Emotions Within the Bounds of Pure Reason: Emotionality and Rationality in the Acceptance of Technological Risks

Dieter Birnbacher

1 Discrepancies in the Assessment of Technological Risks by Experts and Laypeople

One thing that the last decades of technological innovation have shown is that there is a sharp divergence between the perception, assessment and acceptance of technological risks by laypeople and by scientific and technological experts (cf. Slovic et al. 1979; Renn and Zwick 1997, 87 ff.; Renn 2008). It has become evident that the general public, in judging the acceptability of technologies, makes use of mental "heuristics" (Tversky and Kahneman 1974) that differ significantly from those employed by experts. This has become obvious in the case of energy production by nuclear power and of transgenic crop plants in agriculture, two cases that have lead to well-known and enduring social and political schisms. Moreover, it has lead to a deep and lingering distrust between representatives of science, technology and industry on the one side and great portions of the general public on the other. A view widespread among scientists, engineers and industrialists is that some if not even most factors underlying the public acceptance of technologies do not mirror any objective features of these technologies and their prospects but are merely "psychological" and, in the last analysis, irrational. Public attention, according to this view, is focussed on risks that hardly ever affect the life of those who fear them, whereas other, far more dangerous risks, are ignored. The point is nicely expressed by Peter Sandman's dictum that "the risks that kill you are not necessarily the risks that anger and frighten you" (cf. Jungermann and Slovic 1993, p. 80). In this view, the fears, anxieties and worries manifested by the stubborn non-acceptance of technologies like nuclear power and gene-food express reactions that may be psychologically understandable (or at least explainable) but only thinly related to the facts. One form this criticism assumes is that these reactions, however firmly embedded in the human psyche, are "purely emotional". From this diagnosis it is only a short step to the conclusion that it is wrong, and possibly even irresponsible, to give these emotions a role to play in technological planning and development.

D. Birnbacher (✉)
Department of Philosophy, Düsseldorf University, Düsseldorf, Germany
e-mail: Dieter.Birnbacher@uni-duesseldorf.de

S. Roeser (ed.), *Emotions and Risky Technologies*, The International Library of Ethics, Law and Technology 5, DOI 10.1007/978-90-481-8647-1_11,

Accordingly, emotions are a factor politicians have to deal with, not technological planners. Emotions are a matter of strategy, not something that should enter into judgements about what is acceptable or unacceptable.

A similar discrepancy can be found on the level of theory. The scientific debate about acceptable risk largely recapitulates the public debate, with a rational choice approach on the line of a socially extended Bayesianism on the one side and a theory of "qualitative characteristics" (Slovic et al. 1979, p. 36) on the other. On the Bayesian side, risks are assessed from a risk-neutral perspective, and only the two "classical" dimensions of risk are taken into account, (negative) value of possible outcomes and outcome probability. On the other side, a risk-averse perspective is held to be more appropriate and a number of further aspects are brought into the picture, among them voluntariness, potential for catastrophe, naturalness and distributional aspects. There are obvious advantages in the Bayesian approach from a practical point of view. It makes the calculation and comparison of risks considerably more straightforward than an approach that takes into account a large number of (probably weighted) qualitative factors. Whereas in Bayesianism comparisons between the benefits and risks of technological alternatives involve comparing one single value, expected utility, risk assessments by qualitative standards are inherently more complicated, intransparent and controversial. Whereas expected values are derived by a simple arithmetic operation (the summing of the products of each individual outcome with its probability), risk assessments by qualitative factors usually fail to make explicit which factors are included and what relative weight is given to each of them, thus lowering the chances of a systematic, authoritative and consensual assessment. The moot question, however, is how these more or less formal advantages of Bayesianism compare with its material inadequacies and especially its inability to pay tribute to some of the factors paramount in the acceptance or non-acceptance of a technology by the general public. One of these inadequacies is that Bayesianism assesses risky activities by a purely additive criterion and is indifferent between risks with great magnitude and low probability and risks with low magnitude and high probability. That means that it is unable to account for at least one of the structural asymmetries that are relevant in risk acceptance. For the general public, it matters a lot whether the risk profile of a technology includes severe but infrequent accidents or frequent but trivial accidents. And since Bayesianism balances costs with benefits irrespective of the identities of their subjects it cannot take account of distributional features. It cannot distinguish between cases in which the harm resulting from a technology befalls those who benefit from it and others in which it is imposed on third parties.

The conflict between these viewpoints has persisted for decades, and with a high degree of polarization. Both parties insist on the rationality of their respective perspective and do not hesitate to accuse the other of irrationality, blindness and yielding to "pure emotion", often, ironically, with a good deal of emotion. On the Non-Bayesian side, the same "emotional" factors in risk perception and assessment – such as the preference for the natural and the familiar against the technical and unfamiliar – are interpreted as indicators of intuitive wisdom that are denounced on the Bayesian side as sentimentality and ignorance. The tendency to substitute

unanalysed intuition for systematic decision-theoretical approaches is particularly prominent in some variants of "bounded rationality". Gigerenzer and Selten, for one, have pleaded for an approach to complex choices that no longer attempts to integrate qualitative factors into the Bayesian scheme by "tinkering with the utility or probability function, while at the same time retaining the ideal of maximization or optimization" (Gigerenzer and Selten 2001, p. 3). Instead, they recommend to question the monopolistic claims of Bayesianism and to adopt a radically new approach that starts from the heuristics factually employed in choices, mostly on an unconscious and "instinctive" basis, the "adaptive toolbox". In this way, decisions are not only more readily available but also better adapted to their respective contexts. The interesting point for our purposes is that, according to these authors, some of the emotions share these advantages. There are circumstances in which emotions are a more reliable guide to rationality than deliberation and calculation, as is shown in the context of animal behaviour:

> Emotions like disgust or parental love can provide effective stopping rules for search and a means for limiting search spaces. In particular, for important adaptive problems such as food avoidance, an emotion of disgust, which may be acquired through observation of conspecifics, can be more effective than cognitive decision making. (Gigerenzer and Selten 2001, p. 9)

This plea for "relying on instinct" will, however, make no impression on the Bayesian who will be quick to point out that these authors measure the success of the method by an independent standard of rationality ("effectiveness") that is not supposed to be determined by emotional factors. Furthermore, Bayesians will be ready to concede that in many real-life contexts emotions may indeed be more reliable guides to behaviour than deliberation and calculation, but that this does in no way detract from the necessity to establish a standard of rationality on which the instrumentality of "rules of thumb", "gut-feelings", "instincts" and the like can be judged. The standard invoked in judging whether a "tool" from the "adaptive toolbox" is truly "adaptive" (instead of "maladaptive") cannot itself be justified with reference to the tools from the box. After all, Gigerenzer and Selten are far from denying that there are plenty of occasions in which the ready-made "tools" commonly employed in practical matters are of little help, or even positively misleading.

The Bayesian, however, will go further and maintain that there is evidence from a great number of empirical studies that the effects of emotional factors in judgements on risks are predominantly of a distorting kind. This evidence is particularly strong since it does not have to rely on some presupposed standard of rationality but points to features that are incompatible with any coherent standard:

1. There is strong evidence that risk perception and risk assessment are to a high degree culture-relative and depend on non-cognitive factors like habitualization and dissonance reduction. The most plausible explanation for the differences in the judgements about the risks of nuclear energy between the populations of France and Germany, or in the judgements about transgenic food between the populations of Europe and the United States is not that one of these publics is better informed than the other or in a better position to draw the right conclusions

from this information, but that judgements are harmonized with what has become part of their habitual environment. The basis of this harmonization of facts and values is not coherence in the sense of a unified cognitive picture of the world but coherence in the sense of dissonance reduction, the unconscious striving for a view of the world in which action, cognition and emotion are in harmony with each other.

2. There is evidence that there are sharp discontinuities in popular risk perception. There seems to be a threshold in probability (sometimes identified with the probability 10^{-5}) beneath which rare events are judged to be sufficiently improbable to be no longer a cause of concern. We do not care about catastrophes that are "morally (or practically) impossible". However, we care a lot about catastrophic risks the probabilities of which are only slightly higher. Whereas risks below the threshold are discounted and ignored, risks above the threshold are judged as particularly dangerous. On the theoretical level, this dichotomy is mirrored in Nicholas Rescher's theory of acceptable risk that makes a similarly sharp contrast between risks of catastrophe too trivial to be given any attention, and non-trivial risks of catastrophe that must be avoided at all costs, thus making the overall negative value of risks leap from zero to infinite at the threshold (cf. Rescher 1983, p. 76).

What is blurred by the polarization of Bayesians and Non-Bayesians is the prospect of finding a compromise that does justice, as far as it goes, to both sides and attempts to combine what is adequate in both perspectives. This is what I propose to do in the following. The question, in my view, is not whether the one or the other approach is the correct one, but how far the discrepancy between them is real or apparent, and how far the Bayesian model can do justice to the attitudes behind the qualitative risk features, provided these can be shown to encapsulate aspects of rationality not covered by the standard variants of Bayesianism. Before looking at this question, we should, however, first clarify in what exact way emotions can be said to influence judgements about risks and how far they are acceptable.

2 The Role of Emotion in Judgements About Acceptable Risk

"Emotion" serves as an umbrella concept for a large variety of mental items, and in discussing the role of emotion in judgements about risks one should make clear what kind of item one has in mind. The question as to what extent emotions can serve as avenues to the truth where reason is blind, and to what extent they distort and mislead judgement, depends, among others, on what category of emotion one is thinking of. There are two categories of emotion for which it is more or less obvious that their influence on judgement is mainly a distorting one and that they justify, so far as it goes, the view of philosophers like the Stoics for whom emotions were, in the first place, "disturbances" – not only in the sense that emotions disturb one's peace of mind but also in the sense that they disturb one's sound judgement. These

two categories are emotions as *temporally extended moods* and emotions as *episodic forms of excited feeling.*

It is a fact well-known from experience that both kinds of emotional states tend to weaken our faculty of judgement and to engender pessimism or optimism, as the case may be, largely uncontrolled to the facts. In a depressive mood, things appear threatening that are normally seen as indifferent or easy to cope with. In an elevated mood, the "bright side of life" dominates inner experience, shielding from view what might blemish the harmonious picture. Similar observations can be made in cases of acute emotion. In an experiment on the impact of affect on risk assessment Johnson and Tversky investigated to what extent fear, anxiety and worry caused by the reading of dramatic stories have an influence on estimates of the frequencies with which certain risks are believed to occur. The subjects were made to read stories with vivid and detailed portrayals of deaths deliberately designed to induce anxiety and worry. It turned out that there was a considerable influence of the emotional states induced by the stories on the estimates of the frequency of certain risks though the risks had nothing do to with the dramatic events in the stories. The frequency estimates of the group with induced negative mood was significantly higher than those of the control group. An analogous experiment with a group with induced positive mood showed an even higher inverse effect (Johnson and Tversky 1983, p. 28).

