Chapter 2

The Social Background to Technological Risk

2.1 What Is Risk?

Risk As An Everyday Notion In everyday speech, one *risks* something, or one *takes a risk*, if a goal outcome is not certain to be attained by a course of action, or if a course of action has a certain likelihood of engendering deleterious consequences in the course of attempting to reach the goal.

2.2 Risk And Teleological Systems

Risk As Associated With Purposive Behavior Notice first that we are speaking of *purposive* or *intentional behavior.* One has a goal of sorts, although sometimes this goal may be just to execute the particular course of action. Wandering around aimlessly can be considered as much a purposive behavior as running for the bus.

Teleological Systems I call a system a *teleological system* if it has a purpose of some sort, if there is a goal that the system is intended to fulfil. Most artifacts are teleological systems; predator-prey systems are not (when I speak of intentions, I mean human intentions; the intentions of putative "higher beings" as expressed in some religions do not count); indeed even the systems considered by social systems theorists are in large part not teleological - they might have arisen as the product of many individual and collective intentions, but they do not themselves have an explicit goal. Special-interest groups, lobbying groups and even companies are said to have a goal. As such, their efficient and goal-oriented functioning can be questioned. However, I shall mainly concentrate on artifactual teleological systems.

The Risk Background of Teleological Systems When we build teleological systems, the background includes amongst other things

- the operation of the system to achieve the goal,
- the reasons for wanting to achieve the goal,
- the consequences of achieving the goal,
- effects of the course of action implemented to achieve the goal,
- the consequences of failure to achieve the goal,
- effects of the failed course of action utilised
- the range of accidents possible through achieving the goal
- the range of accidents possible through trying and failing to achieve the goal
- the range of accidents possible through trying to achieve the goal through the particular course of action embodied in the system

Typical System Risk Analysis A typical system analysis considers only

- the operation of the system to achieve the goal,
- the range of accidents possible through trying to achieve the goal through the particular course of action embodied in the system,

which is a somewhat restricted subset of the total background questions on safety.

2.2.1 Risk Analysis As Profession

Risk Assessment and Management as Profession It has been mooted [FLSDK81, KH99] that risk assessment is a professional skill by itself. There are a number of reasons for this.

- It involves techniques that one does not learn in the technical professions involved in designing the teleological systems.
- There is a variety of well-developed techniques which may be brought to bear on any of the issues involved in risk assessment, that are significantly more effective than naive approaches to the issues
- The techniques involved in risk assessment in a variety of fields have much in common with each other; more than they have in common with the techniques of teleological system building within the field itself

While [KH99] is primarily concerned with calculation techniques and [FLSDK81] with management, [MH90] deals with all the techniques they have found useful as specialists in handling uncertainty in assessment situations. Further to [FLSDK81]'s view that risk assessment is a decision problem, the practical techniques of so-called Decision Science, that is, rational decision making in uncertain situations, are described in [KKS93]. The fundamentals of rational choice, along with expositions of significant items of theoretical interest such as Arrow's Impossibility Theorem, may be found in [Res87]. Those interested in the foundations of inductive reasoning and the fundamentals of Bayesian decision theory may be interested in [Sky99].

General Difficulties with Risk Assessment Speaking against the acceptance of risk assessment as a profession are the following phenomena.

- The problems often defy precise formulation
- Although there are well-developed techniques, these techniques are often best employed not as decision methods, but in concert with other "competing" techniques as decision aids only.
- A large number of techniques involve subdisciplines of already-established fields such as social-psychological interviewing techniques, elicitation of expert judgements, and statistical evaluation.
- Some effective techniques, although relatively few, such as Fault Tree Analysis and Hazop, are specifically bound to technical disciplines already.
- Because of the nature of the problems, in which many of the data are unknown or very uncertain, the very best and most careful analyses can still lead only to rough, approximate answers; a situation which is theoretically and often socially unsatisfactory.

2.3 Risk Assessment

2.3.1 Two Principles: Know And Consult

Two principles which one finds uniformly in writings on risk assessment are

Know Everything One should inform oneself as thoroughly as possible what the facts of the matter are;

Ask Everyone All "stakeholders" should have an appropriate degree of involvement in deciding whether to implement a system with attendant risks.

2.3.2 Fact And Value

One Principle, Two Views There is one principle which one finds asserted by one group of professionals and denied by another.

Fact and Value Should Be Separated Technical experts in safety analysis should present "decision makers" with the facts concerning the level of risk inherent in a system design; the decision makers should solicit the decision through social processes. For an example of this point of view, see Section 2.4.4.

Fact and Value Cannot Be Separated Implicit in any technical assessment of a system are a series of assumptions about what matters and what does not matter. These assumptions are value judgements and should enter explicitly into an assessment of the values involved in a decision about risk.

A Value-"Fact" It is widely accepted by almost everyone (except certain U.S. Congresspeople) that there is no such thing as *Zero Risk*. But what does this mean? Another way of putting it is that no course of action is risk-free.

2.3.3 "Acceptable Risk": A Confused Concept?

A Confused Concept The idea of "*Acceptable Risk*" has been proposed as an alternative to the concept of zero risk. Various definitions of "acceptable risk" have been proposed.

- A chance of less than 1 in 10^6 of an untimely death during a lifetime [Lew90, pp95,105], see also [Ato76] as quoted in [FLSDK81, p85]
- A chance of less than 1 that a catastrophic accident will happen to any device (aircraft) during the lifetime of the fleet [LT82, p37]
- No significant increase in risk over "background" levels without the technology [Wei79] as quoted in [FLSDK81, p87];
- Nothing scandalous about my behavior making it into the news during my time in office [various heads and former heads of state of Western countries, 1998-9]

A Less Confused Concept Rather than talk about levels of risk being "acceptable" or not, one may prefer to talk about risky "*options*" being acceptable or not:

> Strictly speaking, one does not accept risks. One accepts options that entail some level of risk among their consequences.
> [FLSDK81, p3]

By “options” is meant, for example,

- a course of action (including doing nothing);
- design of a teleological system;
- forms of use of a teleological system.

The following definition from [FLSDK81, p2] suggests exactly this:

> Acceptable-risk problems are decision problems; that is, they require a choice among alternative courses of action. What distinguishes an acceptable-risk problem from other decision problems is that at least one alternative option includes a threat to life or health among its consequences. We shall define *risk* as the existence of such threats.

Notice that the definition of “risk” is more narrow than the technical definition we consider in Part II. That definition suggests that we could define “loss” whatever way we wanted, and not necessarily as a threat to life or health.

2.3.4 Risk As Decision

The Decision Problem The decision problem identified by [FLSDK81] consists of (quote)

> ......the following five interdependent steps
>
> 1. Specifying the objectives by which to measure the desirability of consequences
> 2. Defining the possible options, which may include “do nothing”
> 3. Identifying the possible consequences of each option and their likelihood of occurring should that option be adopted, including, but not restricted to, risky consequences
> 4. Specifying the desirability of the various consequences
> 5. Analyzing the options and selecting the best one
>
> [FLSDK81, p2]

They point out that no known techniques allow each of these processes to be conducted optimally during an analysis. They evaluate the known techniques against this “wish list”.

