SUBJECTIVITY IN RISK ANALYSIS

Felix Redmill
Redmill Consultancy
Email: Felix.Redmill@ncl.ac.uk

Report
July 2001

This report was not published, but it formed the basis of a number of shorter articles that addressed subjectivity in risk analysis.

1 INTRODUCTION

People make decisions about risks without consciously attributing numeric, or even qualitative, values to them. However, if a comparison is to be made between the costs and the benefits of risk reduction - as called for by the Health and Safety Executive (HSE 1992) - risk values must be derived.

Risk values are arrived at via the process of risk analysis. In many quarters, this is assumed to be objective, and its results - the risk values - to be correct. Yet, as will be shown in subsequent sections of this report, all stages of the process involve subjectivity, in some cases to a considerable extent. Always there is reliance on judgement, and, as in all cases in which judgement is called for, there can be no guarantee that it will be made to a reasonable approximation, even by an expert. Indeed, it may be - and sometimes is - made by an inexperienced novice. The need for judgement introduces subjectivity and bias, and therefore uncertainty and the likelihood of inaccuracy. The results obtained by one risk analyst are unlikely to be obtained by others starting with the same information.

Further, there is a natural impediment to arriving at 'correct' risk values. Although definitions of risk do not explicitly refer to time, the future is implicit in them. Risk does not define a current problem or a future certainty, but rather the *potential* for *future* harm. Charette (1989) reminds us that risk management is not about future decisions, but about the future of decisions that we must take now - in other words, about the speculative or uncertain aspects of the outcomes of our decisions.

Thus, risk may be estimated but it cannot be measured (Gould et al 1988). Risk values cannot be assumed to be 'correct'. The United Kingdom Interdepartmental Liaison Group on Risk Assessment recognises this in saying that risk assessment is 'a tool for extrapolating from statistical and scientific data' to arrive at 'a value which people will accept as an estimate of the risk attached to a particular activity or event' (UK-ILGRA 1996). The reference to statistical data suggests quantitative risk assessment, but qualitative risk assessment may also be used. Relevant questions to be asked are whether the estimate is a sufficiently good approximation for the purpose in hand, and what confidence there is in it.

Speaking of the process of risk determination in technological systems, Lowrance (1980)

pointed out that estimates of risk, whether made by scientists or lay people, cannot escape containing elements of subjectivity. These, he said, enter into 'the very defining of the questions, and into the designing of the experiments used in assembling evidence, and then into the weighing of the social importance of the risk.'

While some risk analysts may be unaware of the subjectivity that Lowrance points to, many recognise the role played by human judgement. Indeed, Okrent (1998) says that 'senior risk analysts themselves understand the very major role played by subjective opinion'.

However, in spite of their awareness, engineers do not in general take steps to neutralise subjectivity in their risk analyses or to qualify their results in the light of it. Wharton (1992) advises us that, 'Failures to cope with uncertainty in the management of technological risk abound. Their causes include overconfidence in scientific knowledge, the underestimation of the probability or consequences of failure, not allowing for the possibility of human error and plain irresponsibility concerning the potential risk to others.'

Wharton goes on to say, 'And yet to avoid such risks by adopting an overly conservative attitude to technological innovation may be to deny the potential benefits to shareholders, employees and society.' Indeed, we need to accept risks, but, in doing so, we should recognise and allow for our assumptions and the scope for inaccuracy.

The purpose of this report is to show the ways in which the process of risk analysis is subjective. This is not intended as a destructive dissection, for a major benefit of risk analysis is that it requires thought and judgement and is not (or should not be) merely the application of a set of rules. However, where subjectivity is arbitrary and could be reduced, or where its better understanding could reduce the likelihood of error, its exposure could be beneficial. Further, an understanding of the subjectivity and the scope for error of their analyses could lead risk analysts to recognise their assumptions and consider more fully the confidence that they can reasonably have in their results. We might usefully heed Schneiderman's (1980) words: 'If we know all our numbers are likely to be crude, and behave that way, we may even behave sensibly. At least we will have an escape route to follow when we discover we have made a wrong decision. If our numbers are too firm, or if we believe they are firm when they are not, I am afraid we may bully ourselves into doing bad things, efficiently.'