These two categories of emotions, then, can be left to themselves. They are not what is at stake in the debate about the rationality and irrationality of emotion in risk assessment. The kind of emotion that is at stake can be characterised by the following features:

1. The emotions that are candidates for being honoured in judgements about acceptable risk are of the nature of emotional attitudes rather than of the nature of episodic emotions. Ontologically, they are dispositions rather than events of processes. "Emotional attitude" means that the attitude is not purely cognitive and that its content cannot be adequately expressed by a purely descriptive statement. In this sense, a belief that something is the case is purely cognitive, but not the hope or fear that something is the case. Emotional attitudes necessarily include an element of evaluation, positive or negative.
2. The emotional attitudes in question are intentionally related to the risk in question. They are closely related to the thought of the risk and are not, as the emotional states in the experiments of Johnson and Tversky, induced by independent factors.
3. The emotional content of the attitude is largely, or wholly, unconscious or preconscious. The subject is not aware of this emotional content, or is made aware of this content only by directing attention at it.
4. Emotional content can enter into judgements about risks at several different stages and influence components of these judgements to different degrees. Some emotional factors work primarily on the estimate of frequencies, others primarily on the estimate of the negative values of risks. One mechanism involving emotional factors and primarily influencing frequency estimates is the "availability

heuristic" (Slovic et al. 1979, p. 15). It makes that a certain event is judged as likely or frequent to the degree that it is easy to imagine or recall, which in turn depends, among others, on emotional factors like surprise, irritation and dread. Other mechanisms involving emotional factors work on the valuation component, such as the mechanism that makes risks that are a threat to ourselves look more frightening than risks that are a threat to other people.

It is exactly these emotional factors in the attitudes towards risky technologies that the dispute between Bayesians and Non-Bayesians is about. Both differ radically in their views about the compatibility of these factors with rationality as a standard of judgement and guide to action. It should not, however, be overlooked that there is also a great area of agreement between the parties as far as the compatibility and incompatibility of these factors with rationality is concerned. This agreement pertains primarily to components of judgements about risks that are consequences of more general features and independent of the probabilistic nature of the items judged.

Neither party denies that at least one non-cognitive factor is not only compatible with rational judgements about acceptable risk but even required by them, namely the non-cognitive factors entering in the estimates of the moral and non-moral value of the possible adverse events and their consequences. Statements about risks (like statements about chances) are, as a rule, value judgements and not purely descriptive. They imply that certain possible outcomes are judged to be of negative value. If, however, value judgements, as I think they do, necessarily contain a non-cognitive element and express a "pro-" or "con-attitude" with at least a minimum of emotional content, emotional attitudes are, as it were, interwoven with risk judgements and inseparable from them. By necessarily referring to values of some kind or other, the very concept of risk seems unintelligible if defined in a purely descriptive way. Risk judgements are inherently value judgements and go beyond what can be attained by a purely cognitive approach.

This leaves open the possibility that the value judgements contained in judgement about risk (and therefore about acceptable risk) may be influenced by more specific emotional factors that make them "irrational" in one of various ways. The most important of these factors are *involvement* or *ego-preference*, the tendency to judge risks to be inacceptable in relation to the extent one is threatened by them in one's own person, and the tendency to discount adverse events according to social distance and to distance in time. I will not here discuss if there are circumstances under which these tendencies, which play an important role in common-sense judgements about risks, can be given a rational justification (cf. Birnbacher 2003). Let it suffice to say that if these tendencies are criticized as "irrational" this is not because this follows from the fact that they are non-cognitive or emotional tendencies, but because they constitute specific emotional tendencies that tend to distort judgements about the acceptability (from a moral and impersonal perspective) of actions and strategies not only in the domain of risks but likewise in non-probabilistic domains.

Furthermore, both parties are agreed that there are a number of non-cognitive factors that are incompatible with the rationality of judgements about acceptable risk but which notoriously enter into such judgements. One such very general factor is the *framing effect* explored by Kahneman and Tversky that tends to determine reactions to risks by the way risk statements are formulated. The risks of an operation tend to be judged as more acceptable if they are expressed in positive terms, i. e. in terms of the probability of survival, than if the same risks are given a negative wording in terms of the probability of death. The striking thing about this effect is that it works even with people who think they are intelligent and critical enough to be immune to verbal deception. Again, this effect does not depend on the probabilistic nature of the events in question. The framing effect seems to be a general phenomenon of communication and not specific to communication about chances and risks (cf. Tversky and Kahneman 1981, p. 457).

Another point on which both parties are agreed is that there are emotional attitudes to risks that are "irrational" in so far as they are incompatible with the considered judgements of the judging person himself. Thus, a person may develop a generalized fear of dogs after having been severely bitten by a dog that co-exists with the belief that most dogs are innocuous, and even with the knowledge that this particular fear is neurotic and unfounded. This shows that emotions and emotional attitudes can be "irrational" in more than one way. They can be "irrational" or inadequate, as this last example shows, by becoming autonomous, dissolving the ties that normally bind them to the faculty of judgement. And they can be "irrational" or inadequate by being in harmony with the faculty of judgement, which in turn is mislead, either by the impact of the emotional factors involved or by independent factors. It is one of the weaknesses of most traditional theories of emotions that they do not distinguish clearly between these two kinds of irrationality. This holds even of Spinoza's theory of emotions which in other respects goes a long way to do justice to the cognitive components of emotion and to distinguish between adequate ("active") and inadequate ("passive") emotions, but which nevertheless tends to regard all inadequate or "irrational" emotions as indistinguishably pathological ("delirii species", Ethics, IV, 44 Scholium). But, of course, there is an important difference between the kind of irrationality involved in phobias and that involved in fears based on deficient judgement. If A has an "irrational" fear of a certain dog because he has a generalized phobia in respect of dogs, this can safely be categorized as pathological. This is definitely not the case if B has an equally irrational fear of a certain dog because he has been misinformed about the dog's dangerousness or because he mistakes the dog for another dog that is in fact dangerous.

Correspondingly, there are at least two ways in which emotions and emotional attitudes are open to criticism: They can be criticised because they are neurotic, and they can be criticised because they are based on false judgments. This latter possibility is open because emotions and emotional attitudes, in contrast to moods, feelings and feeling dispositions, have judgemental components. In the case of risks, these components include, among others, judgements about the kind of consequences to be expected from a certain action or event, the values of these consequences, their frequency, and the degree of certainty associated with the estimates

of each of these dimensions. Some of these judgemental components in turn contain emotional or non-cognitive elements, such as the valuation of possible outcomes. If these judgemental components go wrong at a certain point, the emotion or emotional attitude based on it will, as a rule, go wrong as well.