	Certain Knowledge	Uncertain Knowledge
Complete Consent	Problem: Technical Solution: Calculation	Problem: Information Solution:Research
Contested Consent	Problem: (dis)Agreement Solution: Coercion or Discussion	Problem: Knowledge and Consent Solution: ??

Table 2.1: Douglas and Wildavsky's Problem Table

2.4 Alternative Conceptions of Risk

2.4.1 Risk as Interplay of Knowledge and Consent

[DW82] suggest that

> Risk should be seen as a joint product of *knowledge* about the future and *consent* about the most desired properties.
> [DW82, p5]

They pose the problem of assessing risk in the form of a table, shown in Table 2.1.

Some Components Are Missing It should be clear from a few moments thought that the explanation of [DW82] was not intended as a definition. The notion of *loss* is absent, and it is hard to see how this could be explained in terms of knowledge and consent. Supposing I were to wish to steal a sheep, know that I can do so, and have perfect knowledge of the likelihood that I would be caught. I would presumably be as unlikely to consent to the consequence that I would pay a fine of $ 10,000 as I would to consent to the consequence that my hand would be chopped off. However, the latter would be regarded by most as a more *severe* consequence and as a greater loss should it come about. (It may well be that notions of loss and of severity of consequence are interdefinable.) This preference cannot be explained by the notions of knowledge and consent, since by hypothesis these are the same in the two cases. However, it can obviously be explained by the notions of loss or severity, as the very phrasing of the example shows.

The Subjects Are Missing Knowledge doesn't exist in a vacuum. People have knowledge, and different people can have very different knowledge about outcomes. Consent exists even less in a vacuum – although one can reasonably speak of accumulated knowledge without supposed this accumulation is realised by any one person, it is hard to speak of consent without asking whose consent. One may assume that consent of a stakeholder is meant; the further question becomes how to identify "stakeholders".

A Tricky Example Consider the Jonestown massacre in Guyana, in which a nominally Christian cult which originated in San Francisco committed apparently willing mass suicide under the direction of the so-called "Reverend" Jim Jones. The participants are presumed to know what they were doing, to know that their actions of drinking the poisoned liquid would result in their deaths, and to have consented to this action. This would put them in the top left box in Table 2.1, which suggests that all they would need to do to find out their "risk" is to calculate. This cannot be so simple.

- Determining the consequences for the participants themselves might be straightforward, but
- determining the social consequences of their actions upon their relatives would not be so straightforward,
- and upon the social effectiveness in their roles of former religious colleagues of Jones in San Francisco,
- and upon the social tolerance of cults in formerly tolerant California, would be even less straightforward

This hangs on the assumption that the stakeholders in the action were also relatives, former colleagues, and members of Northern California society, and not just the participants themselves.

The Example Doesn't Fit the Proposed Paradigm Douglas and Wildavsky might reply by noting that I am saying that knowledge of the consequences was incomplete, and therefore this example fits rather in the upper right (we presume there is no dispute over consent). But their "solution" there involves "research". I doubt that research in advance of the mass suicide would have revealed the social consequences upon San Francisco and California society. So their prescriptions in Table 2.1 appear too facile.

An Extended Metaphor One might view Douglas and Wildavsky's contribution in [DW82] as an extended metaphor, intended to persuade readers of the essential dependence of risk on cultural norms, and emphasising the primacy of risk perception as a full-fledged component of risk. This view entails that fact and value cannot be separated.

2.4.2 The Royal Society's View

The Douglas-Wildavsky proposal stands in stark contrast to the "scientific" or "engineering" view, which tries to separate fact from value and which regards risk as the topic of calculations, such as in a Risk Cost-Benefit Analysis (RCBA), and perceptions of risk to be illusory in so far as they depart from the conclusions of such an RCBA . Compare the Royal Society's 1983 definition [Roy83], quoted in [Ada95, p8]:

> The Study Group views "risk" as the probability that a particular adverse event occurs diring a stated period of time, or results from a particular challenge. As a probability in the sense of statistical theory, risk obeys all the formal laws of combining probabilities.
>
> [.....]
>
> [*Detriment* is] a numerical measure of the expected harm or loss associated with an adverse event .... it is generally the integrated product of risk and harm and is often expressed in terms such as costs in £s, loss in expected years of life or loss of productivity, and is needed for numerical exercises such as cost-benefit analysis or risk-benefit analysis.

But Even Here Things Change Thanks to the arguments concerning separation of fact and value, the Society's view had changed by 1992 [Roy92]. The Study Group's terms of reference, quoted in [Ada95, p9]:

> [to] consider and help to bridge the gap between what is stated to be scientific and capable of being measured, and the way in which public opinion gauges risks and makes decisions.

The Study Group concluded, amongst other things, that

> the view that a separation can be maintained between "objective" risk and "subjective" or perceived risk has come under increasing attack, to the extent that it is no longer a mainstream position [Roy92], quoted in [Ada95, p9].

2.4.3 The National Research Council's View

Risk Characterisaction As Process The National Research Council Committee On Risk Characterization published seven principles for "implementing the [risk characterization] process" [Nat96, p2]:

1. Risk characterisation should be a *decision-driven activity*, directed toward informing choices and solving problems [Nat96, p2].

2. Coping with a risk situation requires a *broad understanding* of the relevant losses, harms, or consequences to the interested and affected parties [Nat96, p2].

3. Risk characterization is the outcome of an *analytic-deliberative process*. Its success depends critically on systematic analysis that is appropriate to the problem, responds to the needs of the interested and affected parties, and treats uncertainties of importance to the decision problem in a comprehensible way. Success also depends on deliberations that formulate the decision problem, guide analysis to improve decision participants' understanding, seek the meaning of analytic findings and uncertainties, and improve the ability of interested and affected parties to participate effectively in the risk decision process. The process must have an appropriately diverse participation or representation of the spectrum of interested and affected parties, of decision makers, and of specialists in risk analysis, at each step [Nat96, p3].

4. The analytic-deliberative process leading to a risk characterisation should include early and explicit attention to *problem formulation*; representation of the spectrum of interested and affected parties at this early stage is imperative [Nat96, p6].

5. The analytic-deliberative process should be *mutual and recursive*. Analysis and deliberation are complementary and must be integrated throughout the process leading to risk characterization: deliberation frames analysis, analysis informs deliberation, and the process benefits from feedback between the two [Nat96, p6].

6. Those responsible for a risk characterization should begin by developing a provisional *diagnosis of the decision situation* so that they can better match the analytic-deliberative process leading to the characterization to the needs of the decision, particularly in terms of level and intensity of effort and representation of parties [Nat96, p7].

7. Each organisation responsible for making risk decisions should work to *build organizational capability* to conform to the principles of sound risk characterization. At a minimum, it should pay attention to organisational changes and staff training efforts that may be required, to ways of improving practice by learning from experience, and to both costs and benefits in terms of the organization's mission and budget [Nat96, p8].

If the NRC Committee thought that risk could be characterised as a likelihood coupled with a severity, it is unlikely they would need to suggest such a complicated recursive, mutual, analytic-deliberative, feedback-oriented, stakeholder-intensive, proactive, organizationally-structurally-supported process. A few engineers with calculators coupled with a few stakeholders to tell how much it hurts should have sufficed. One can conclude that the Committee accepts the inevitable intertwining of fact and value in risk characterisation, and thereby the necessity of a political process to elicit those values and the consent of stakeholders. One only wishes they could have expressed it more succinctly.

