplines have supplied overarching approaches to risk, intended to be applicable to risks of different kinds. Statistics, epidemiology, economics, psychology, anthropology, and sociology are among the disciplines that have developed general approaches to risk.

Philosophers did not have a big role in the early development of risk analysis. Most of the philosophical contributions to the area were in fact outsiders' criticisms of risk analysis. There was a strong tendency in the early development of risk analysis to downplay value issues. Risk assessments were presented as objective scientific statements, even when taking a stand on value-laden issues such as risk acceptability. Most of the early philosophical work on risk had as its main purpose to expose the value-dependence of allegedly value-free risk assessments (Thomson 1985b; MacLean 1985; Shrader-Frechette 1991; Cranor 1997; Hansson 1998). This was an important task, and it was also undertaken with some success. Although hidden value assumptions are still common in risk assessments, there is now much more awareness of their presence. Philosophers who took part in these discussions certainly contributed to the steps that have been taken to keep facts and values apart as far as possible in the assessment of risk, in particular attempts to divide the risk-decision process into a fact-finding risk-assessment part and a value-based risk-management phase (National Research Council 1983).

In the 1990s, philosophers increasingly discovered many other risk-related issues in need of philosophical clarification. Philosophers have studied the nature of risks, the specific characteristics of knowledge about risk, the ethics of risk taking, its decision-theoretical aspects, the implications of risk in political philosophy, and several other areas. It is too early to write the history of these developments, but a pattern emerges in which most of the major subdisciplines of philosophy turn out to have important risk-related issues to deal with. These developments will be introduced in section ➤ Philosophical Perspectives in Risk Theory, but before that we are going to look more closely at the nature of the philosophical contribution.

What Philosophy Can Contribute

Philosophy is unique in having potential connections with virtually every other academic discipline. Philosophical concepts and methods have proven to be applicable to a wide variety of problems in other academic disciplines. When you probe into almost any field of learning, interesting problems of a philosophical nature tend to emerge. Unfortunately, this potential is underused, largely due to intellectual isolation and to the "two-cultures" phenomenon that separates philosophers from empirical scientists.

The contributions of philosophy to other disciplines and to interdisciplinary cooperations can be of many kinds, but experience shows that there are certain ways in which philosophy has particularly often turned out to be useful. Three of them are especially important in risk research:

- *Terminological clarifications*: Philosophy has a long tradition of constructing precise definitions and developing new distinctions, often beyond the limits of what can readily be expressed with current linguistic means. Armed with standard tools and distinctions from philosophy, philosophers can often contribute to conceptual clarification in other disciplines.

- *Argumentation analysis*: Arguments, as we express them in scientific or social debates, tend to depend on unstated assumptions. Using the tools of logic and conceptual analysis philosophers can often exhibit hidden assumptions and clarify the structure of arguments.
- *The fact–value distinction*: Factual input from science has a large and increasing role in debates on social issues. This applies to virtually all branches of science: economics, behavioral science, environmental science, climatology, medical science, technological sciences, etc. But even if scientists try to make their statements as value-independent as possible, they do not always succeed in this. Philosophical tools are useful in identifying the values that are inherent in science-based information.

In the following three subsections, we will have a brief look at each of these types of contribution, in order to show how philosophical method can contribute to investigations of risk that are performed primarily by researchers in other fields.

Terminological Clarification

As in many other research areas, the terminology in risk research is often imprecise. This applies even to key terms such as "risk" and "safety." The word "risk" has been taken over from everyday language, where it is used (often somewhat vaguely) to describe a situation in which we do not know whether or not some undesired event will occur. In risk analysis, two major attempts have been made to redefine risk as a numerical quantity. First, in the early 1980s, attempts were made to identify risk with the probability of an unwanted event (Fischhoff et al. 1981; Royal Society 1983, p. 22). This usage has some precedents in colloquial language; we may for instance say that "the risk that this will happen is one in twenty." Secondly, in more recent years, several attempts have been made to identify risk with the statistical expectation value of unwanted events. By this is meant the product of an event's probability with some measure of its undesirability. If there is a probability of 1 in 100 that three people will die, then "the risk" is said to be 0.03 deaths. Currently, this is by far the most common technical definition of risk (International Organization for Standardization 2002; Cohen 2003).

From the viewpoint of philosophical definition theory (Hansson 2006b), this terminology is problematic in at least two ways. First, it conflates "risk" with "severity of risk." It makes sense to say that a probability of 1 in 1,000 that one person will die in a roller coaster accident is "equally serious" as a probability of 1 in 1,000,000 that 1,000 people will die in a nuclear accident, but it does not make sense to say that these two are the same risk (namely 0.001 deaths). They are in fact risks with quite different characteristics.

Secondly, it is a controversial value statement that risks with the same expectation value of undesirable events are always equally serious. Some authors have claimed that serious events with low probabilities should be given a higher weight in decision making than what they receive in the expected utility model (O'Riordan and Cameron 1994; O'Riordan et al. 2001; Burgos and Defeo 2004). The identification of "risk" with expectation values has the unfortunate effect of ruling out this view on the severity of risk by means of a terminological choice. In order to achieve clarity in discussions on risk, we need to make a clear distinction between a risk and its severity, and we also need to avoid terminology that takes controversial standpoints on what constitutes risk severity for granted. Therefore, a term such as "expected damage" is much preferable to "risk" as a designation of the statistical expectation values employed in risk analysis (Hansson 2005).

Besides "risk," several other terms used in risk studies are in need of terminological clarification. Prominent among these are "safety" and "precautionary principle."

"Safety" has sometimes been defined as a situation without accidents (Tench 1985) and on other occasions as a situation with an acceptable probability of accidents (Miller 1988). In a recent philosophical analysis of the concept, it was shown that usage of the terms "safe" and "safety" vacillates between an absolute concept ("safety means no harm"), and a relative concept that only requires such risk reductions that are considered to be feasible and reasonable. It may not be possible to eliminate either of these usages, but it is possible to keep track of them and avoid confusing them with each other (Möller et al. 2006).

The "precautionary principle" is a principle for decision making under scientific uncertainty that has been codified in a several international treaties on environmental policies. Its major message is that policy decisions in environmental decisions can legitimately be based on scientific evidence of a danger, even if that evidence is not strong enough to constitute full scientific proof that the danger exists. There has been considerable controversy on the precise meaning of the principle. A careful philosophical analysis showed that the major definitions of the precautionary principle contain four major components, namely (1) a threat to the environment or to human health, (2) a degree of uncertainty that is sufficient for action (such as "even before scientific proof is established"), (3) the action that is then taken (e.g., "warn" or "forbid"), and (4) the level of prescription (e.g., "is mandatory") (Sandin 1999). 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). Although this analysis does not resolve the controversies on the principle, it facilitates a precise understanding of these controversies.

Argumentation Analysis

Ever since Aristotle, logical and argumentative fallacies have been an important topic in philosophy (Walton 1987). It is not difficult to find examples of traditional fallacies such as ad hominem in discussions on risk. In addition, there are fallacies that are specific to the subject matter of risk. The following is a sample of such fallacies:

Risk X is accepted.
Y is a smaller risk than X.
∴ Y should be accepted.

Risk X is natural.
∴ X should be accepted.

X does not give rise to any detectable risk.
∴ X does not give rise to any unacceptable risk.

There is no scientific proof that X is dangerous.
∴ No action should be taken against X.

Experts and the public do not have the same attitude to risk X.
∴ The public is wrong about risk X.

A's attitude to risk X is emotional.
∴ A's attitude to risk X is irrational.

For examples of the first five of these fallacies and clarifications of why they are fallacies, see Hansson (2004b). For the last of these fallacies, see Roeser (2006).