3 Which "Qualitative Risk Factors" can be Integrated into the Bayesian Scheme?

There remain a number of factors about which Bayesians and Non-Bayesians differ, and these differences must now be considered. We should be careful to make two kinds of distinction from the start. The first distinction is that between those factors in judgements about acceptable risk that are in principle amenable to an analysis and evaluation within the Bayesian scheme, but are only rarely given attention in practice, and those that resist integration and require a revision of that scheme. If it turns out that the qualitative risk factors which the Non-Bayesians appeal to lend themselves to a reconstruction within this scheme, then what deserves to be criticized is the practice rather than the theory of Bayesianism. The upshot is not that thinking about technological risk in terms of values and frequencies is misguided in principle, but that the potential of this thinking is only insufficiently made use of in practice. The case is different if it turns out that some of the qualitative factors cannot in principle be reconciled with this approach.

The second distinction is that between factors in judgements about acceptable risks for which it is at least prima facie plausible that they should be included in a theory of acceptable risk and others for which this is less clear. We have already referred to the empirical fact that the riskiness of technologies is to some degree influenced by the estimated extent to which one is threatened by a certain risk in one's own person. It is clear that a person-relative criterion of this kind cannot legitimately figure in impersonal judgements about risks. The fact that certain consequences may be dangerous for *me* cannot be relevant to the judgement about whether a certain risky technology is acceptable from the perspective of society or of all who are positively or negatively affected by it.

Bayesianism can in principle include many of the factors that tend to be quoted by Non-Bayesians as examples of the context-sensitivity and adequacy of risk judgement of the general public. Bayesianism is an extremely flexible instrument. If many of the qualitative factors are not normally included in formal analyses this is mainly because they do not lend themselves to easy calculation. In principle, however, the costs and benefits taken account of in the Bayesian analysis are not restricted to those easy to quantify, such as money or the number of deaths or injured. Instead, they can include aesthetic, social and political benefits and harms such as the loss of amenities that go with many forms of "big" technology, the obsolescence of skills in the wake of new technologies and the loss of democratic control often involved in centralised production. On a more fundamental level, Bayesianism is not bound to one particular system of valuation of outcomes. There is no necessary connection between Bayesianism and utilitarianism, nor, for that matter,

between Bayesianism and consequentialism. Against Harsanyi who thought that the Bayesian rationality postulates "entail utilitarian ethics as a matter of mathematical necessity under relatively weak conditions" (Harsanyi 1978, p. 223) it must be said that this follows only under the presupposition that utility can be adequately conceptualized along Neumann-Morgenstern lines and that all possible values collapse into preferences. Neither is Bayesianism entailed by utilitarianism. A utilitarian can have good utilitarian reasons for a more risk-averse approach than the risk-neutral approach implied by the maximization of expected values. Though consequences matter in Bayesianism (as in all theories of acceptable risk), this does not prejudge how the (possible) consequences are valued. First, these consequences need not be valued on exclusively consequentialist terms. Even if the concept of risk is inherently a consequentialist concept in so far as it involves uncertain consequences, this does not imply that these consequences have to be assessed in accordance with an ethic that measures the severity of negativities by the value of consequences. The consequences might alternatively be measured by deontological features such as the extent to which they constitute violations of rights. As Ralph Keeney emphasized (cf. Keeney 1984, p. 120) the technical apparatus of Bayesianism is indifferent to the kind of values assigned to the possible outcomes and is able to provide even for the extreme case that the outcomes are only valued according to their moral instead of their non-moral value. In brief: Though one may easily agree to Harsanyi's thesis that result-orientation is a central principle of rationality in responsible decision-making (Harsanyi 1978, p. 225) and that the question whether a risky activity is acceptable from a prudential or moral point of view cannot be determined by its inherent or purely symbolic features, this does not settle the issue which values, and which kinds of value, are assigned to the results. Even if it is beyond question that in decisions under risk consequences matter, this leaves open whether the standards by which the consequences are evaluated are of a consequentialist or deontological kind, and whether, if they are of a consequentialist kind, the value of the consequences is determined only by non-moral values such as life, health and quality of life or by moral values like morally good actions, morally good intentions or the exercise of virtue.

The adaptability of the Bayesian model to different systems of valuation goes even further. Apart from taking account of goods or bads that befall individuals, it is also able to incorporate structural values such as equality, equity or distributional justice provided these are operationalised in a way that makes them commensurate with individual values. Thus, it is perfectly able to incorporate the intuition that a distribution of risks is highly unfair if A gets the whole profit from a risky activity and B bears all the burdens. Empirical surveys show that judgements on acceptable risk react to his kind of unfairness (cf. Renn and Zwick 1997, p. 92). An account that aggregates only individual goods or bads cannot represent this kind of unfairness. But, as the example of Rainer Trapp's "non-classical" brand of utilitarianism (Trapp 1988) and the "person-trade-off" approach in health economics (Nord 1999) show, these structural features can be integrated into a consequentialist scheme by treating them as a dimension of chances and risks that supplement the "classical" individualist dimensions and can be handled along the same formal lines. From an

ethical point of view, it does not at all seem incompatible with rationality to take distributional features into account. On the contrary, it seems imperative to give features like equality and fairness some role to play in the evaluation of consequences. It would be far-fetched to think that to care for equality and fairness in the distribution of risks and chances is pure sentimentality or, in this sense, purely "emotional".