2.4.4 A Software Safety Expert's View

Leveson writes that

> Making decisions such as how safe is safe enough involves addressing moral, ethical, philosophical, and political questions that cannot be answered fully by algebraic equations or probabilistic evaluations [Lev95, p17]. ...... We must also realise that decisions about safety will cause legitimate disagreements that cannot be resolved by simple utilitarian arguments [Lev95, p18].

Leveson quotes Alvin Weinberg, former head of Oak Ridge National Laboratory, as suggesting that it is the scientist's duty to

> ...inject some order into this often chaotic debate by distinguishing scientific from trans-scientific problems [Lev95, p18].

Rather than attempt to characterise reasoning as "moral", "ethical", "philosophical", "political" or "scientific", I would prefer to say that probabilistic reasoning is one form of reasoning that one can expect to use when reasoning about risk, and one should use it where appropriate. If there are reasons why one should clearly delimit where is is appropriate to use such forms of reasoning, as Weinberg suggests is the "scientist's duty", then presumably those reasons are good reasons for distinguishing "political" from "moral" or other forms. I doubt whether such a line can be drawn; utilitarian calculations belong as much to moral reasoning as they do to "scientific" calculations. Furthermore, reasoning is reasoning, and valid reasoning takes the same form in any subject matter. The question in any domain is more one of hidden assumptions than it is of any distinction in the notion of valid reasoning between "moral", "ethical", "philosophical" or "scientific" deliberations.

2.4.5 Risk Decisions As A Feedback System

Adams [Ada95] proposes to characterise risk decisions as a personal feedback system, in which a balance is sought between the rewards of risk-taking, one's personal propensity to take risks, the perceived danger, and knowledge of related accidents. Risk-taking decisions are the result of balancing these competing factors. He emphasises risk compensation (below) as an important factor in the system which is discounted in many risk analyses, and provides strong evidence for its existence.

A Quick Example of Risk Compensation I have just obtained a reclining bicycle, which I delight in riding. I wear a tie to work, and noticed that when I mounted the bicycle to ride to work, as I leaned over, the tie came perilously close to the oily chain, which is at thigh-level on the bicycle and only partly protected. I was very careful in mounting.

I then bought myself tie clips, to keep my ties attached to my shirt. I am much less careful about mounting the bicycle on my way to work. One day, if my tie clip fails to perform its function, because I have not attached it securely, I will suffer an oily tie.

It has been speculated that risk compensation may be a strong factor in the behavior of cyclists with helmets [Ada95, pp144-151], [Hil93].

I think we may conclude that risk compensation behavior is apparent. The question is, how significant a factor is it in assessing behavior while making decisions under risk?

2.4.6 Perception is an Irreducible Component of Risk

An Example Consider an example from [Ada95, p9]. Slipping and falling on ice is a game for young children, but potentially fatal for old people. The probability of such an event is influenced directly by the perception of its probability: old people see the risk of slipping on an icy road to be "high"; they take avoiding action, thereby reducing the probability for their group. The young people take minimal action, or even encourage it. Furthermore, older people share experiences and perception of the risk; so do young children in their peer group. Behavior is different; perception is different; consequences are different; probably even the mechanics are different. The role that such an event plays in the lives of these two groups is thoroughly different. This is sufficient grounds to speak of a *cultural difference.*

Intertwining of Perception and Risk: Another Example Adams quotes the author Roald Dahl, relating how he excitedly rode his new tricycle to school each day:

> All this, you must realise, was in the good old days when the sight of a motor car on the street was an event, and it was quite safe for tiny children to go tricycling and whooping their way to school in the center of the highway [Dah86], quoted in [Ada95, p11].

To put Dahl's feeling that it was "quite safe" in context, Adams notes that between 1922, the period about which Dahl was writing, and 1986, the number of children under the age of 15 killed annually on the roads in England and Wales *fell* from 736 to 358, although the amount of motorised traffic increased by a factor of 25. The child road death rate is now about half what it was then; per motor vehicle it has fallen 50-fold.

Changes in Exposure Before one puts this risk assessment down to a perception, one might query whether this reduction in number is not because the roads have become "objectively" safer, but primarily because the exposure of children to traffic is much reduced. Some figures suggest this: 80% of children made their way unaccompanied to school in 1981, for example, compared with only 9% in 1990. The difference, according to surveyed parents, was mostly their worries about the danger of traffic.

Changes in Behavior; Vigilance Adams also wonders how much may be due to changes in behavior: playing *alongside* the street rather than *in* it. The *vigilance* of motorists towards children may have changed, also the children's reaction to the speed, volume and variability of traffic. Measuring changes in exposure effectively presents all but insurmountable problems [Ada95, p13]. The general problem

> ...for those who seek to devise objective measures of risk is that people to varying degrees modify both their levels of vigilance and their exposure to danger in response to their *subjective* perceptions of risk [Ada95, p13].

2.4.7 Risk Compensation

Purpose of Characterising Risk Is To Manage It According to Adams, the Royal Society's purpose in devising risk assessment procedures is to manage the risk. When people respond to their perception of risk by altering their behavior, then management of risk does not operate against a static background that can be measured, but against people's adjustment to a newer risky situation. This adjustment is termed *risk compensation.*

A Model of Risk Compensation Adams proposes a model of risk compensation that he attributes to Wilde in 1976 [Ada95, pp14–16]. This model is based on the following propositions

- Everyone has a propensity to take risks
- This propensity varies between individuals
- The propensity is influenced by the rewards of risk-taking
- Perceptions of risk are influenced by the experiences of self and others with accident losses
- Individual risk-taking decisions represent a balance between the perception of risk and the propensity to take risks

- accident losses are a consequence of taking risks.

Because of the feedback from consequences to perception and the mixing with propensity, it follows that managing risk is an interactive phenomenon. Adams illustrates this idea with what he calls the "*risk thermostat*", Figure 2.1. Just how complicated matters can be to assess when two risk managers, one riding a bicycle and the other driving a truck, meet on a wet curve in the road can be seen in Figure 2.2. One can imagine how complicated this gets with many "risk managers" all at once. With this, Adams hopes to illustrate how simplistic the current assessment methods are in comparison with reality.