The Fact-Value Distinction

As already mentioned, the task of uncovering hidden value assumptions in risk assessments often requires philosophical competence. Implicit value components of complex arguments have to be discovered, and conceptual distinctions relating to values have to be made. Often, other competences are required as well. A thorough understanding of the technical contents of risk assessments is needed in order to determine what factors influence their outcomes (Hansson and Rudén 2006). This is an area in which cooperations between philosophy and other disciplines can be very fruitful.

For the philosophical part of this work, two distinctions are particularly important. The first is the seemingly trivial but, in practice, often overlooked distinction between being value-free and being free of controversial values. There are many values that are shared by virtually everyone or by everyone who takes part in a particular discourse. Medical science provides good examples of this. When discussing analgesics, we take for granted that it is better if patients have less rather than more pain. There is no need to interrupt a medical discussion in order to point out that a statement that one analgesic is better than another depends on this value assumption. Similarly, in economics, it is usually taken for granted that it is better if we all become richer. Economists sometimes lose sight of the fact that this is a value judgment. Obviously, a value that is uncontroversial in some circles may be controversial in others. This is one of the reasons why values believed to be uncontroversial should be made explicit and not treated as non-values.

The other distinction is that between epistemic and non-epistemic values. Most of the values that we usually think of in connection with risk policies are non-epistemic. The epistemic values are those that rule the conduct of science. Among the most commonly mentioned examples of such values are the attainment of truth, the avoidance of error, simplicity, and explanatory power. It was Carl Hempel who pointed out that these should be treated as values, although they are not moral values (Hempel 1960; Levi 1962; Feleppa 1981; Harsanyi 1983). Epistemic values are not necessarily less controversial than non-epistemic ones, but these are different types of controversies that should be kept apart.

The following are three examples of values that are often implicit or "hidden" in risk assessments:

1. *Values of error-avoidance*: Two major types of errors can be made in a scientific statement. Either you conclude that there is a phenomenon or an effect that is in fact not there. This is called an error of type I (a false positive). Or you miss an existing phenomenon or effect. This is called an error of type II (a false negative). In scientific practice, errors of type I are the more serious ones since they make us draw unwarranted conclusions, whereas errors of type II only make us keep an issue open instead of adopting a correct hypothesis. As long as we stay in the realm of pure science, the relative weights that we assign to the two types of error express our epistemic values, and they need not have any connection with our non-epistemic values. However, when scientific information is transferred to risk assessment, values of error-avoidance are transformed into non-epistemic and often quite controversial values. Consider the question "Does Bisphenol A impair infant brain development?" In a purely scientific context, the level of evidence needed for an affirmative answer to this question is a matter of

epistemic values. (How close to certainty should we be in order to take something to be a scientific fact?) In a risk assessment context, the relevant issue is what level of evidence we need to act as if the substance has this effect. This is a matter of non-epistemic values. (How much evidence is needed for treating the substance as toxic to infants?) In a case like this, a focus on epistemic values will usually lead to more weight being put on the avoidance of type I errors than type II errors, whereas a focus on non-epistemic values can have the opposite effect.

The distinction between a scientific assessment and a judgment of what should be done given the available scientific information is both fundamental and elementary, but it is nevertheless often overlooked (Rudén and Hansson 2008). It is not uncommon to find scientists unreflectingly applying epistemic standards of proof in risk assessment contexts where they are not warranted. It should be said to their defense that this is often more difficult to avoid than what one would perhaps think. Scientists are educated to focus on type I errors, and the tools of science are often ill suited to deal with type II errors. As one example of this, standard statistical practices for the evaluation of empirical data that have been tailored to the epistemic issue need to be adjusted in order to deal adequately with the risk assessment situation in which type II errors are usually more important (Krewski et al. 1989; Cranor and Nutting 1990; Leisenring and Ryan 1992; Hansson 1995, 2002).

2. *The value of naturalness*: In public debates, risks associated with GMOs or synthetic chemicals are often denounced as "unnatural." This argument is seldom used in risk assessments, but a converse version of it can sometimes be found, most often in connection to radiation. Radiation levels are frequently compared to the natural background with the tacit assumption that exposures lower than the natural background are unproblematic. In health risk assessments, this is a very weak argument. That something is natural does not prove that its negative effects on human health are small (Hansson 2003a). In ecological risk assessments, an argument referring to naturalness may be more relevant. If we want to protect the natural environment, then it is important to know what is natural. However, appeals to naturalness or unnaturalness are often made in a perfunctory way in discussions on ecological risk, and there is much need for clarification and analysis.
3. *Attitudes to sensitive individuals*: Risk assessments tend to focus on individuals with average sensitivity to the exposure in question. However, individual sensitivity differs and in many cases it is possible to identify groups of exposed persons who run a larger risk than others. According to the best available estimates, the radiogenic cancer risk is around 40% higher for women than for men at any given level of exposure. There are also small groups in the population who run a much higher risk. However, the recommended exposure limits are based on a population average rather than data for subpopulations (Hansson 2009a). From an ethical point of view, this is problematic. Exposing a person to a high risk cannot be justified by pointing out that the risk to an average person would have been much lower. Nevertheless, sensitive groups are often overlooked or disregarded in risk assessments, and the ethical implications of doing so are seldom discussed. It often takes careful study to reconstruct and analyze the underlying value assumptions.

Philosophical Perspectives in Risk Theory

In the previous section, we encountered several ways in which philosophers can contribute to interdisciplinary risk studies. Such contributions can be seen as applications of philosophy, mostly without much influence on the core of philosophical research. But the interaction between philosophy and risk studies does not end there. In recent years, it has become increasingly clear that risk has implications in many if not most of the philosophical subdisciplines. In what follows, we will have a look at several of these subdisciplines. In some of them, there is already an established tradition of studying risk. In others, little has yet been done, but interesting issues for future research can nevertheless be pointed out.

Epistemology

Risks are always connected to lack of knowledge. If we know for certain that there will be an explosion in a factory, then there is no reason for us to talk about that explosion as a risk. Similarly, if we know that no explosion will take place, then there is no reason either to talk about risk. What we refer to as a risk of an explosion is a situation in which it is not known whether or not an explosion will take place. In this sense, knowledge about risk is knowledge about the unknown. It is therefore a quite problematic type of knowledge. It gives rise to several important epistemological questions that have not been much studied. Two of them will be mentioned here.

The Limits of Epistemic Credibility

Some issues of risk refer to possible dangers that we know very little about. Recent debates on biotechnology and nanotechnology are examples of this. It is easy to find examples in which many of us would be swayed by considerations of unknown dangers. Suppose that someone proposed to eject a chemical substance into the stratosphere in order to compensate for the anthropogenic greenhouse effect. It would not be irrational to oppose this proposal solely on the ground that it may have unforeseeable consequences, even if all specified worries can be neutralized.

But on the other hand, it would not be feasible to take the possibility of unknown effects into account in all decisions that we make. Given the unpredictable nature of actual causation, almost any decision may lead to a disaster. We therefore have to disregard many of the more remote possibilities. It is easy to find examples in which it can be seen in retrospect that it was wise to do so. In 1969, *Nature* printed a letter that warned against producing polywater, polymerized water. The substance might "grow at the expense of normal water under any conditions found in the environment," thus replacing all natural water on earth and destroying all life on this planet (Donahoe 1969). Soon afterward, it was shown that polywater does not exist. If the warning had been heeded, then no attempts would have been made to replicate the polywater experiments, and we might still not have known that polywater does not exist. In cases like this, appeals to the possibility of unknown dangers may stop investigations and thus prevent scientific and technological progress.