In the next section the process of risk analysis is briefly summarised. Next, its stages are examined in more detail and the subjectivity of some of the techniques used in risk analysis considered. Then, the subjectivity of human reliability assessment (HRA), the risks posed by management but not addressed in risk analysis, and the psychological biases introduced by human mental heuristics are discussed. Finally, the report is concluded with a discussion of where further research might bring improvements to risk analysis.

2 THE PROCESS OF RISK ANALYSIS

The process of risk analysis can be divided into stages or sub-processes. In most literature, for example Shrader-Frechette (1991), the international safety standard IEC 61508 (IEC 2000) and Railtrack's safety management manual, 'The Yellow Book' (Railtrack 2000), three stages are named:

- Hazard identification;
- Hazard analysis;
- Risk assessment (or evaluation).

The purpose of the first stage is to identify the hazards that could lead to breaches of safety. That of the second stage is to analyse the identified hazards so as to estimate the frequency and severity of potential harm and thus define the risks that they pose. That of the third stage is to assess the risks against defined criteria in order to determine their tolerability.

The terms 'risk analysis' and 'risk assessment' are not consistently defined. They are used differently by different authors, and sometimes they are used synonymously or interchangeably. Here 'risk analysis' refers to the total process and 'risk assessment' to its final stage.

Risk analysis is 'generic' and may be applied to any situation and any form of decision-making, from defining policy and strategy, through all levels of planning, to day-to-day tactical decision-making. Both the nature of the application and the purpose of the analysis influence a number of factors, such as the level of formality, the techniques used, and whether a quantitative or a qualitative approach is taken. Thus, it is both useful and important to define an initial planning stage to precede the three technical stages defined above - as recommended by the Department of the Environment (1995). The four stages of risk analysis, in order, are therefore:

- Definition of scope;
- Hazard identification;
- Hazard analysis;
- Risk assessment (or evaluation).

3 DEFINITION OF SCOPE

Whether or not the definition of scope is defined as a stage of risk analysis, it must be done. It necessarily involves the judgement of those planning the analysis, it will influence the nature and direction of the analysis, and it will be a predisposing factor on its results. Decisions must be made about both the study itself and the system to be studied, and both sets of decisions involve considerable discretion. Kasper (1980) says, 'The very choice of questions to be asked, issues to be considered, and methods to be used involves judgement.'

The terms of reference of a study may place limitations on where information is sought or the sources of admissible information. The way in which a study is conducted (regardless of the terms of reference) can have the same effects. For example, if the public is not consulted, certain perspectives and opinions, and perhaps the main sources of opposition to a proposal, may be precluded. Wynne (1980, 1982) reports on how the Windscale Public Inquiry in 1977 was predisposed toward the evidence of the 'experts' and how the approach of the project's opponents, from a different frame of reference, set them at a disadvantage.

The 'system' to be analysed may be a proposed policy as well as a tangible system, and, in any case, its boundaries - physical, geographical and logical - need to be defined. Then the study sets out to identify and analyse the risks that may be posed to people, property or the environment on the other side of the boundary. These, in turn, depend on the inputs and

outputs across the boundaries, some of which may only be observable from some perspectives, or acknowledged by some participants, so any constraint on what may be considered in the analysis could severely affect the results.

The terms of reference may also place an outer boundary on the scope of the analysis. If, as in many cases, the system is an industrial plant (equipment), and the study is limited to the risks posed within the factory, then, by definition, any risks to the public are not considered.

Similarly, risk analyses may be limited to one type of risk, say financial risks that management are interested in, while excluding other types, say safety risks that the public may be more concerned about.

Thus, the results of risk analyses may be distorted or predisposed by the definition of the terms of reference, the exclusion of certain types of evidence, the definition of the study and system boundaries, and other aspects of the definition of scope. However objective or accurate the technical aspects of the analysis may be, the nature of its results are to a more or less extent predetermined by its subjectively defined scope and, in general, by the strategic planning of the analysis itself.