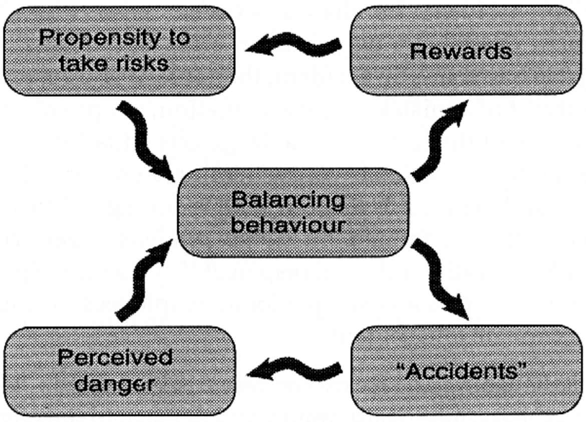

Figure 2.1: Adams's "Risk Thermostat", after [Ada95, Fig. 2.2]

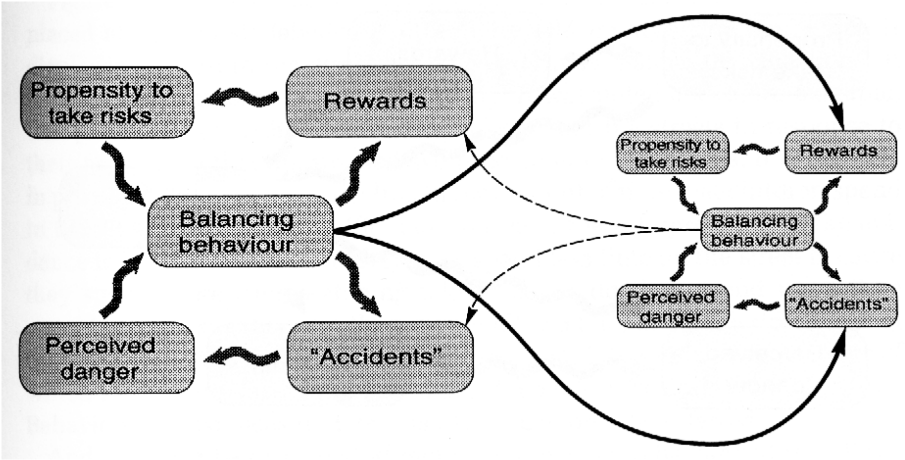

Figure 2.2: Risk Thermostats Interacting, after [Ada95, Fig. 2.3]

2.4.8 Summary: Risk As Cultural Artifact

We may take it, as the Royal Society suggested in 1993, that there is nowadays significant weight given to the two theses that

- assessment of risk is culturally dependent, and
- risk perception is an irreducible and inseparable component of risk itself.

2.5 Cultural Theory

2.5.1 Attitudes to Nature and Risk

Myths of Nature Adams [Ada95] identifies four anthropomorphic attitudes to nature, which stem from the observations of Holling [Hol79, Hol86] concerning different management strategies for managed ecosystems that appeared to be explicable in terms of the managers beliefs about nature. He identified three belief styles, extended to four by Schwarz and Thompson [ST90] and developed into so-called *cultural theory* of risk in [TEW90]. These myths are

nature benign: nature is stable, robust and forgiving of human insult. In the technical vocabulary of dynamics, the state of nature is a *stable equilibrium.* The appropriate management style is *laissez-faire.*

nature ephemeral: nature is fragile, precarious, unforgiving. We must tread carefully on the earth. The state of nature is an *unstable equilibrium.* The appropriate management style is *precautionary.*

nature perverse/tolerant: Within limits, nature can be relied upon, but care must be taken not to exceed those limits. The state of nature is a *local stable equilibrium that is not global.* The appropriate management style is *interventionist*

nature capricious: Nature is unpredictable. The state of nature is that there are *no equilibria.* The appropriate management style is *resignation: do nothing.*

It should be clear that the four models refer to various features of so-called dynamical systems, a field of mathematics which uses the analysis of differential equations to study predator-prey and other ecological systems. Adams suggests this with his diagram illustrating the four myths, reproduced in Figure 2.3. It is clear to those engaged in such modelling that dynamical systems include and are included in other dynamical systems, and is perfectly mathematically in order to consider the union of all such natural dynamical systems. One can identify this with "nature", and plausibly ask about its equilibrium properties, as one can for

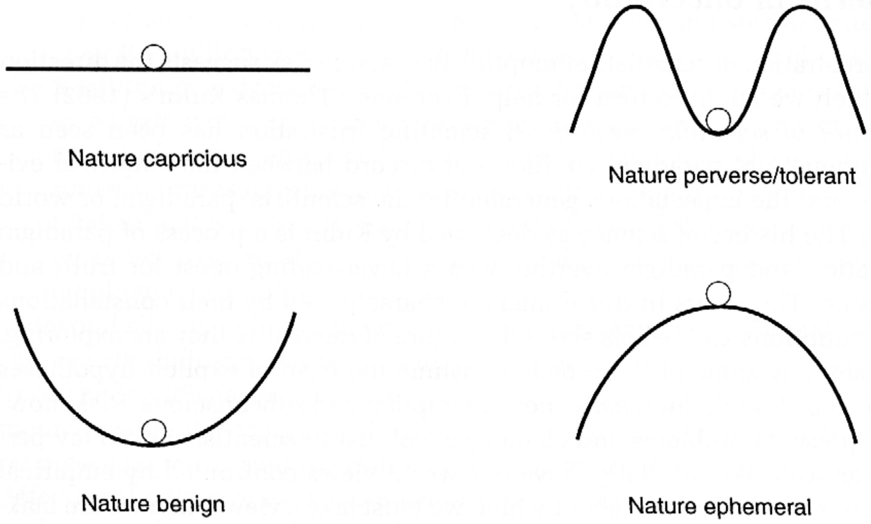

Figure 2.3: The Four “Myths” of Nature, after [Ada95, Fig. 3.1]

any dynamical system. But because the system is so complex, one cannot hope to answer this question definitively. Hence the various “myths of nature” correspond to the variety of possible beliefs about the global equilibrium properties of “nature”. They need not stem from anthropomorphic roots after all. All very reasonable so far.

Applied to Risk-Taking If we regard risk-taking as the management of uncertainty, and this uncertainty concerns the way the “world” is, one can plausibly identify “the world” with “nature”, as long as nature includes human activity also. Thus can the four views of nature be adapted as the background to risk-taking decisions, as in [ST90, TEW90].

- If nature is in stable equilibrium, then I can take risks as I like, strive to exert control over my environment and people in it, and the “world” will accomodate. I am an *individualist* about risk.

- If nature is in unstable equilibrium, then I must manage my risk by consensus to ensure uniformity of action by all and avoid disturbing the equilibrium. I work by consensus under strong group cohesion; leaders arise through force of personality and persuasion; rules from outside the group don’t apply. I am an *egalitarian.*

- If nature has local but not global stable equilibria, then I may act inside defined boundaries, and outside these boundaries others must have the say. I construct management structures; I am a *hierarchist.*

- If nature has no equilibria, then it is not possible for me to manage my risk, that is, to affect the "world" in such a way as to get it to respond more favorably to my wishes. Management is impossible; I am a *fatalist.*

It is clear that, although the "myths" of nature are bound up with speculation about the global nature of a well-defined model, the attitudes to risk just enumerated are indeed myths. They are socially constructed parables about how the world behaves which lead to management paradigms.

Grounding These Paradigms The paradigms may, however, be grounded in abstract social views. Consider the two dimensions, denoted by their extrema, of

- Individualist - collectivist. This dimension describes one aspect of the nature of the human animal. Eagles are individualist, rabbits are collectivist. We use the acronyms *I/C*
- Prescribed/unequal - prescribing/equal. This dimension describes organisation. At one end, social choices are constrained, prescribed, by a superior authority and social and economic transactions are characterised by inequality. At the other end, transactions are negotiated by participants as equals, without externally prescribed constraints on choice. We use the acronyms *U/E*

Given this typology of social organisation, we can categorise the four views of risk as

- Individualists are IE.
- Egalitarians are CE.
- Hierarchists are CU.
- Fatalists are IU.