It appears to be an unavoidable conclusion that we should take some but not all remote possibilities seriously. But which of them? What about the warnings that global warming might soon be aggravated by feedbacks that lead to a run-away greenhouse effect totally beyond our control, the warnings that the greenhouse effect may not exist at all, the warnings that mobile phones might have grave health effects, and that high-energy physics experiments might lead to an apocalypse? We are in need of concepts and criteria to discuss such issues in a systematic way, but as yet very little research has been performed on how to assess epistemic credibility in cases like this (Hansson 1996, 2004d).

The Legitimacy of Expertise in Uncertain Issues

In many issues of risk, we have seen wide divergences between the views of experts and those of the public. This is clearly a sign of failure in the social system for division of intellectual labor. However, it should not be taken for granted that every such failure is located within the minds of the nonexperts who distrust the experts. Experts are known to have made mistakes. A rational decision-maker should take into account the possibility that this may happen again. This will be particularly important in cases when experts assign very low probabilities to a highly undesirable event. Suppose that a group of experts have studied the possibility that a new microorganism that has been developed for therapeutic purposes will mutate and become virulent. They have concluded that the probability that this will happen is 1 in 100,000,000. Decision-makers who receive this report should of course consider whether this is an acceptable probability of such an event, given the advantages of using the new organism. But, arguably, this is not the most important question they should ask. The crucial issue is how much reliance they should put on the estimate. If there is even a very small probability that the experts are wrong, say a probability that we in some way estimate as 1 in a million, then that will be the main problem to deal with. In cases like this, reliance on experts creates serious epistemic problems that we do not yet seem to have adequate tools to analyze.

Decision Theory

A risk (in the informal sense of the word) is a situation in which some undesirable event may or may not occur, and we do not know which. Probability theory is a tool for modeling such situations. However, it should not be taken for granted that all such situations can be adequately modeled in that way. In many cases, our knowledge is so incomplete that no meaningful probability estimates are obtainable. In other cases, the situation may have features that make it unsuitable for probabilistic modeling. This applies in particular to risks that depend on complex interactions between independent agents. We all try both to influence the choices that others make and to foresee them and adjust to them. Therefore, our choices will depend in part on how we expect others to react and behave, and conversely their choices will depend on what they expect from us. Such interpersonal interactions are extremely difficult to capture in probabilistic terms.

This applies not least to malevolent action, such as the actions of an enemy, a saboteur, or a terrorist. Such agents try to take their adversaries with surprise. It is in practice impossible – and perhaps even counterproductive – to make probability estimates of their actions. For most purposes, a game-theoretical approach that makes no use of probabilities is more adequate to deal with inimical actions than models that employ probability estimates.

The use of probabilistic models is also problematic in situations where we have to take a whole series of decisions into account. The crucial issue here is whether or not one should treat one's own future choices and decisions as under one's present control (Spohn 1977; Rabinowicz 2002). The consequences at time t_3 of your actions at time t_1 are not determinate if you have an intermediate decision point t_2 at which you can influence what happens at t_3. In a moral appraisal of your actions at t_1, you have to decide whether to treat your actions at t_2 as under your own control at t_1 or as beyond your control at that point in time. In the former case, a decision at t_1 can bind your actions at t_2; in the latter case, it cannot do so. Consider the following two examples:

Example 1
A nonsmoker considers the possibility to smoke for just 1 week and then stop in order to achieve a better understanding of why so many people smoke. When making this decision, should she regard herself as being in control of the future decision whether or not to stop after a week? Or should she make a probabilistic appraisal of what she will do in that situation?

Example 2
A heavy cigarette smoker considers whether or not to try to quit. When making this decision, should she regard herself as being in control of future decisions whether or not to start smoking again? Or should she make a probabilistic appraisal of her future decisions? From such a viewpoint, quitting may seem to have a too meager prospect of success to be worth trying (Hansson 2007a).

Probably, most of us would recommend the non-control (probabilistic) approach to future decisions in Example 1 and the control (non-probabilistic) approach in Example 2. However, no general rule seems to be available to determine when a probabilistic approach to one's own future decisions is appropriate. Since most risk issues seem to require decisions on more than one occasion, this is a problem with high practical relevance. We have access to sophisticated decision-theoretical models that employ probabilities, but we do not have tools to determine when we should use these models and when we should instead use non-probabilistic approaches.

Philosophy of Probability

An average person's yearly risk of being struck by lightning is somewhat below one in a million (Lopez and Holle 1998). Risks of that magnitude have often been considered "negligible." After the Deepwater Horizon oil spill in 2010, BP's chief executive, Tony Hayward, said that the risk of this spill had been "one in a million" (Cox and Winkler 2010). However, there was no sign that the public – or the public authorities – were willing to discuss BP's responsibility from the premise that the accident was almost as unlikely as being struck by lightning. The fact that the accident had occurred was generally taken as proof that the company had not taken sufficient measures to prevent it from happening.

This example reveals a common pattern in how we argue about probabilities in the context of risks. 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. But, such an argument is very rarely heard in practice. Once a serious accident has happened, not much credence is given to the calculations showing that it was highly improbable. Instead, the very fact that the accident happened is taken as evidence that its probability was higher than estimated. Such reasoning has been disparaged by some as a fallacy, "hindsight bias" (Levi 1973; Fischhoff 1977). But it is not a fallacy.

Suppose that we know that a certain accident took place yesterday. Then the probability that it *did* happen was 1. Nevertheless, the probability that it *would* happen can have been much lower, say 1 in 100. As was observed by Blackburn in a different context, these are two distinctly different types of probabilities. "We can say that the probability of an event was high at some time previous to its occurrence or failure to occur, and this is not to say that it is now probable that it did happen" (Blackburn 1973, p. 102).

A simple example serves to show that our estimates of past probabilities can legitimately be influenced by information about what happened after the events in question. Suppose that a die was tossed 1,000 times yesterday. My original belief about the chance of a six on the first toss was that it was 1/6. When I learn that all the 1,000 tosses yielded a six, I change my opinion and assign a probability close to 1 to the event to which I previously had assigned 1/6. This is sensible since it is much more plausible that a die is biased than that a fair die yields the same outcome in 1,000 tosses.

The same principle applies to the probabilities referred to in risk analysis, such as the probability of an accident. Suppose that a new type of nuclear reactor is built. Nuclear engineers argue persuasively that it is much safer than previous designs. They convince us that the probability of a core damage ("meltdown") is 1 in 10^8. However, after the first reactor of the new type has been in service for only a couple of months, a serious accident involving core damage occurs. Probably, most people would not see this as an example of an extremely improbable event taking place. Instead, they would see the accident as a very strong indication that the probability estimate 1 in 10^8 was wrong. Just as in the example with the die, it would be perfectly rational to substantially revise one's estimate of the probability, perhaps from 1 in 10^8 to 1 in 10^5 or even higher. However, this is not an ordinary (Bayesian) revision of probabilities (A Bayesian revision refers to the probability that the accident actually took place, and thus takes us all the way from 10^{-8} to 1). This is a nonstandard form of probabilistic revision. It can be accounted for in terms of second-order probabilities (Hansson 2009b, 2010b), but its properties and its implications in assessments of risk remain to be investigated.

Philosophy of Science

In order to understand the relationship between risk assessments and scientific knowledge, it is useful to take intrascientific knowledge production as a starting point. The production of scientific knowledge begins with data that originate in experiments and other observations. Through a process of critical assessment, these data give rise to the scientific corpus (See ➋ *Fig. 2.1*). The corpus consists of that which is taken for given by the collective of researchers in their continued research and, thus, not questioned unless new data give reason to question it (Hansson 2007b). Hypotheses are included into the corpus when the data provide sufficient evidence for them, and the same applies to corroborated generalizations that are based on explorative research.