Risk-tolerability decisions to be taken in the fourth (risk assessment) stage of the analysis are also influenced by the first stage, for criteria against which they will be made are defined then. Decision-makers need to be identified, the decision-making process defined, and its mechanisms put in place. If politicians, the public, or other non-experts are to be involved, plans need to be made about what risk information will be communicated to them, and how and when it will be communicated. Clearly these decisions at the first stage of the analysis, taken subjectively, will influence the risk-tolerability decisions to be made later.

4 HAZARD IDENTIFICATION

The purpose of this activity is to identify the sources of risk - the things that can go wrong and lead to a breach of safety. The nature of the hazards depends on the circumstances. For example, in an industrial plant hazards might include failures of equipment, human error, and the use of equipment outside its design specification, whereas in the formation of high-level policy they may be the potential causes of societal impact or environmental problems.

In any case, the aim should be to maximise the identification of hazards. An individual can perform a rudimentary means of hazard identification simply by pondering the circumstances, and this may be adequate in a low-risk situation. But in the fields of industrial and environmental safety, it is expected that all 'reasonably foreseeable' hazards (IEC 2000) should be identified, the implication being that professionalism - both in the relevant field and in hazard identification - should be brought to bear. To assist the professionals in their study, a number of techniques have been developed (see Storey 1996 for an overview).

In some well understood situations or systems, the use of a checklist may be an adequate method. For example, the annual Ministry of Transport (MOT) test of motor vehicles is based on testing, against predefined criteria, a list of components that would be hazardous if in poor condition. However, the adequacy of a checklist depends on a thorough understanding on what could go wrong. Thus, without extensive past experience, careful observation, and

documented fault and hazard logs, a checklist would not be soundly based. Moreover, its adequacy also depends on the circumstances of its use being the same as those in which it was created; if they differ, the checklist could be out-of-date and dangerously misleading. Checklists, even when appropriate, need to be reviewed periodically (the MOT checklist has been updated many times).

In systems which are not so well understood, perhaps because they are only now being planned or designed, techniques which employ the creativity of human investigation are required. Brainstorming is sometimes used, but although it is creative there is usually little formality in the process. Information for hazard identification may also be derived from audits and formal or informal interviews with staff.

The most powerful method in use today is HAZOP (hazard and operability studies). First developed in the chemical industry (CIA 1977), it was later extended for use with systems involving software (MOD 1996, Redmill et al 1999). In recognition of the fact that no individual is likely to identify all possible hazards (Redmill et al 1997), this technique calls for a number of viewpoints to be represented. Not only is a team essential, but the leadership of the team and the planning of the study are also crucial to success. Because the method depends heavily on the judgement and activity of humans, formality is a key to its efficiency and, thus, to the identification of the maximum number of hazards.

Yet, the formality, the need for a team, and a methodological approach - the features essential to success - can also be the seeds of failure. A study can take a long time (in some cases, several weeks) and be expensive, so it is not unnatural that managers often seek to reduce the time and effort spent on HAZOP studies. Key features for success are the holding of a sufficient number of study meetings, the presence of appropriate team members (rather than simply those who are available), and the leadership of the study by a trained and competent moderator. But all these aspects are open to the discretion of management and therefore subjectively determined.

If they, and other study parameters, are compromised in order to reduce costs, the inevitable result is both an inefficiently conducted study and ineffective hazard identification. Further, it is common for this to be blamed on the hazard identification process, which is then discredited. This influences all subsequent risk-analysis activities, for hazard identification is the foundation of risk analysis and those hazards not identified are not analysed and will not be mitigated.

Another factor prejudicial to maximising the identification of hazards is a tendency to pay particular attention to the distinguishing aspects of the case in hand and to neglect or reject those aspects which relate it to other cases of the same type or which resemble them. There is a human tendency to perceive problems as unique when we would benefit by seeing them as examples of a wider class (Kahneman and Lovallo 1993). Thus, we take the 'inside view' rather than the 'outside view'. Taking the latter would lead us to ask such questions as: 'What happened on the last occasion that we did something like this?' and 'Has anyone else done something like this and, if so, what happened?'

By taking the inside view, we fail to consider, or even to recognise, relevant information and, thus, we neglect the lessons that we might have learned and the experience that could be appealed to. We are likely to be overconfident in our plans (e.g. our system's design) and to

overemphasise their virtues, thus overlooking their weaknesses.