This placement is illustrated in Figure 2.4 This provides what amounts to a theory of how risk attitudes arise. But does it work to explain risk behavior?

Empirical Evidence is Lacking It has proven hard to identify these attitudes with groups of risk-takers. Adams reports [Ada95, p64] that [Dak91] had "limited success" in trying to substantiate the hypothesis that social concerns are predictable given people's cultural biases. Recent studies assessing the fruitfulness of the four cultural categories in predicting risk attitudes have also doubted whether the categories best describe individuals, and have suggested that a given individual may well have a mix of attitudes to different risk situations. Marris and colleagues [MLO98] distributed psychometric-style questionnaires to residents of

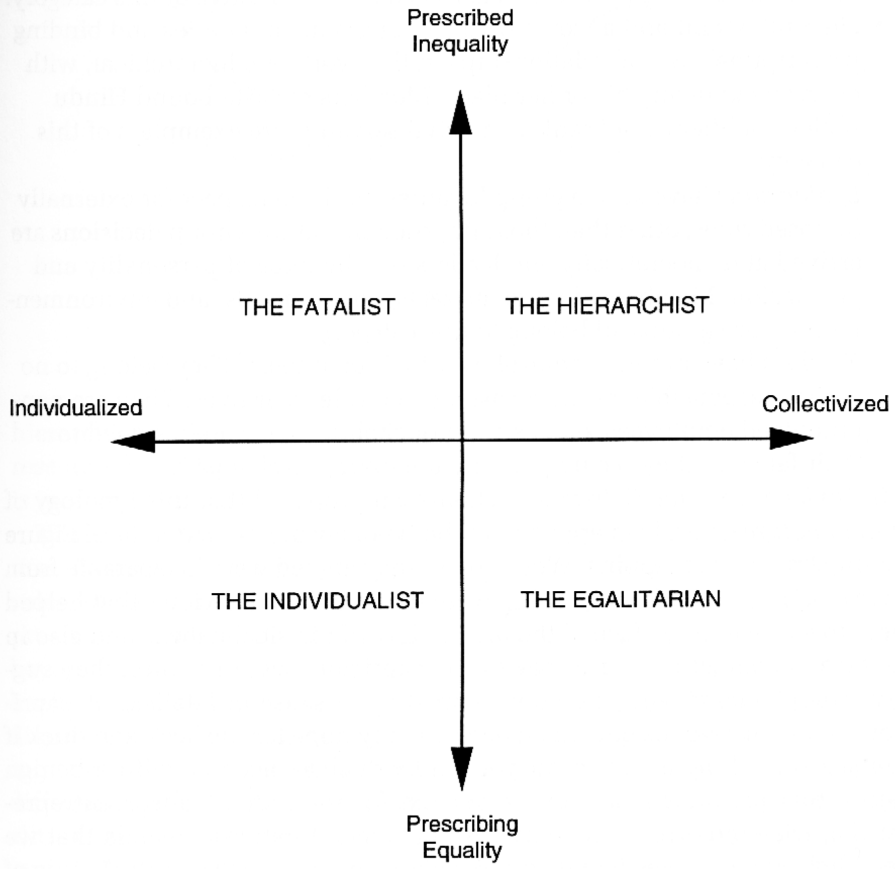

Figure 2.4: The Myths As Four Quadrants in Two Dimensions, after [Ada95, Fig. 3.2]

Norwich, England, and found that psychometrics explained a “far greater proportion” of risk variance than cultural biases as explained by cultural theory. However, they discovered a “key point” that cultural biases were associated with concern about distinct types of risk, and that the pattern of responses in these cases was compatible with that predicted by cultural theory. Furthermore, the psychometric questionnaire could only allocate individual respondents unequivocally to a unique cultural category in 32% of the cases. Brenot and colleagues used a version of a questionnaire developed by Dake [Dak91, Dak92] to test the correlation between cultural bias and 20 social and environmental risks. They found a “weak positive” correlation: cultural bias explained just 6% of the risk variance. They compared with other studies in other countries, and concluded that “new methods, more qualitative and contextual”, are needed to investigate cultural perceptions of risk.

"More Studies Are Needed" So the jury is out on cultural theory. The model is based more firmly on structure and less on parable than some proponents have credited it with. Maybe one can usefully compare the situation with that of individual political views versus party systems.

The "Party System" Analogy A cultural category is like a party manifesto. But I can still have strong political views on a number of issues without following the party line. Party X believes as person A does that that untrammeled libertarianism is the best model for social welfare economics; quite in distinction to party Y, which believes in putting jobless and homeless people up in the local 5-star hotel at tazpayers' expense as compensation for not having a home or salary. However, party Y also believes that the new autobahn should not be built at all, let alone in front of A's house, which is where party X decided to put it. Assigning a party affiliation to A on the basis of this information would be unwise, I propose. Maybe we can pursue an analogy with risk characterisation.

Distinct Attitudes for Distinct Risks Although I may ride my bicycle with relatively great care, I do believe that whether I am assaulted by an auto is largely due to chance rather than under my control, and the residual variance I can affect is limited. By contrast, it is apparently the case that many car drivers believe themselves to have greater control than they actually do – most drivers believe themselves to be "above average" [NS75, Sve81] quoted in [KST82, p469], which is a collective contradiction. However, I may well believe that my career is largely under my control, through my performance, although of course it is significantly affected by my age, where I choose to work, what choices I have that suit my capabilities, and what potential colleagues think of my personal presentation, as well as what political role I play for them. Further, I tend to think that paper acceptance at academic conferences is largely a matter of chance, whereas paper acceptance by journals is much more determined by the relative quality of the contents. Also, while I talk with my colleagues and plan action about matters of mutual concern, I don't necessarily believe that what they agree to and what they do are perfectly correlated. I am thus an egalitarian about bicycle riding, largely an individualist about my career and about journal papers, and largely a fatalist about conference papers and my neighbors' neighborliness quotient. As a whole, my attitudes to risk management appear to be diverse. I would be surprised to find that I were atypical. The research results of Marris and colleagues are unsurprising.

2.6 Perception Heuristics

2.6.1 Problem Presentation Affects Choice

In a series of classic experiments, Daniel Kahneman and Amos Tversky, amongst others, have investigated the probability reasoning of laypeople (i.e., those who are not probability theorists or statisticians, but who might be aware of probability calculations, such as students). One experiment asked practicing physicians to answer the following [TK81]:

1. Imagine that the United States is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the consequences of the programs are as follows:

   - If program A is adopted, 200 people will be saved.
   - If program B is adopted, there is one-third probability that 600 people will be saved, and two-thirds probability that none will be saved

   Which of the two programs would you favor?

2. Imagine that the United States is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the consequences of the programs are as follows:

   - If program C is adopted, 400 people will die.
   - If program D is adopted, there is one-third probability that nobody will die, and two-thirds probability that 600 people will die.

   Which of the two programs would you favor?