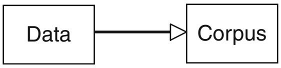


Fig. 2.1
The knowledge-formation process in pure science

The scientific corpus is a highly complex construction, much too large to be mastered by a single person. Different parts of it are maintained by different groups of scientific experts. These parts are all constantly in development. New statements are added, and old ones removed, in each of the many subdisciplines, and a consolidating process based on contacts and cooperations among interconnected disciplines takes place continuously. In spite of its complex structure, the corpus is, at each point in time, reasonably well defined. In most disciplines, it is fairly easy to distinguish those statements that are, for the time being, generally accepted by the relevant experts from those that are contested, under investigation, or rejected. Hence, although the corpus is not perfectly well defined, its vague margins are fairly narrow.

The process that leads to modifications of the corpus is based on strict standards of evidence that are an essential part of the ethos of science. Those who claim the existence of an as yet unproven phenomenon have the burden of proof. In other words, the corpus has high entry requirements. This is essential to protect us against the importation of false beliefs into science.

But as we noted in section ➋ The Fact – Value Distinction, scientific information is often used not only to guide the progress of science but also to guide practical decisions. As one example of this, studies of the anthropogenic greenhouse effect are used both to achieve more reliable scientific knowledge about what happens to the climate and to determine what practical decisions to take in climate policies. In this and many other cases, two decisions have to be based on the same scientific information: the intrascientific decision concerning what to believe and an extrascientific (practical) decision concerning what to do. These are two different decisions, although they make use of the same scientific data.

➋ *Figure 2.2* illustrates the practical use of scientific information (Hansson 2004c). The obvious way to use science for decision-guiding purposes is to employ information from the corpus (arrow 2). In many cases, this is all that we need to do. The high entrance requirements of the corpus have the important effect that the information contained in it is dependable enough to be relied on in almost all practical contexts. Only on very rare occasions do we need, for some practical purpose, to apply stricter standards of evidence than those that regulate corpus inclusion.

However, the high entry requirements of the corpus also have another, more complicating implication. On some occasions, evidence that was not strong enough for corpus entry may nevertheless be strong enough to have legitimate influence in some practical matters. To exemplify this, suppose that a preservative agent in baby food is suspected of having a negative health effect. The evidence weighs somewhat in the direction of there being an effect, and most scientists consider it to be more probable that the effect exists than that it does not. Nevertheless, the evidence is not conclusive, and the issue is still open from a scientific point of view. Considering what is at stake, it would be perfectly rational for a food company or a government agency to cease the use of the substance. Such a decision would have to be informed by scientific information that did not satisfy the criteria for corpus entry. More generally speaking, it would not seem rational – let alone morally defensible – for a decision-maker

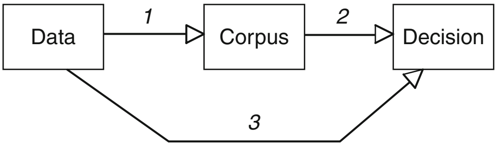


Fig. 2.2
The use of scientific data for decision-making

to ignore all preliminary indications of a possible danger that do not amount to full scientific proof. We typically wish to protect ourselves against suspected health hazards even if the evidence is much weaker than what is required for scientific proof. As was indicated in section The Fact – Value Distinction, in order to guide the type of decisions that we want to make, these decisions have to be based on standards of evidence that differ from the criteria used for intrascientific purposes. Evidence that is weaker than the requirements for corpus entry cannot influence decisions in the "standard" way that is represented in *Fig. 2.2* by arrows 1 and 2. In cases like this, we need to take a direct way from data to practical decision-making (arrow 3).

Just like the process represented by arrow 1, the bypass route represented by arrow 3 involves an evaluation of data against criteria of evidence. However, the two evaluation processes differ in being calibrated to different criteria for the required strength of evidence. The process of arrow 1 is calibrated to the standard scientific requirements, whereas that of arrow 3 is calibrated to criteria corresponding to the needs of a practical decision. However, the latter process is nevertheless in important respects a scientific one. From the viewpoint of philosophy of science, it is a challenge to clarify the nature of argumentation and decision processes like this that contain a mixture of scientific and policy-related components.

From a somewhat more practical point of view, it is essential to ensure that the bypass route does not lead to inefficient use of the available scientific information. In order to see what this requires, it is instructive to compare the processes represented by arrows 1 and 3. First of all, there should be no difference in the type of evidence that is taken into account. Hence, in the baby-food example, the same experimental and epidemiological studies are relevant for the intrascientific decision (arrow 1) and for the practical one (arrow 3). The evidence is the same, although it is used differently. Furthermore, the assessment of how strong the evidence is should be the same in the two processes. What differs is the *required* level of evidence for the respective purposes (Hansson 2008).

The term "precautionary principle" has often been used to designate the process illustrated by arrow 3 in our diagram (cf. section Terminological Clarification). But the need for a special principle can be put in doubt. Once it is recognized that the principle applies to practical decisions, it will be seen that the importation of practical values that this route makes possible is not only legitimate but in many cases also rationally required. 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. 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. More generally speaking, it is compatible with – and arguably required by – practical rationality that decisions be based on the available evidence even if it is incomplete.

Although the account given here is a reasonable ideal account, it is far from easy to implement in practice. If we want to take uncertain indications of toxicity seriously, then this has implications not only on how we interpret toxicological tests but also on our appraisals of more basic biological phenomena. If our main concern is not to miss any possible mechanism for toxicity, then we must pay serious attention to possible metabolic pathways for which there is insufficient proof. Such considerations in turn have intricate connections with various issues in biochemistry and, ultimately, we are driven to reappraise an immense number of empirical conclusions, hypotheses, and theories. Due to our cognitive limitations, this cannot in practice be done. In practice, we will have to rely on the corpus in most issues and use the detour (arrow 3) only in a limited number of selected issues. It remains to clarify how such partial adjustments are best made.

Philosophy of Technology

Since the nineteenth century, 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. However, most of the discussion on safety engineering is fragmented between different areas of technology. The same basic ideas or "safety philosophies" seem to have been developed more or less independently in different areas of engineering. Therefore, the same or similar ideas are often discussed under the different names for instance by chemical, nuclear, and electrical engineers. But a recent study has shown that there is much unity in this diversity. In spite of the terminological pluralism, and the almost bewildering number of similar or overlapping safety principles, much of the basic thinking seems to be the same in the different areas of safety engineering (Möller and Hansson 2008). In order to see what these basic ideas are, let us consider three major principles of safety engineering: inherent safety, safety factors, and multiple barriers.

Inherent Safety

Also called primary prevention, inherent safety consists in the elimination of a hazard. It is contrasted with secondary prevention that consists in reducing the risk associated with a hazard. For a simple example, consider a process in which inflammable materials are used. Inherent safety would consist in replacing them by noninflammable 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.

The major reason to prefer inherent safety to secondary prevention 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, if inflammable materials are present, some unforeseen chain of events can start a fire. 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 meeting 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 nonexplosive one or the inventories of explosive and inflammable substances have been drastically reduced, then the plant will be much less attractive to terrorists and therefore also a less likely target of attack (Hansson 2010a).

Most of the development of techniques for inherent safety has taken place within the chemical industry. Another major industry where inherent safety is often discussed is the nuclear industry, where it is 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 temperatures will not be high enough to trigger the release of radioactive fission products (Brinkmann et al. 2006).

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. The use of safety factors, i.e., numerical factors for dimensioning safety reserves, originated in the latter half of the nineteenth century (Randall 1976). Their use is now well established in structural mechanics and in its many applications in different engineering disciplines. Elaborate systems of safety factors have been specified in norms and standards (Clausen et al. 2006).