There is a further dimension of technological risk that must be taken account of in any adequate calculation and comparison of risk: the benefits of security and the harm of insecurity. The bad thing about risks is not only that they involve a bad of some kind in case they materialize. The bad thing is also the psychological threat their existence implies for those subject to the risk. The psychological benefits and harms of a technology are not exhausted by the psychological goods and bads involved in the materialization of its chances and risks. They include, in addition, the benefits and harms connected with their anticipation, especially if this undergoes a process of "social amplification" by which the psychological effects spread through society (cf. Kasperson 1988; Renn 1991). The prospect of a future possible good is, as a rule, itself a good, the prospect of a future possible bad itself a bad. Therefore, the fear and insecurity generated by the existence of a risk should be taken as serious as the feeling of insecurity generated by its materialization. In my view, they should be included in the risk profile of a technology even in cases in which these feelings seem, from an objective point of view, exaggerated or "hysterical", or are based on severely distorted risk perceptions. As far as these feelings are immune to enlightenment they must be added to the items in the negative side of the balance, irrespective of whether they are rationally justified or not. It is significant, moreover, that according to the psychologists of risk the prospect of a future bad is worse to a higher degree than the prospect of a future good is good. Future harms, whether certain or probable, arouse more fear than future benefits, whether certain or probable, arouse joyous expectation. Obviously, we react to future harm or risk as born optimists for whom the good is the normal thing. This is an additional reason to give some weight to the feelings of insecurity generated by risks.

There is, then, a certain range of factors in laypeople's risk perceptions and assessments that can be reconciled with Bayesianism, at least with a suitably refined version of it. It remains to be shown that this is true for all factors that can be rationally justified. Of course, as with other attempts to reconcile doctrine and common sense, we should not mislead ourselves into thinking that there is a pre-established harmony between the common sense and the scientific approach. We should take serious Amos Tversky's warning that "in the absence of any constraints, the consequences can always be interpreted so as to satisfy the axioms" (Tversky 1975, p. 171). The best thing to keep clear of this temptation is to adopt as far as possible the point of view from which the general public judges on acceptable risk and only then make the second step to ask how this fits into the Bayesian picture.

Among the qualitative characteristics that play a role in common sense risk perception some concern primarily the *value* of the consequences of a risky activity or event, others the level of *insecurity* generated, and others both. This gives us

a principle by which we can classify the main candidates among the qualitative characteristics for integration into the Bayesian scheme.

Among the first group, the characteristic that it seems easiest to reconcile with a Bayesian approach, is *irreversibility*. Irreversible harms are commonly given more weight than reversible harms, and risks with irreversible outcomes are commonly feared and avoided to a higher extent than risks with reversible outcomes. It is evident that this factor can and must be honoured in a Bayesian analysis. In general, the fact that a harm is irreversible means that the consequences of the harm are more severe than those of a corresponding reversible harm, partly because of their scope and partly because of their opportunity costs. Irreversible harm can be expected to stay for a longer period of time than reversible harm. Moreover, it narrows the options available and in this way compromises freedom. In order to compensate for the harm in terms of subjective well-being, one usually has to invest more labour and time than in the case of reversible harm, provided that compensation is possible in the first place. Apart from material costs, psychological costs are usually higher. Whereas a house destroyed by a fire can be reconstructed, human victims cannot be revived, objects belonging to the cultural heritage cannot be restored in the original. Coping with irreversible losses of what one valued requires patience and humility, and is usually accompanied by a longer period of suffering.

Another member of the second group is (perceived) *control*. The psychology of risk has shown that risky activities that are subject to control by whoever engages in it are commonly judged to be more acceptable than comparable activities over which the subject has no control. To many people, the risks of using a ski-lift seem to be more severe than the risks of running-down skiing, the risks of travelling by airplane more severe than the risks of driving. The important variable seems to be the extent to which the subject believes to be autonomous in the direction the risky activity takes while it is running. Whereas voluntariness concerns the freedom to engage in a risky activity by one's own choice, control concerns the freedom to change the course of events at will while it lasts. Can this factor be integrated into the Bayesian scheme? Surely, at least to a certain extent. As far as control is a relevant psychological variable, it must be included in an adequate calculation of outcome values. Even if people overestimate the extent to which they are able to control the process which they think they can control, this feeling substantially contributes to the subjective feeling of safety, at least within the "Faustian" culture of the West (that successively seems to govern the world) that values active control of social and natural processes more than passive acceptance.

The most important members of the third group are *voluntariness* and *potential for catastrophe*. Both factors tend to modify both the value of possible outcomes and the extent to which risks by their very existence generate feelings of security or insecurity. *Voluntariness* has proved to be highly relevant for perceived acceptability of risks. One of the pioneers in the scientific study of risk-taking, Chauncey Starr, went so far to maintain that voluntary risks can be one thousand times as great as risks of an involuntary nature to be judged acceptable (Starr 1969, p. 1237). This conclusion, however, was derived exclusively from revealed preferences, measured in monetary terms. Though Starr's interpretation seems exaggerated, it is a

fact that people attach very great importance to having a choice instead of having risks imposed on them by others. The relevance of this factor is evident. It is evident that what matters in choices between risks is not only the estimated value of outcomes but the autonomy in taking risks in the first place. This is reflected in the striking differences in the emotional attitudes towards voluntary and involuntary risks. The harm we suffer from a self-imposed risk "feels" differently from a harm from a risk imposed by others without our consent. The risk of death by murder has an emotional quality strikingly different from that of the risk of death by suicide. Both violate our integrity, but only the former violates our autonomy. Whereas risks imposed by others or by natural factors limit our autonomy, risks imposed on ourselves by ourselves increase our autonomy. On the one hand, we have a strong interest in not being subjected to risks by others without our consent. On the other hand, we have an equally strong interest in being free to impose risks on ourselves if we so want, for example by risky kinds of pastimes and sports. Voluntariness is itself a utility, involuntariness a disutility. The attractiveness of voluntary activities lies partly in their very being voluntary, the unattractiveness of involuntary activities in their being involuntary. The same activities often assume completely different values according to whether they are freely chosen or constrained.