Most physicians preferred A over B, and D over C. Note that the two problems are formally equivalent. Programs A and C save 200 and let 400 die; programs B and D give one-third chance that all will be saved and none will die, and two-thirds chance that none will be saved and all will die. Those paying attention to the actual outcomes, if they prefer A over B, should also prefer C (identical with A) over D (identical with B). But they don't, despite being au fait with the numbers. That seems to be a simple contradiction, and would render the majority choices irrational. Furthermore, the expected number of deaths is identical for all four programs. Someone choosing strictly according to expected number of deaths (one common measure of risk in cooperative situations, when a qualitative measure is called for) would have no preference amongst the four choices. One might exhibit a preference for certainty, or on the contrary for the chance of a jackpot, but this would also entail consistent choice, which the majority do not exhibit.

The outcome of this experiment is reproducible, with different populations, in different formulations, and roughly in the proportions of respondents preferring which alternatives. It is a *result* of social cognitive psychology, if anything is. It appears to demonstrate a certain kind of irrationality in choices under uncertainty.

One conclusion is clear.

- The mode of presentation of an uncertain choice, a risk, affects the choice. And how.

2.6.2 Prospect Theory

Exactly how this presentation affects choice is explained by *Prospect Theory* [ST95, pp80–81]. First observe that "saving people" is a *gain* and "people dying" is a *loss*. A preference for a risky outcome over a "sure thing" with the same expected value is termed *risk seeking*; a preference for a "sure thing" over a risky outcome with the same expected value is termed *risk aversion*. The experiment demonstrated that risk aversion holds for gains and risk seeking for losses. This is true in general, except for choices involving very small probabilities. Prospect theory posits the following three phenomena:

Diminishing sensitivity: I am more sensitive to a difference in expected outcome varying between, say, $50 and $150 than I would be to a difference in expected outcome varying between, say, $8,050 and $8,150.

Relative Value: I am sensitive to gains and losses rather than to total wealth.

Loss Aversion: I am more sensitive to losses than I am to gains of equal magnitude.

It turns out that prospect theory can explain many of the apparently irrational, but reproducible, preferences expressed in choice problems. Various other phenomena to complement those explained by prospect theory have been identified.

2.6.3 Other Heuristics

[SFL82] describe other heuristics of risk perception.

Availability: People using this heuristic judge an event as likely or frequent if instances of it are easy to imagine or recall. Aircraft accidents, shark attacks (after *Jaws*), atomic powerplant accidents (after Brown's Ferry, Three Mile Island and Chernobyl). It surprises people to realise that twice as many people were killed on the roads in Northern Ireland than were killed in the sectarian violence over the last quarter century [Ada95, p62]. Since rare events tend to get reported and discussed, in constrast to relatively common events, one might expect people to overestimate the frequency of rare events and underestimate the frequency of common events. Such a result can be seen in Figure 2.5, in which participants were asked to estimate the frequency of various causes of death in the U.S.

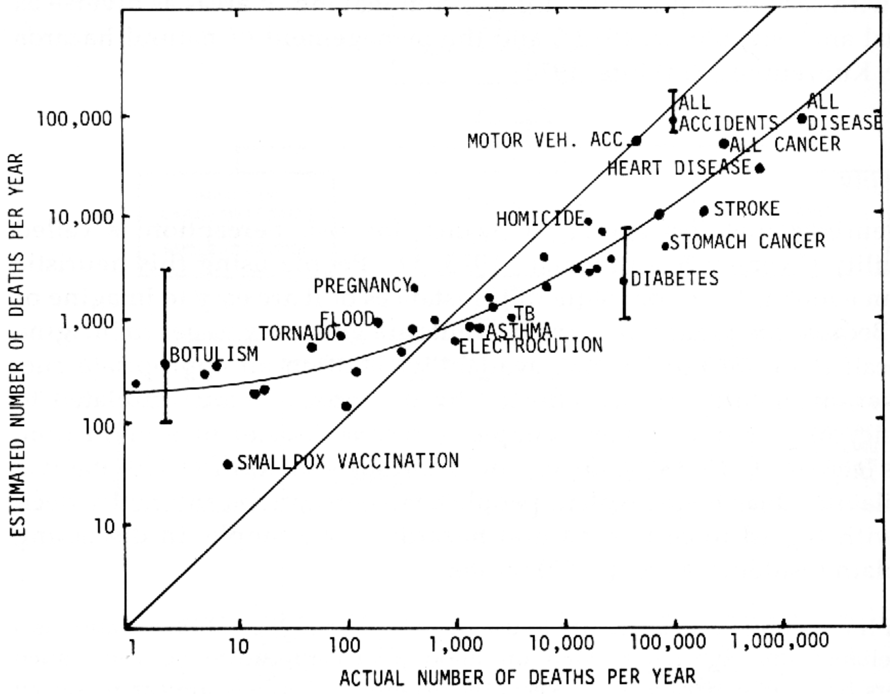

Figure 2.5: Estimates of Event Frequency Plotted Against True Frequency, after [SFL82, Fig. 2]

Overconfidence: People typically have greater confidence in judgements under uncertainty than warranted. One notable result was remarked by Hynes and Vanmarcke [HV76], who asked seven "internationally known" geotechnical engineers the height at which an earth embankment would cause the clay foundation to fail, and to specify "confidence bounds" around this value that were wide enough to have a 50% chance of enclosing the true failure height. In other words, they were asked to guess and hedge their guess to 50% likelihood. None of the intervals from any of the seven experts enclosed the true failure height. The results from this experiment are illustrated in Figure 2.6

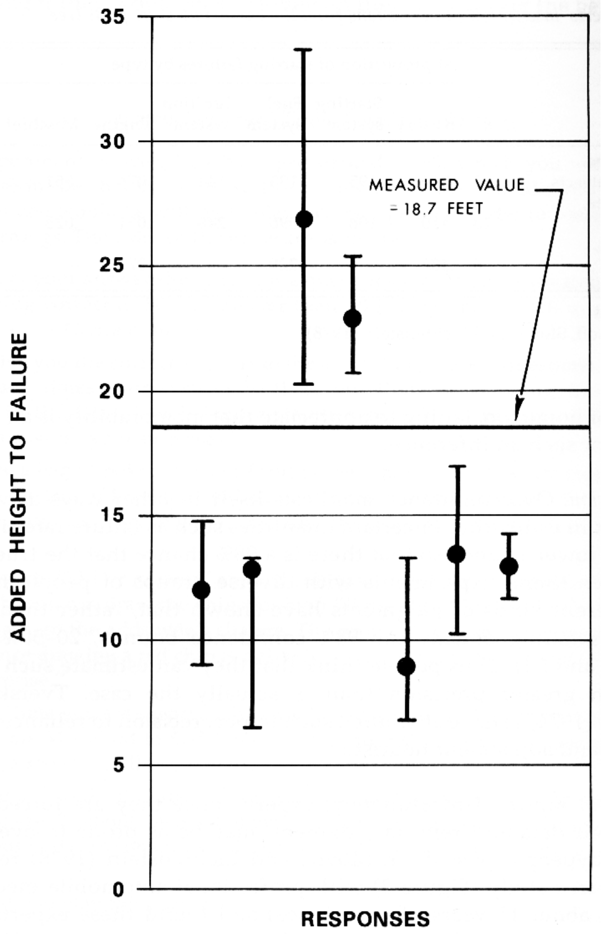

Figure 2.6: Estimates of Dam Failure Height With 50% Confidence Levels, after [SFL82, Fig. 4]

Anchoring: Judgements are "anchored" to initially presented values. For example, individuals were asked to estimate the frequency of death in the U.S. from 40 different causes. When told initially that total annual driving deaths were about 50,000, people tended to give higher estimates of fatalities for all causes than if told initially that 1,000 people die annually in the U.S. from electrocution [SFL82, p481].