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. A safety factor is typically intended to protect against a specific 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.

According to standard accounts of structural mechanics, safety factors are intended to compensate for five major categories of sources of failure:

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.
5. Human error (e.g., in design) (Knoll 1976; Moses 1997).

The first two of these can in general be classified as variabilities, that is, they refer to the variability of empirical indicators of the propensity for failure. They are therefore accessible to probabilistic assessment (although these assessments may be more or less uncertain). The last three failure types refer to eventualities that are difficult or impossible to represent in probabilistic terms and, therefore, belong to the category of (non-probabilizable) uncertainty. They are not easily amenable to probabilistic treatment. It is, for instance, difficult to see how a calculation could be accurately adjusted to compensate self-referentially for an estimated probability that it is itself wrong. However, these difficulties do not make these sources of failure less important. Safety factors are used to deal both with those failures that can be accounted for in probabilistic terms and those that cannot (Doorn and Hansson 2011).

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.

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 can for instance 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 it against small movements in the rock and absorbs leaking radionuclides. 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 of 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 copper canister will fail for some unknown reason not included in the calculations. Then, hopefully, the radionuclides will stay in the bentonite, etc.

The notion of multiple safety barriers can also refer to safety barriers that are not placed in a spatial sequence like the defense 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 sulfide 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 a further illustration of this principle, suppose that a shipbuilder 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 lifeboats is above 1,000 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. (This is a lesson that we should have learned from the *Titanic* disaster.)

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, 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. 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 incapacitation.

These three principles of engineering safety – inherent safety, safety factors, and multiple barriers – are quite different in nature, but they have one important trait in common: They all aim at protecting us not only against risks that can be assigned meaningful probability estimates, but also against dangers that cannot be probabilized, such as the possibility that some unforeseen event triggers a hazard that is seemingly under control. It remains, however, to investigate more in detail the principles underlying safety engineering and, not least, to clarify how they relate to other principles of engineering design.

Ethics

Moral theorizing has mostly referred to the values of certain outcomes. The evaluation of uncertain outcomes is conventionally referred to decision theory, where it is treated as means-ends (instrumental) reasoning directed toward the attainment of given ends. Hence, moral philosophy refers primarily to human behavior in situations when the outcomes of actions are well defined and knowable. Decision theory takes assessments of these cases for given and derives from them assessments for situations involving risk and uncertainty. In this derivation, it operates, or so it is assumed, exclusively with criteria of rationality and does not add any new moral values. The dominating framework for these deliberations is expected utility theory.

Consider a person who risks a sleeping person's life by playing Russian roulette on her. In a moral assessment of this act, we need to consider (1) the set of consequences that will ensue if the person is killed, and (2) the set of consequences that will fall out if the person is not killed. In addition to this, we should also take into account (3) the act of risk imposition, which in this case takes the form of intentionally performing an act that may develop into an instance of either (1) or (2). In many people's moral appraisal of this misdeed, (3) has considerable weight. The act of deliberate risk-taking is perceived as a wrongdoing against the sleeping person, even if she is not killed and even if she never becomes aware of this episode or any disadvantage emanating from it. However, in the standard decision-theoretic approach, only (1) and (2) are taken into account (weighed according to their probabilities), whereas (3) is left out from the analysis.

This can be expressed with somewhat more precision in the following terminology: In a conventional decision-theoretical appraisal of this example, (1) and (2) will be replaced by their *closest deterministic analogs*. (1) is then evaluated as the act of discharging a fully loaded pistol at the sleeping person's head and (2) as that of letting off an unloaded pistol at her head. The composite act of performing what may turn out to be either (1) or (2) is assumed to have no other morally relevant aspects than those that are present in at least one of these two acts, both of which have well-determined consequences. The additional moral issues in (3), i.e., the issues concerning risk-taking per se, have no place in this account.

It is a general feature of this form of decision-theoretical analysis that if an act has moral aspects that are not present in the closest deterministic analog of any of its alternative developments, then these aspects are left out from the analysis. The crucial (but usually unstated) underlying assumption is that an adequate appraisal of an action under risk or uncertainty can be based on the values that pertain to its closest deterministic analogs. But as we saw from the Russian roulette example, this assumption has the disadvantage of excluding from our consideration the moral implications of risk-taking per se. This exclusion is unavoidable since risk-taking is by definition absent from the closest deterministic analogs that are used in the analysis.

The exclusion of risk-taking from consideration in most of moral theory can be clearly seen from the deterministic assumptions commonly made in the standard type of life-or-death examples that are used to explore the implications of moral theories. In the famous trolley problem, you are assumed to know that if you flip the switch, then one person will be killed, whereas if you do not flip it, then five other persons will be killed (Foot 1967). In Thomson's (1971) "violinist" thought experiment, you know for sure that the violinist's life will be saved if he is physically connected to you for 9 months, otherwise not. In Williams's (1973) example of Jim and the Indians, Jim knows for sure that if he kills 1 Indian, then the commander will spare the lives of 19 whom he would otherwise kill, etc. This is in stark contrast to ethical quandaries in real life, where action problems with human lives at stake seldom come with certain knowledge of the consequences of the alternative courses of action. Instead, uncertainty about the consequences of one's actions is a major complicating factor in most real-life dilemmas.

There are no easy answers to questions such as what risks you are allowed to impose on one person in order to save another or what risks a person can be morally required to take in order to save a stranger. These are questions that present themselves to us as moral questions, not as issues for decision-theoretical reckoning to take place after the moral deliberations have been finished. The exclusion of such issues from most discussions in moral philosophy has the effect of removing essential aspects of actual moral decision-making from our deliberations on moral theory. In order to include them, we have to give up the traditional assumption that the valuation of risk should take the form of applying decision-theoretical – and thus nonmoral – reasoning to values that refer to the moral evaluation of non-risky outcomes (in the form of closest deterministic analogs). Instead, we have to treat risk-taking per se as an object of moral appraisal.

The obvious way to develop an ethical theory of risk would be to generalize one of the existing ethical theories so that it can be effectively applied to situations involving risk. The problem of how to perform this generalization can be specified in terms of *the causal dilution problem*. It was presented by Robert Nozick (1974) as a problem for deontological ethics but is equally problematic for other moral theories.

- *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?

In utilitarian moral theory, 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, in stark contrast to common moral intuitions.

But actualism is not the standard decision-theoretical solution to the causal dilution problem for utilitarianism. The standard approach is to maximize expected 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.

This is a much more credible solution, and it has the important advantage of being 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).

However, this 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 leveled out as in the safety 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.

Even when the leveling-out argument for expected utility maximization is valid, compliance with this principle is not required by rationality. It is quite possible for a rational agent to refrain from minimizing total damage in order to avoid imposing high-probability risks on individuals. This can be exemplified with an example involving an acute situation in a chemicals factory (Hansson 1993). There are two ways to repair a serious gas leakage that threatens to develop into a disaster. 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. Hence, we have to go beyond expected utility theory in order to do justice to important moral intuitions about the rights of individuals.

As already mentioned, the causal dilution problem was originally formulated for rights-based theories by Robert Nozick. He asked: "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 the question as follows:

- *The causal dilution problem for deontological/rights-based moral theories:*
  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?

A rights-based moral theory can be extended to indeterministic cases by just prescribing that if A has a right that B does not bring about a certain outcome, then A also has a right that B does not perform any action that has a nonzero 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. For instance, you would not be allowed to drive a car in the town where I live since this increases my risk of being killed by you.

Hence, rights and prohibitions have to be defeasible so that they can be canceled 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 canceled. 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.

Finally, let us consider contract theories. They 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 fraught with problems. 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 other 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 (any type of) fuel 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 risks, 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.