Voluntariness belongs to the third category because it affects both the valuation of the possible outcomes and the dimension of security. Voluntariness is not only a utility in its own right, it also has an impact on the degree to which we feel secure. Insecurity is primarily dependent on the extent to which we are subject to risks imposed by nature or by others without our free consent. It is true, the more people are prone to uncontrollable impulses the more reason they have to fear the consequences, for themselves and for others, of their own passion, rashness and foolishness. But to the same extent that they have those reasons it is doubtful whether these risks can be classified as fully voluntary.

Something similar can be said about the *potential for catastrophe*. This characteristic, too, tends to aggravate both the harm in case the risk materializes and the feeling of insecurity it generates by its existence. Risks involving harms that occur rarely but in catastrophic dimensions are much less accepted than risks with the same number of victims where these are distributed over time and each single harm is too trivial to arouse public attention. Thus, air traffic accidents are given much more publicity than car accidents though the total number of victims is considerably lower. One single accident with fifteen thousand deaths is much more spectacular than the same number of deaths by domestic accidents. But even if we discount the factor of public attention there is a strong intuitive tendency to judge risks with the potential for disaster less acceptable than risks with a more distributed pattern of incidence.

Is this intuition open to reconstruction within the Bayesian model? Certainly it is, at least to a certain extent. What distinguishes the harm caused by a catastrophic event with thousands of deaths at a time from a sequence of thousand individual deaths distributed over time is exactly that the harm occurs simultaneously and has a more thoroughgoing impact, both on the material and the psychological resources of a society. On the material side, non-linearities in the disutility of concentrated

harm have to be taken into account. One accident with the great number of victims can transcend the capacities of a society in terms of medical, technical, financial and human support. A society which can come to terms with one hundred similar incidences of disease per week is not necessarily in a position to deal adequately with five thousand incidences a day. In the long term, a catastrophic event often has lasting effects on the economy, e. g. by companies going out of business, unemployment, and costs due to rising safety standards. On the psychological side, the impact can be worse: the collapse of the economic basis of a whole region, social upheavals, loss of trust. Take as an example the impact of the Tschernobyl accident on the prestige of nuclear energy even in nations in which an imprudence similar to the one that caused the accident is hard to imagine. (To do justice to these additional effects, R. Wilson once proposed to calculate the social costs of accidents with n deaths by n^2, cf. Starr et al. 1976, p. 657).

At the same time, the additional factor of decreased perceived security dictates that catastrophic possibilities are assigned a special weight over and above the weight they receive in expected value analysis. The very existence of the possibility that a technology can lead to catastrophic harm is a significant psychological item in the overall risk of a technology. Given the "desire for certainty" (Slovic 1978, p. 101), it makes a world of difference whether the probability of disaster is 0.00 or 0.01. This difference is much more significant than a difference between, say, a probability of 0.50 and 0.51. There are, then, excellent reasons to handle catastrophic possibilities differently from medium-sizes risks and not to level them down in the way they inevitably will be in expected value analysis as it is commonly applied. In this respect, then, the intuitive and "emotional" reactions to disastrous risks can serve as a clue. They point to the fact that the frame in which analyses are commonly carried out has to be extended so as to take account of these additional factors.

It goes without saying that paying tribute to these factors considerably complicates the Bayesian picture. The simplicity and the elegance in the evaluation of risks that recommends Bayesianism especially to engineers and technological planners would have to be sacrificed. A good deal of the sensitivity to context present in intuitive judgements of acceptable risk would have to be integrated into the Bayesian frame. But there are good reasons to justify these complications. On the one hand, any simpler version of Bayesianism would be less adequate. On the other hand, any model that renounces calculation and deals with risks on a purely intuitive basis would lack the transparency that goes with an explicit analysis and balancing of factors.

4 Leaving the Bayesian Picture Behind

One dimension that looms large in common sense risk perception and assessment has not yet been mentioned, the dimension of the perceived *uncertainty* in the probability estimates. In general, the more uncertain the probability estimate is

perceived to be, the more risk-averse is our attitude towards the risk in question. This is an important fact because with controversial technologies estimates of benefits and risks are nearly always based on limited experience and essentially depend on subjective probability estimates by experts that lack the certainty about relative frequencies available for lotteries and games.

With technologies about which there is too little experience to give reliable estimates of the frequencies with which possible harmful consequences might result, there is in fact not only one, but two kinds of uncertainty to consider, each on a different level of knowledge and ignorance: uncertainty about the probability with which certain kinds of possible adverse events are to be expected, and uncertainty about whether the list of possible consequences considered in the calculation of risks exhausts the possibilities. The first kind of uncertainty concerns, among others, risks that can be identified but cannot be calculated by standard methods like fault-tree analysis or simulation, such as common mode failures, external effects and human factor risks (cf. Kates 1981, p. 93 f.). Who, for example, would have thought of the possibility that in 1975, a technician checking for an air leak in Brown's Ferry Nuclear Plant on the Tennessee river would do this, in violation of standard operating procedures, with a lighted candle, thus causing a disastrous fire? The second kind of uncertainty is likewise hard to avoid. There is nearly always a small probability that certain risks have been overlooked or could not be known in advance, either for contingent or for principle reasons. Well-known examples from the history of technology teach us that some causes of disaster are, and can, only be identified after the event.

It can fairly be said that situations of choice with elements of uncertainty are more common than situations in which all risks are completely known. Complete knowledge of probabilities of all possible eventualities is as rare as complete uncertainty. This is especially true of situations in which technologies are at stake for which limited experience does not allow a final judgement about how safe they are under critical conditions. For these situations, the "emotional" reserve about new technologies that have not yet stood the test of proving their safety under real-world conditions, has some measure of truth in it. The generally lower acceptance of technologies with incompletely known risks (provided the chances benefits these technologies offer are not seen as substantial enough to outbalance the risks) has a *fundamentum in re* and cannot be attributed to excessive conservatism. There seems to be a "rational core" in the conservative instincts that permeate the emotional attitudes to new technologies in the general public, except in areas like medicine or communication where attention is primarily focussed on the benefits.