That experts as well as "laypersons" are subject to the same heuristics and biases (i.e., these are cognitive phenomena) makes the elicitation of expert opinion under uncertainty far from the objective assessment that one hopes it might be. Various tricks, or "elicitation protocols" have been devised to obtain estimates as free from the effects of heuristics and biases as possible [MH90, Chapter 7].

2.7 Difficulties With the Numbers

Biases make it harder to obtain what one would like to believe were roughly accurate judgements about event frequencies and likelihoods. Further problems are

- often a dearth of available statistics from which desired conclusions could be drawn, and
- the wide variance in calculated values, even given apparently sufficient statistics.

2.7.1 An Example: The Value of a Life

Various attempts have been made to compare how much has been spent overall, across many different industries and social themes, to save how many lives. If an estimate can be made of how much has been spent on safety measures, and how many lives have been saved, one can divide the one number by the other and call it, somewhat crudely, the "value of a life saved". It represents the marginal average cost that society has been willing to spend in the past not to forgo a life.

One of the earliest estimates came up with $200,000 per life saved [TR76]. An estimate a few years later came up with $2m [Rap81]. This is an order of magnitude higher. Nearly twice as high again is the estimate of [Mar92], quoted in [Ada95, p103] of £2m–3m. Thes are hardly figures on which one can place much faith.

2.7.2 Example: Cigarette Smoking Deaths

In a well-known comparison of various ways to increase one's chances of dying prematurely by 1 in 10^6, 1 in a million, [Wil79] (quoted in [FLSDK81, p81]) includes smoking 1.4 cigarettes.

Another estimate can be obtained as follows. Let us take the average male lifetime to be 75 years. According to [CL79] (quoted in [FLSDK81, p82]), cigarette smoking will reduce a life expectancy by 2,250 days, which is roughly 6 years. So let us take the average lifetime of a cigarette smoker to be 70 years. Assume he started smoking at age 15 years, giving a smoking duration of 55 years. The average German smokes 5 cigarettes a day, and since about one-third of the population claim to smoke, we may obtain an estimate of 15 cigarettes a day per smoker, which is about 5,000 cigarettes per year, and thus 275,000 per lifetime. Every second smoker may expect to die from smoking-related causes, so the average number of cigarettes per death is 550,000. This is approximately one-third the figure given by [Wil79]. While not an order of magnitude difference as the the "value of a life saved", this is still a notable difference.

Don't Forget: Probability Allows Anything One should not forget that a probability estimate is compatible with most individual outcomes. If there is an infinitesimal likelihood that I will receive a dose of pigeon dropping on the head today, that does not rule out that I'll be hit by a pigeon every day of the next year. As Chauncey Starr is reported to have said concerning Three Mile Island [Ada95, p51]:

> On the technical side, this accident, while no one wanted it, has a statistical probability that falls within the predicted probability of this type of accident.

2.8 Excessive Prudence Is Disadvantageous

One may wonder if safety problems arise primarily because of a lack of will to fix them. In fact, there can be considerable disadvantage to excessive prudence. [Ada95, p55] lists some.

- People may spend more money on insurance, needlessly
- Motorists may drive more slowly and with more space between vehicles if they believe that there is "black ice" on road, hindering traffic flow
- The construction industry may waste money and resources on "overbuilding", for example, building to earthquake safety standards in regions which have little or no earthquake risk
- On the railways in Britain, excessive expenditure on safety measures raises ticket prices and encourages people to use even less safe modes of transportation such as cars instead.
- An inordinate fear of physical attack leads some women and elderly people not to venture outdoors as often as they would prefer to.

2.9 How Biases May Affect Assessments

2.9.1 Cultural Biases

One way in which the four cultural types may be seen to affect assessments follows from the types of error they may make. Suppose one is attempting to evaluate a hypothesis such as

> Hypothesis: CO_2 emissions threaten a runaway greenhouse effect

- a *Type 1 error* is made when a hypothesis is accepted that should be rejected;

- a *Type 2 error* is made when a hypothesis is rejected that should be accepted.

The four types distribute themselves amongst the error categories thus:

- Egalitarians are at high risk of a Type 1 error and low risk of a Type 2 error
- Individualists are at high risk of a Type 2 error and low risk of a Type 1 error
- Hierarchists would reject the statement of the hypothesis as unspecific on critical limits
- Fatalists would ignore the hypothesis and not attempt to determine its truth or falsity

2.9.2 Evaluation Biases

A common elicitation technique used to attempt to set a uniform value (usually a monetary value) on factors in a risk problem is to ask the value of compensation. This can take two forms:

- What is one *willing to pay* (WTP) for a certain advantage that one does not have.
- What is one *willing to accept* (WTA) in compensation for loss of a resource or capability that one values.

These quantities are used in an attempt to achieve an equitable distribution of risk or of consequences of a course of action. The quantities are not dual. In a successful transaction, the range of amounts that one party is willing to pay overlaps the range of amounts that the other party is willing to accept, but not all risk and compensation problems are of this type. For example, consider asking a fatally ill person what could compensate him for loss of his life. The answer might well be that no amount of money would suffice for him to consider himself suitably compensated. However, the amount he would be willing to pay to have his life saved is rigorously limited by his assets.

In general, WTPs can be very much less that WTAs, and this leads to bias in distribution [Ada95, p99].

2.9.3 An Example: Negotiating a Smoke

Consider two rules for smoking in a compartment of a railway carriage [Ada95, p99].

- Under the *permissive rule*, one may smoke. In this case
  - A smoker may consider a WTA for giving up smoking for his journey
  - A non-smoker may consider a WTP for experiencing a smoke-free journey
- Under the *restrictive rule*, one may not smoke, unless all parties are agreed to it. In this case
  - A smoker may consider a WTP for smoking during his journey
  - A non-smoker may consider a WTA for suffering smoke during his journey

The consequences for smoker and non-smoker alike of the preexisting rules are different, given that WTPs are less than WTAs. Whoever has the right of WTA is likely to prevail. Thus the permissive rule favors the smoker and the restrictive rule favors the non-smoker.

A Personal Comment I am a non-smoker who strongly does not like to breathe air polluted with cigarette smoke, and avoid it wherever possible. But I do believe the decision to smoke or not is a personal one, so have nothing against smoking per se. The difference between the number of smokers in the U.S. and in Germany is tiny. About a quarter of Americans say they smoke and about a third of Germans. The difference is one-twelve - about 8% of the population. However, in the U.S. I have no trouble avoiding smoke when I wish. Restaurants are completely non-smoking or have adequately ventilated non-smoking areas; offices are mostly or entirely smoke-free. This is supported by Federal and State regulations. However, in Germany, there are few regulations. Restaurants are to me so unpleasantly smoky that I do not go to eat in restaurants any more, although that was a hobby when in the U.S. and I went out most nights. My office, although nominally in a non-smoking corridor, is invaded by the strong odor of cigarette smoke many times daily, and by the end of a working day I have noticeable physical effects from it. At bus stops or in train stations with more than two or three people waiting, there will be cigarette smoke. People and exchange students who visit from Great Britain, the U.S. or Ireland have also remarked on the comparative pervasiveness of smoke here. The difference between the biases implicit in the permissive and restrictive rules is real and palpable.