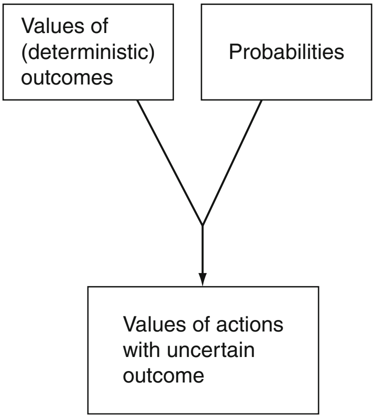
We are left then with hypothetical consent. However, as the debate following John Rawls's *Theory of Justice* has shown, there is no single decision rule for risk and uncertainty that all participants in a hypothetical initial situation can be supposed to adhere to (Hare 1973; Harsanyi 1975). It remains 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 (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). Apparently, this has not been done and, hence, contract theory does not either have a solution to the causal dilution problem.

The difficulties that we encounter when trying to solve the causal dilution problem within the frameworks of the 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 if the moral appraisals of actions with deterministic outcomes are given, then we can derive from them moral appraisals of actions whose outcomes are probabilistic mixtures of such deterministic outcomes. In utilitarian approaches, it is furthermore assumed that probabilities and (deterministic) utilities are all the information that we need (➔ *Fig. 2.3*). 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 1985a). 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 (➔ *Fig. 2.4*).

Based on this analysis, the causal dilution problem can be replaced by a *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?

The defeasance problem is a truly moral problem, not a decision-theoretical one. As far as I can see, it is the central ethical issue that a moral theory of risk has to deal with. Obviously, there are many ways to approach it, only few of which have been developed. It remains to investigate and compare the various solutions that are possible. My own preliminary solution refers to reciprocal exchanges of risks and benefits. Each of us takes risks in order to obtain benefits for ourselves. It is beneficial for all of us to extend this practice to mutual exchanges of risks and benefits. If others are allowed to drive a car, exposing you to certain risks, then in exchange you


Values of (deterministic) outcomes
Probabilities
Values of actions with uncertain outcome

■ **Fig. 2.3**
The traditional account of how values of indeterministic outcomes can be derived

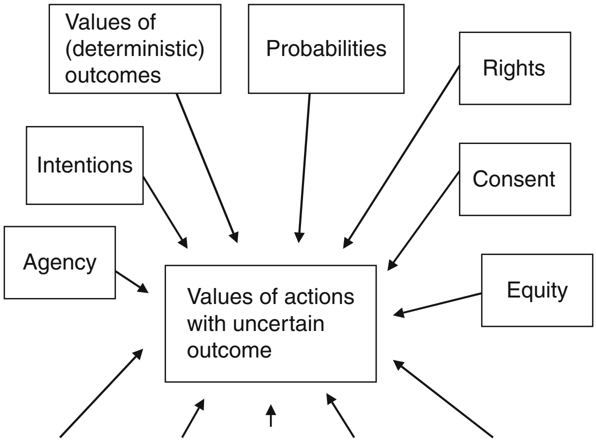

Values of (deterministic) outcomes
Probabilities
Rights
Intentions
Consent
Agency
Values of actions with uncertain outcome
Equity

Fig. 2.4
A less incomplete picture of the influences on the values of indeterministic options

are allowed to drive a car and expose them to the corresponding risks. This (we may suppose) is to the benefit of all of us. In order to deal with the complexities of modern society, we also need to apply this principle to exchanges of different types of risks and benefits. We can then *regard exposure of a person to a risk as acceptable if it is part of a social system of risk-taking that works to her advantage and gives her a fair share of its advantages.*

This solution is only schematic, and it gives rise to further problems that need to be solved. Perhaps the most difficult of these problems is how to deal with large differences among the members of society in their assessments of risks and benefits. But with the approach presented here, we have, or at least so I wish to argue, a necessary prerequisite in place, namely, the right agenda for the ethics of risk. According to traditional risk analysis, in order to show that it is acceptable to impose a risk on Ms. Smith, the risk-imposer only has to give sufficient reasons for accepting the risk as such, as an impersonal entity. According to the proposal just presented, this is not enough. The risk-imposer has to give sufficient reasons why Ms. Smith – as the particular person that she is – should be exposed to the risk. This can credibly be done only by showing that this risk exposure is part of some arrangement that works to her own advantage. For a more detailed discussion of this approach, see Hansson (2003b).

Philosophy of Economics

Risks have a central role in economic theory, and there are obvious parallels between the problems of economic risk and the problems concerning other types of risk such as risks to health and the environment. Let us have a look at two interesting issues in the philosophy of economic risk: the aggregation problem and the problem of positive risk-taking.

The aggregation problem concerns how we compare risks accruing to different individuals. Standard risk analysis follows the principles of classical utilitarianism. All risks are summed up in one and the same balance irrespectively of whom they accrue to. Thus, all risks are taken to be fully comparable and additively aggregable. In risk-benefit analysis, benefits are added in the same way and, finally, the sum of benefits is compared to the sum of risks in order to determine whether the total effect is positive or negative. In such a model, just as in classical utilitarianism, individuals have no other role than as carriers of utilities and disutilities, the values of which are independent of whom they are carried by.

An obvious alternative to this utilitarian approach is to treat each individual as a separate moral unit. Then risks and benefits pertaining to one and the same person can be weighed against each other, whereas risks and benefits for different persons are added or otherwise aggregated since they are considered to be incomparable. Such "individualistic" risk weighing is quite different from the total aggregations that are standard in risk analysis. But individualistic risk weighing dominates in medicine. It is applied for instance in ethical evaluations of clinical trials. It is an almost universally accepted principle in research ethics that a patient should not be included in a clinical trial unless there is genuine uncertainty on whether or not participation in the trial is better for her than the standard treatment that she would otherwise receive. That her participation is beneficial for others (such as future patients) cannot outweigh a negative net effect on her own health; in other words, her participation has to be supported by an appraisal that is restricted to risks and benefits for herself (London 2001; Hansson 2004a).

The two traditions in risk assessment differ in the same way as the "old" and "new" schools of welfare economics. In Arthur Pigou's so-called old welfare economics, the values pertaining to different individuals are added up to one grand total. This is also the approach of mainstream risk analysis. The new school in welfare economics that dominates mainstream economics since the 1930s refrains from adding individual values. Instead, it treats the welfare of different individuals as incomparable. This became the standard approach after Lionel Robbins had shown how economic analysis can dispense with interpersonal comparability (Pareto optimality is the central tool needed to achieve this). The individualist approach that was exemplified above with clinical trials is based on the same basic principles as those applied by the new school in welfare economics (Hansson 2006a).

Mainstream risk analysis and mainstream economics represent two extremes with respect to interindividual comparisons. The aggregations of total risk that are performed routinely in risk analysis stand in stark contrast to the consistent avoidance of interindividual comparisons that is a major guiding principle in modern economics. This difference also has repercussions in the ideological uses of the respective disciplines. It is an implicit message of risk-benefit analysis that a rational person should accept being exposed to a risk if this brings greater benefits for others. The implicit message of modern (new school) welfare economics is much more appreciative of self-interested behavior.

The issue of *positive risk-taking* appears to be more or less specific for economic risks. Risk is by definition undesirable, and we expect a rational person to avoid risk as far as possible. But in economics, risk-taking is often considered to be desirable. The capitalist's risk taking is acknowledged as essential for the efficiency of a capitalist system, and it is also taken to justify the owner's prerogative to exert the ultimate control over companies and to reap the profits. As said already by Adam Smith in his *Wealth of Nations*, "something must be given for the profits of the undertaker of the work who hazards his stock in this adventure" (Smith [1776] 1976, p. 1:66).