What does this imply for the assessment of technological risk? I think that it calls for a revision of the Bayesian model, not so much because of the inevitable uncertainty of probability estimates on the first but because of the inevitable uncertainties on the second level. There are a number of conditions that constrain Bayesianism as an appropriate strategy in risk assessment: that the risky activity or event is iterated so many times that adverse outcomes are compensated by favourable outcomes; that no outcome is so disastrous that it overthrows the system; and that all relevant benefits and risks are identified. Uncertainty on the first level can be made consonant

with these conditions by reconstructing it as a range of probability along the lines formulated, e. g., by Rescher (1983, p. 94 ff.) or by identifying a "tolerable window" (Posner 2004, p. 176 ff.). In this way, choices under uncertainty and comparisons of risk with uncertain probabilities can be treated by the same expected-value assessment appropriate to reliable probabilites. A more serious limitation is the second condition that the risks of which we are uncertain must at least admit of identification, a condition that fails to be fulfilled in many cases in which a technology is controversially discussed. This fact, indeed, is a strong argument for adopting a more risk-averse strategy than Bayesianism. This is not to say that the adequate strategy should be as conservative as the maximin rule that ranges options according to worst possible outcomes irrespective of probabilities (cf. Rescher 1983, p. 161, Leist and Schaber 1995, p. 56). Such a principle would be excessively prohibitive of technological progress, which requires a minimum of preparedness to gamble even with grave risks. But the principle should at least restrict the possibility, present in the Bayesian model, to balance severe harm for the victims of technological progress by the benefits provided to the rest of mankind.

There is an additional reason for questioning the adequacy of Bayesianism in determining acceptable technological risk. Decision-making on risky technologies can be conceived in two ways, each with different consequences for the criteria of ethical legitimacy. On the one hand, it can be conceptualised as the self-imposition of risks by a collective such as a nation or a transnational unit. On the other hand, it can be conceptualised as an act by which an authority imposes risks on others, e. g. on those parts of the population that are positively or negatively affected by the respective technology. In the first case, the decision to carry out the activity follows the decision theoretical model of subjective rationality. Since the deciding agent is identical with the agent who bears the benefits and the risks of the decision, the question is how to optimize the relation between benefits and risks given the preferences of the collective agent. In the second case, the appropriate model is the model of justified imposition of risks on others. The question is no longer a question of subjective rationality but a question of ethics. The question is whether it is morally legitimate for the authority to impose risks on others who may have preferences widely different from those deciding on or carrying out the risky activity.

If one adopts the latter, individualistic, point of view, as I think we should, it is plausible to take account of all relevant preferences of those affected by a technology, including their risk preferences. Even if the agent himself is a Bayesian, convinced that the best thing is to choose the option by which the expected value for all affected by the option is maximized in the long run, it is doubtful whether he is justified in generalizing this preference and to impose risk profiles on others which they, from their own risk preferences, want to steer clear of. As soon as others are affected by the agent's choices the question arises whether it is legitimate to orient the imposition of risks exclusively on one's own risk preference. Even if the agent, as far as he is concerned, is perfectly willing to have risks imposed on him in accordance with his own risk preferences, it is doubtful whether this legitimatizes imposing corresponding risks on others. After all, it is not usually the case that we

are allowed to impose on others what we allow others to impose on us. (Think of G. B. Shaw's travesty of the Golden Rule: "Do not do unto others as you would be done by them. Their tastes might be different.") A physician who, as far he is concerned, would be willing to undergo a certain risky operation, cannot assume a priori that his patients share his risk preference and prefer the risky operation to a more conservative treatment involving less risks and less chances. If his obligation, in deciding about how to proceed, is to respect the preferences of his patients, it is part of the same obligation to respect their risk preferences.

It is a well-known fact that as far as the risk profile of technological options include risks of a certain severity, the risk attitudes prevailing in the general public are risk-averse rather than risk-neutral. This fact constitutes a further reason to modify the Bayesian approach in the direction of an approach that attaches more negative weight to adverse events and restricts the balancing of risks and benefits, without, however, unduly obstructing technological progress.

5 Conclusion

Risk is inherently a value concept and cannot be analyzed in purely descriptive terms. This is one reason why emotions in a broad sense including emotional attitudes are a central component in the assessment of the risks of technological options. Likewise, emotional factors play a part in the explicit or implicit valuation of the distribution of benefits and risks, in risk preference and in the characteristics shown by psychologists of risk to contribute to the acceptance or non-acceptance of risky technologies. Not all of these emotional factors are compatible with rationality. For reasons of economy and limitation of resources, emotions, just as perceptions, make use of simplifying heuristics that are often useful and sometimes misleading.

In this article, I have mainly dealt with emotional factors for which it is plausible to assume that they are compatible with rationality at least to a certain extent: voluntariness, control, potential for catastrophe, and uncertainty. These factors, however, are only a selection from the "qualitative characteristics" that have been found to determine the popular perception and acceptance of risks. Other factors in this list are, I think, not amenable to a reconstruction in the Bayesian or any other model of rationality. They are "emotional" not only in the sense that they are rooted in spontaneous and non-cognitive tendencies but also in the polemical sense designed to deny them intellectual respectability. Rather than useful heuristics in situations where the tools of formal analysis fail to be helpful, they mislead our thinking and misdirect our actions. Among these are: 1. symbolic values, 2. salience; 3. familiarity, and 4. naturalness. *Symbolic content* seems to play a significant role in the valuation of risk. For example, energy production from nuclear fission is inevitably associated with the nuclear bomb and solar energy with the life-giving role of the sun, so that it appears "natural" that the latter is less risky than the former. *Salience* is a factor in the assessment of technological risks as it is a factor in the individual's perception and evaluation of possible diseases. The frequencies of dramatic and sensational events are overrated, the frequencies of trivial events are underestimated (Slovic