2.10 Professional Attitudes To Risk Management

2.10.1 Engineering Codes of Ethics and Their Consequences

Engineers have to manage risk, whether they are familiar with the tools of technical risk management or not. [Ada95, pp186–189] reports on a conference [Fel90] of engineers concerning “preventable disasters”. The Rules of Conduct for Chartered Engineers require engineers to pay due regard to

- the safety of the public,
- the interests of their client or employer,
- the reputation of other engineers,
- the standing of the profession.

Adams notes that only the first of these has anything to do with safety. The other three are political or social group interests. It is well possible, of course, that failing to pay due regard to safety could ultimately influence the standing of the profession. The point here is, as has been well-documented in, for example, the history of DC-10 carog-door failures [FB92] and the decision to launch the Challenger space shuttle in low temperatures with a known temperature-affected weakness, that engineers who warn of problems are very often ignored by management or have their concerns submerged in the flow of the organisations involved. That is, very often the criteria above may lead to contradictory choices of action. Client or employer *versus* safety, for example. Safety *versus* (sometimes undeserved) public trust in engineering capability (that may, for example in the case of very large and complex software systems, not even exist in appropriate measure).

2.10.2 An Example of What Counts: The Therac-25

But what do practicing engineers expect from the development of safety-critical devices? The history of a series of accidents caused by the Therac-25 radiation therapy machine in the mid-1980’s in [Lev95, Appendix A] indicates what the authors, as well as their readers presumably, single out as problems.

One Possible Attitude

A thought experiment. Suppose the makers of the Therac-25 had said, OK, our machine is killing people because of the way it has been used. However, it has saved many more lives than that. So on balance (using, if you like, an RCBA), there are benefits. Let’s leave everything as it is.

This fits with “standard” evaluation techniques, an RCBA , by hypothesis. What exactly, if anything, would be wrong with such an attitude?

One might say, not all the interests of all the stakeholders are taken into account [Nat96, FLSDK81]. The interests of the people who were accidentally irradiated much more than they should have been were not taken into account. But they could have been, consistently with this attitude. Suppose each of them was informed of the chances of successful radiation and of overexposure beforehand, and they had consented (as we suppose many of them still would have). Then their interests had been taken into account.

Other stakeholders include insurance companies, hospitals, regulators, the company itself, and users of the machine, amongst others. I think it is fair to say that the patients are the primary stakeholders, however. So the claim, that stakeholders' interests have not adequately been taken into account, could well fail. The question remains: what exactly is wrong with the utilitarian argument for doing nothing?

The Implicit Critique

Leveson and Turner analyse the design of the machine. The machine made much more extensive use of software control than its purported predecessors [Lev95, p516].

Turntable Positioning

The authors noted that positioning of the turntable holding the patient was crucial, and that protection against inappropriate positioning, or inappropriate activation of the device with the turntable in a disadvantageous position, was traditionally provided by mechanical interlocks. In the Therac-25, software checks were substituted for many of the hardware interlocks [Lev95, pp517–518].

Operator Interface

The design of the operator interface, displayed on a 25 line by 80 character computer screen, left a lot to be desired. Error messages were "cryptic", containing codes (numbers 1 through 64) for various types of malfunction. The codes were not explained in the operator's manual. Apparently malfunction messages were commonplace and did not usually involve patient safety [Lev95, pp591–520].

Hazard Analysis

A hazard analysis was performed by the manufacturer. The analysis excluded the software. Three assumptions were explicit:

- Programming errors had been reduced by testing on simulator hardware and under field conditions. Any residual software errors were not included in the analysis.
- Software does not degrade due to wear, fatigue, or the reproduction process.
- Computer execution errors are caused by faulty hardware or by "soft" random errors caused by alpha particles or electromagnetic noise.

Information-Gathering About the First Accidents

The first accident led to a lawsuit from the patient involved. It was not officially investigated. The company claimed the first it had heard was when a lawsuit was filed against it by the patient about 9 months later. Others claim that the company was officially notified of a lawsuit about five months after the accident. The accident was not reported to the U.S. Food and Drug Administration, the reponsible government authority, until adter further accidents in the next year.

After the second accident, the company was informed and sent a service engineer to investigate. Regulators and users were informed that there was a problem, although users claimed not to have been told of an accident.

Company Response

The company investigated, made some hardware and software modifications, and informed users that the hazard rate of the new system offered a *five-orders-of-magnitude* improvement over the old system. They had, however, been unable to reproduce the reported hardware behavior in their investigations.

Further Accidents and Response

There were further accidents with the machine. Eleven machines had been installed altogether, six in the U.S. and five in Canada. Altogether, there were six accidents at four different sites. Over a third of the installed sites suffered an accident.

The Software Bugs

Two different software bugs were found in response to two different accident scenarios. Both involved so-called "race conditions", conditions in which a particular instruction execution sequence is required for correct operation, but in which the instructions could and did execute in a different order. The first involved a hardware operation taking about eight seconds. The operator was able to change certain settings, and these changes were reflected on the terminal screen, but the machine could not correctly attend to the desired changes until after the eight-second hardware cycle. A missequenced series of operations followed [Lev95, pp534–537].

The second bug also required an operator action, which triggered the missequencing of operations when it occurred simultaneously with an internal software operation [Lev95, pp542–544].

The Causal Factors

The authors invoke the following causal factors:

- Overconfidence in Software
- Confusing Reliability with Safety
- Lack of Defensive Design
- Failure to Eliminate Root Causes
- Complacency
- Unrealistic Risk Assessments
- Inadequate Investigation or Followup on Accident Reports
- Inadequate Software Engineering Practice

One remarks first that all these causal-factor statements are value judgements or negative human attributes: *overconfidence, confusion. failure, complacency, unrealism, inadequate practice.* This suggests

- that there is a standard, or many standards, which this particular machine and its development did not meet
- that many of the causal factors were human failures which need not have occurred

There is thus a strongly moral tone to this assessment. No one is saying "well, this machine is really complicated and we know it fails but no one knows how to do better than this". On the contrary, the authors are saying "best engineering practice was not followed in this and this and this respect".

Conclusion: What Counts To Engineers

At the heart of the reports of many accident investigations lie similar attributions. Overall, they can be summarised thus.

> *We know how to do better and we could have done better in this case.*

This is strictly a moral judgement. I believe it may distinguish engineering safety concerns from other technological areas in which risk and uncertainty assessments need to be taken into account.

On a final, not completely satisfactory, note, I remark that the injunction to

- Use best practice and perform as well as possible

is not generally part of codes of engineering ethics, for example in Section 2.10.1 above. It is, however, becoming enshrined in increasingly many standards governing certification of teleological systems.