The risk taking that Smith referred to was a substantial one, namely the risk of bankruptcy. According to Smith, becoming bankrupt is "perhaps the greatest and most humiliating calamity which can befal an innocent man" (Smith [1776] 1976, p. 1:342). This was the risk that the capitalist was supposed to take and to be compensated for. Its seriousness was essential for Smith's argument as we can see from his negative attitude to arrangements that reduce the risk from that of bankruptcy to that of losing the invested capital. The most important such arrangement was the joint-stock company (with limited liability) to which Smith was decidedly averse (Smith [1776] 1976, p. 2:741).

But since Smith's time capitalism has been fundamentally transformed, two major reductions in capitalist risk-taking have taken place. The first of these occurred in the latter half of the nineteenth century when corporations with limited liability became the dominant legal form of private companies in the industrialized parts of the world (Handlin and Handlin 1945; Prasch 2004). Due to the massive spread of limited liability, personal risk taking in most major industrial and financial endeavors was brought down from bankruptcy to loss of the original investment, which was exactly what Adam Smith had warned against.

The second reduction in economic risk-taking took place about 100 years later. Beginning in the late twentieth century, private investment in companies has to an increasing extent been mediated by institutions and funds that diversify their securities in a sophisticated way to reduce risk taking. Portfolio theory and modern financial marketplaces have combined to make risk spreading much more efficient than what was previously possible. Today, an owner who has applied prudent risk spreading only runs a risk that approximates the general background risk of the economy. In terms of risk-taking, his situation is arguably less akin to that of businesspeople risking everything they own than to that of the nineteenth century landlords who according to John Stuart Mill "grow richer, as it were in their sleep, without working, risking, or economizing" (Mill [1848] 1965, pp. 3:819–820).

Risk-spread ownership has thoroughly transformed the economic system, but its philosophical implications have not been much discussed. It is for instance not unreasonable to ask what effects this development has on the legitimacy of the owner's prerogative that was previously based at least in part on the risk-taking role of owner.

Political Philosophy

Although risk and uncertainty are ubiquitous in political and social decision-making, there has been very little contact between risk studies and more general studies of social decision processes. Public discussions of risk are dominated by a way of thinking that is markedly different from how democratic decision-making is commonly discussed. We can see this from the frequent references in discussions on risk to three terms that are not much used in general discussions on democratic decision-making.

"Consent" is one of these words. Consent by the public is often taken to be the goal of public communications on risk. The following quotation is not untypical:

> ► Community groups have in recent years successfully used zoning and other local regulations, as well as physical opposition (e.g., in the form of sitdowns or sabotage), to stall or defeat locally unacceptable land uses. In the face of such resistance, it is desirable (and sometimes even necessary) to draw forth the consent of such groups to proposed land uses (Simmons 1987, p. 6).

To consent means in this context "voluntarily to accede to or acquiesce in what another proposes or desires" (Oxford English Dictionary). This is very different from the role of the citizen as a decision-maker in a democracy.

The second of these words is "acceptance." The goal of risk communication is often taken to be public acceptance of a presumedly rational act of risk-taking. Hence, in discussions on the siting of potentially dangerous industries, "public acceptance" is usually taken to be the crucial criterion. This usage signals the same limitation in public participation as the word "consent."

The third word is "trust." Much of the discussion in the risk-related academic literature on the relationship between decision-makers and the public takes the public's trust in decision-makers to be the obvious criterion of a well-functioning relationship. This is, again, very different from discussions on democracy in political philosophy. In a democratic constitution, the aim is the public's democratic control over decision-makers, rather than their trust in them (Hayenhjelm 2007).

In contrast to the limited approach to public participation in risk decisions that is indicated by the keywords "consent," "acceptance," and "trust," let us consider the ideal of full democratic participation in decision-making. This ideal was very well expressed by Condorcet in his vindication of the French constitution of 1793. Condorcet divided decision processes into three stages. In the first stage, one "discusses the principles that will serve as the basis for decision in a general issue; one examines the various aspects of this issue and the consequences of different ways to make the decision." At this stage, the opinions are personal, and no attempts are made to form a majority. After this follows a second discussion in which "the question is clarified, opinions approach and combine with each other to a small number of more general opinions." In this way, the decision is reduced to a choice between a manageable set of alternatives. The third stage consists of the actual choice between these alternatives (Condorcet [1793] 1847, pp. 342–343).

The discussion on public participation in issues of risk has mostly been restricted to Condorcet's third stage, and – as we have seen – often also to a merely confirming or consenting role in that stage. This approach is obviously untenable if we wish to see decisions on risk in the full context of public decision-making in a democratic society. Without public participation at all three stages of the decision-making process, risk issues cannot be dealt with democratically. Therefore, the discussion needs to be shifted away from special procedures for dealing with risk. Instead our focus should be on how the special characteristics of risk-related issues can best be dealt with in our general decision-making processes.

Further Research

Far from being an unusual oddity in philosophy, the topic of risk connects directly to central issues in quite a few subdisciplines of philosophy. In some of these subdisciplines, a significant amount of research on risk has already taken place. In others, only the first steps toward systematic studies of risk have been taken. But in all of them, important philosophical issues related to risk remain unexplored. The following are ten of the most important issues for further research that have been pointed out above:

- When is trust in experts on risks justified, and when is distrust irrational?
- How remote possibilities of disaster should be taken seriously?
- How can we account for probabilistic reasoning that seems rational but is not compatible with the standard theory of probability?
- To what extent, and in what ways, should practical consequences have influence on scientific assessments of risk?
- How can the principles of safety engineering be accounted for, and how do they relate to probabilistic risk analysis?

- How can a risk or safety analysis take into account the possibility that the analysis itself is wrong?
- When is it ethically permissible to expose another person to a risk?
- How can utilitarianism be extended or adjusted so that it provides us with a reasonable account of the ethics of risk-taking?
- Do we have a right not to be exposed to risks and, in that case, when can it be overruled?
- What role should those exposed to a risk have in democratic decisions on that risk?

References

Ahteensuu M (2008) In dubio pro natura? PhD thesis in philosophy, University of Turku

Blackburn S (1973) Reason and prediction. Cambridge University Press, Cambridge

Brinkmann G, Pirson J, Ehster S, Dominguez MT, Mansani L, Coe I, Moormann R, Van der Mheen W (2006) Important viewpoints proposed for a safety approach of HTGR reactors in Europe. Final results of the EC-funded HTR-L project. Nucl Eng Des 236:463–474

Burgos R, Defeo O (2004) Long-term population structure, mortality and modeling of a tropical multi-fleet fishery: the red grouper Epinephelus morio of the Campeche bank, Gulf of Mexico. Fish Res 66:325–335

Carlson E (1995) Consequentialism reconsidered. Kluwer, Dordrecht/Boston

Clausen J, Hansson SO, Nilsson F (2006) Generalizing the safety factor approach. Reliab Eng Syst Saf 91:964–973

Cohen BL (2003) Probabilistic risk analysis for a high-level radioactive waste repository. Risk Anal 23:909–915

Condorcet ([1793] 1847) Plan de Constitution, presenté a la convention nationale les 15 et 16 février 1793. Oeuvres 12:333–415

Cox R, Winkler R (2010) Spill may prompt energy mergers. *New York Times* June 2, 2010. http://www.nytimes.com/2010/06/03/business/03views.html. Accessed 9 June 2011

Cranor CF (1997) The normative nature of risk assessment: features and possibilities. Risk Health Saf Environ 8:123–136

Cranor CF, Nutting K (1990) Scientific and legal standards of statistical evidence in toxic tort and discrimination suits. Law Philos 9:115–156