1978, p. 100), with the consequence that people and politicians are more prepared to invest in security against murder and terrorism than in everyday causes of death like coronary infarction and infections from hospitalisation. *Familiarity*, again, makes that we hardly worry about risks that have become habitual features of our life-world even if they are far more substantial than less familiar risks. Interestingly, the emotional attitude underlying familiarity corresponds to the *absence* of emotion in the episodic sense. In a sense, we react to risks that have become familiar with less emotion than would be appropriate. In this way, deaths and injuries from car traffic have become familiar, whereas deaths and injuries by radiation have not. Without the factor of familiarity it seems difficult to explain that energy production by burning coal, with 15,000 deaths in German coal mines since 1948, is widely accepted whereas energy from nuclear power, with 0 deaths in German reactors, is not. It may be thought that more familiar risks are easier to tolerate because society has had time to adapt to these risks and to establish corresponding means to come to terms with them. This consideration, however, is hardly relevant because mechanisms and institutions to control and to correct these risks are already part of the calculation. The existence of a fire brigade is already part of the overall risks of fire, the institution of hospitals already part of the overall risks of car traffic. In consequence, the fact that more familiar risks are more easily accepted than unfamiliar ones has to be explained by habituation effects and cannot be rationalised along the lines of the dimension of certainty. At last, *naturalness* is an important factor in the acceptance of risks, for which, again, it is difficult to see how it can be interpreted as rationally defensible (cf. Hansson 2003). Natural causes of harm are given what might be called a "nature bonus". Natural harms are less feared than anthropogenic harms, possibly because there is nobody in particular to blame for inflicting it. One cancer patient dying from the radioactivity emitted by a nuclear power plant will attract more attention than ten or hundred patients dying from natural radiation. Again, in the preference for natural above technical or other anthropogenic risks, emotions seem to play a central part, possibly due to evolutionary constraints such as the impossibility, over long periods in the history of mankind, to control natural risks (cf. Birnbacher 2006, p. 21 ff.).

All in all, then, emotions are a mixed blessing – in the assessment and acceptance of risks no less than in other domains of life.

References

Birnbacher, D. 2003. Can discounting be justified? *International Journal of Sustainable Development* 6: 42–51.

Birnbacher, D. 2006. *Natürlichkeit*. Berlin/New York: de Gruyter.

Gigerenzer, G., and R. Selten. 2001. Rethinking rationality. In *Bounded Rationality. The Adaptive Toolbox*. G. Gigerenzer, and R. Selten, eds., 1–12, Cambridge, MA: The MIT Press.

Hansson, S. O. 2003. Are natural risks less dangerous than technological risks? *Philosophia Naturalis* 40: 43–54.

Harsanyi, J. C. 1978. Bayesian decision theory and utilitarian ethics. *Economics and Ethics* 68: 223–228.

Johnson, E. J., and A. Tversky. 1983. Affect, generalization, and the perception of risk. *Journal of Personality and Social Psychology* 45: 20–31.

Jungermann, H., and P. Slovic. 1993. Charakteristika individueller Risikowahrnehmung. In *Riskante Technologien, Reflexion und Regulation*. W. Krohn, and G. Krücken, eds., 79–100, Frankfurt/M: Suhrkamp.

Kasperson, R. E., et al. 1988. The social amplification of risk. A conceptual framework. *Risk Analysis* 8: 177–187.

Kates, R. W. 1981. *Risk Assessment of Environmental Hazards*. Chichester: John Wiley & Sons.

Keeney, R. L. 1984. Ethics, decision analysis, and public policy. *Risk Analysis* 4: 117–129.

Leist, A., and P. Schaber. 1995. Ethische Überlegungen zu Schaden, Risiko und Unsicherheit. In *Risikobewertung im Energiebereich*. M. Berg et al., ed., 47–70, Zürich: Verlag der Fachvereine.

Nord, E. 1999. *Cost-Value Analysis in Health Care, Making Sense of QALYs*. New York: Oxford University Press.

Posner, R. A. 2004. *Catastrophe, Risk and Response*. Cambridge: Cambridge University Press.

Renn, O. 1991. Risk communication and the social amplification of risk. In *Communicating Risk to the Public*. R. E. Kasperson, and P. J. Stallen, eds. , 287–324, The Netherlands: Dordrecht.

Renn, O. 2008. Concepts of risk, an interdisciplinary review. *Gaia* 17(50–66): 196–204.

Renn, O., and M. Zwick. 1997. *Risiko- und Technikakzeptanz*. Berlin: Springer.

Rescher, N. 1983. *Risk. A Philosophical Introduction to the Theory of Risk Evaluation and Management*. Lanham, MD: University Press of America.

Slovic, P. 1978. Judgement, choice and societal risk taking. In *Judgement and Decision in Public Policy Formation*. K. A. Hammond, ed., 98–111, Boulder, CO: Westview Press.

Slovic, P., B. Fischhoff, and S. Lichtenstein. 1979. Rating the risks. *Environment* 21: 14–39.

Starr, C. 1969. Social benefit versus technological risk. *Science* 165: 1232–1238.

Starr, C., R. Rudman, and C. Whipple. 1976. Philosophical basis for risk analysis. *Annual Review of Energy* 21: 629–662.

Trapp, R. W. 1988. *"Nicht-klassischer" Utilitarismus. Eine Theorie der Gerechtigkeit*. Frankfurt/M: Klostermann.

Tversky, A. 1975. A critique of expected utility theory, descriptive and normative considerations. *Erkenntnis* 9: 163–173.

Tversky, A., and D. Kahneman. 1974. Judgement under uncertainty, heuristics and biases. *Science* 185: 1124–1131.

Tversky, A., and D. Kahneman. 1981. The framing of decisions and the psychology of choice. *Science* 211: 453–458.