Donahoe FJ (1969) 'Anomalous' water. Nature 224:198

Doorn N, Hansson SO (2011) Should safety factors replace probabilistic design? Philos Technol 24:151–168

Feleppa R (1981) Epistemic utility and theory acceptance: comments on Hempel. Synthese 46:413–420

Fischhoff B (1977) Perceived informativeness of facts. Hum Percept Perform 3(2):349–358

Fischhoff B, Lichtenstein S, Slovic P, Derby SL, Keeney RL (1981) Acceptable risk. Cambridge University Press, Cambridge

Foot P (1967) The problem of abortion and the doctrine of the double effect. Oxford Rev 5:5–15. Reprinted in her *Virtues and Vices*, Oxford: Basil Blackwell, 1978

Handlin O, Handlin MF (1945) Origins of the American business corporation. J Econ Hist 5:1–23

Hansson SO (1993) The false promises of risk analysis. Ratio 6:16–26

Hansson SO (1995) The detection level. Regul Toxicol Pharmacol 22:103–109

Hansson SO (1996) Decision-making under great uncertainty. Philos Soc Sci 26:369–386

Hansson SO (1998) Setting the limit: occupational health standards and the limits of science. Oxford University Press, New York/Oxford

Hansson SO (2002) Replacing the no effect level (NOEL) with bounded effect levels (OBEL and LEBEL). Stat Med 21:3071–3078

Hansson SO (2003a) Are natural risks less dangerous than technological risks? Philos Nat 40:43–54

Hansson SO (2003b) Ethical criteria of risk acceptance. Erkenntnis 59:291–309

Hansson SO (2004a) Weighing risks and benefits. Topoi 23:145–152

Hansson SO (2004b) Fallacies of risk. J Risk Res 7:353–360

Hansson SO (2004c) Philosophical perspectives on risk. Techne 8(1):10–35

Hansson SO (2004d) Great uncertainty about small things. Techne 8(2):26–35

Hansson SO (2005) Seven myths of risk. Risk Manage 7(2):7–17

Hansson SO (2006a) Economic (ir)rationality in risk analysis. Econ Philos 22:231–241

Hansson SO (2006b) How to define – a tutorial. Princípios, Revista de Filosofia 13(19–20):5–30

Hansson SO (2007a) Philosophical problems in cost-benefit analysis. Econ Philos 23:163–183

Hansson SO (2007b) Values in pure and applied science. Found Sci 12:257–268

Hansson SO (2008) Regulating BFRs – from science to policy. Chemosphere 73:144–147

Hansson SO (2009a) Should we protect the most sensitive people? J Radiol Prot 29:211–218

Hansson SO (2009b) Measuring uncertainty. Studia Log 93:21–40

Hansson SO (2010a) Promoting inherent safety. Process Saf Environ Prot 88:168–172

Hansson SO (2010b) Past probabilities. Notre Dame J Formal Logic 51:207–233

Hansson SO, Rudén C (2006) Evaluating the risk decision process. Toxicology 218:100–111

Hare RM (1973) Rawls's theory of justice. Am Philos Quart 23:144–155 and 241–252

Harsanyi JC (1975) Can the maximin principle serve as a basis for morality – critique of Rawls, J theory. Am Pol Sci Rev 69(2):594–606

Harsanyi JC (1983) Bayesian decision theory, subjective and objective probabilities, and acceptance of empirical hypotheses. Synthese 57:341–365

Hayenhjelm M (2007) Trusting and taking risks: a philosophical inquiry. Ph.D. thesis, KTH, Stockholm

Hempel CG (1960) Inductive inconsistencies. Synthese 12:439–469

International Organization for Standardization (2002) Risk management – vocabulary – guidelines for use in standards, ISO/IEC Guide 73/2002

Knoll F (1976) Commentary on the basic philosophy and recent development of safety margins. Can J Civ Eng 3:409–416

Krewski D, Goddard MJ, Murdoch D (1989) Statistical considerations in the interpretation of negative carcinogenicity data. Regul Toxicol Pharmacol 9:5–22

Leisenring W, Ryan L (1992) Statistical properties of the NOAEL. Regul Toxicol Pharmacol 15:161–171

Levi I (1962) On the seriousness of mistakes. Philos Sci 29:47–65

Levi I (1973) Gambling with truth. MIT Press, Cambridge, MA

London AJ (2001) Equipoise and international human-subjects research. Bioethics 15:312–332

Lopez RE, Holle RL (1998) Changes in the number of lightning deaths in the United States during the twentieth century. J Climate 11:2070–2077

MacLean D (ed) (1985) Values at risk. Rowman & Allanheld, Totowa

Mill JS ([1848] 1965) The principles of political economy with some of their applications to social philosophy. In: Robson JM (ed) Collected works of John Stuart Mill, vol 2–3. University of Toronto Press, Toronto

Miller CO (1988) System safety. In: Wiener EL, Nagel DC (eds) Human factors in aviation. Academic, San Diego, pp 53–80

Möller N, Hansson SO (2008) Principles of engineering safety: risk and uncertainty reduction. Reliab Eng Syst Saf 93:776–783

Möller N, Hansson SO, Peterson M (2006) Safety is more than the antonym of risk. J Appl Philos 23(4):419–432

Moses F (1997) Problems and prospects of reliability-based optimisation. Eng Struct 19:293–301

National Research Council (1983) Risk assessment in the federal government: managing the process. National Academy Press, Washington, DC

Nozick R (1974) Anarchy, state, and utopia. Basic Books, New York

O'Riordan T, Cameron J (eds) (1994) Interpreting the precautionary principle. Earthscan, London

O'Riordan T, Cameron J, Jordan A (eds) (2001) Reinterpreting the precautionary principle. Cameron May, London

Prasch RE (2004) Shifting risk: the divorce of risk from reward in American capitalism. J Econ Issues 38:405–412

Rabinowicz W (2002) Does practical deliberation crowd out self-prediction? Erkenntnis 57:91–122

Randall FA (1976) The safety factor of structures in history. Prof Saf 1976(January):12–28

Roeser S (2006) The role of emotions in judging the moral acceptability of risks. Saf Sci 44:689–700

Royal Society (1983) Risk assessment. Report of a Royal Society Study Group, London

Rudén C, Hansson SO (2008) Evidence based toxicology – 'sound science' in new disguise. Int J Occup Environ Health 14:299–306

Sandin P (1999) Dimensions of the precautionary principle. Hum Ecol Risk Assess 5:889–907

Shrader-Frechette K (1991) Risk and rationality: philosophical foundations for populist reforms. University of California Press, Berkeley

Simmons J (1987) Consent and fairness in planning land use. Bus Prof Ethics J 6(2):5–20

Smith A ([1776] 1976) An inquiry into the nature and causes of the wealth of nations. In: Campbell RH, Skinner AS, Todd WB (eds) The Glasgow edition of the works and correspondence of Adam Smith, vol 2. Clarendon, Oxford

Spohn W (1977) Where Luce and Krantz do really generalize Savage's decision model. Erkenntnis 11:113–134

Tench W (1985) Safety is no accident. Collins, London

Thomson JJ (1971) A defense of abortion. Philos Public Aff 1:47–66

Thomson JJ (1985a) Imposing risk. In: Gibson M (ed) To breathe freely. Rowman & Allanheld, Totowa, pp 124–140

Thomson PB (1985b) Risking or being willing: Hamlet and the DC-10. J Value Inquiry 19:301–310

Walton DN (1987) Informal fallacies: towards a theory of argument criticisms. J. Benjamins, Amsterdam

Williams B (1973) A critique of utilitarianism. In: Smart JJC, Williams B (eds) Utilitarianism: for and against. Cambridge University Press, London