A procedural way of neutralising the inside view is for a team rather than an individual to engage in hazard identification. However, the team needs to be carefully chosen (Redmill et al 1999). Members must have different experience, responsibilities, and perspectives, for they need to complement each other. Beware the 'groupthink' of individuals with similar experience and outlook (Janis 1982), for, ironically, they strengthen the conviction that their inside view is both correct and good.

A further technique that is often used for hazard identification is fault modes and effects analysis (FMEA) (IEC 1985, BSI 1991a), often called 'failure modes and effects analysis'. This seeks hazards by examining the effects of the failure of each component of a system. The need for a team is not often emphasised in this method, so it is often carried out by a single person. However, an individual lacks the multiple viewpoints required in hazard identification, is subject to the inside view, and is unlikely to carry out a thorough investigation.

FMEA is likely to miss hazards that result from the interactions of components rather than from the failure of the components themselves. As Leveson (1995) has pointed out, such hazards are frequent in modern complex systems, particularly those in which control is provided by software.

Thus, not only is hazard identification in general dependent on subjective judgment, but different propensities to error are introduced by the different techniques that are available. It is important, at the definition-of-scope stage, to determine which techniques are most appropriate to the study, given the nature of the system to be studied. Then, in planning the study, the neutralisation of subjectivity should be considered. The range and types of hazards in even small enterprises or projects are so large that no single method of identification is likely to uncover them all. A combination of methods is most likely to be successful.

Whether the subject of risk analysis is a high-level policy or an industrial system, hazard identification can never be considered to be complete. Lowrance (1980) observed, 'We simply commit the sin of pride when we think we have been so smart as to have forestalled absolutely every possibility of failure.' Not only should the search for hazards never cease, but also a formal means of hazard identification should be carried out at several stages of the life cycle (Redmill et al 1999).

It should first be performed at the earliest possible stage so as to inform the design of the policy or system. At this point, when only a high-level plan is available, the hazards identified will be 'system-level', but addressing these at this early stage should allow many major problems to be avoided later. Indeed, the hazard-avoidance measures devised at this stage will become specified constraints on subsequent stages of development of the system or policy.

Hazard identification should also be carried out by inspection of the design after this has been prepared, and at later stages in the life cycle too, such as when the policy or system has been operational for some time. At each of these points in time, a new set of circumstances prevails and there is new information to be considered. In all cases, however, the aim of the study is contrary to that of the designer. The latter is concerned with making sure that the

policy or system meets its objectives - perhaps efficiently - while the hazard analyst is concerned with identifying how it may go wrong (Ballard 1992) or be unsafe.

Not only should feedback from audits and interviews be continuously screened for indications of hazards, but also a culture should be nurtured such that all hazards are recorded and reported and not denied or 'left till later'. Hazards not identified are not analysed or managed. Indeed, one of the greatest sources of error in risk analysis is the failure to identify all the hazards or all the ways in which they occur. Ever greater effort is expended on attempts to improve the accuracy of the estimates of the probabilities and consequences of hazards that have been identified while, in many cases, even greater hazards lie unseen (Kletz 1999).

5 HAZARD ANALYSIS

Identified hazards are analysed to determine the risks that they pose, so analysis must be preceded by hazard identification. This does not guarantee that all possible hazards have been uncovered, for that would be almost impossible to achieve, and even if it were achieved it could not be proved. Nor does it mean that all the hazards that will be identified have been identified, as it would be extraordinary if further hazards were not revealed throughout the life of a system or policy. But it does mean that no analysis can proceed if no hazards have been found. It also means that the coverage of analysis is constrained by the thoroughness of identification.

Given that risk is taken to be a function of probability and consequence, analysis involves examining each hazard to determine its likelihood of leading to or maturing into an accident, and the potential consequences if it did. The event that gives rise to a hazard (e.g. the failure of a component or person) may be local, but the hazardous effect is almost always at the system level - i.e. at the boundary between the system and 'the rest of the world'. Thus, the chain of cause and effect between the hazard and its safety-related consequence needs to be determined.

The more sophisticated hazard identification techniques mentioned above (e.g. HAZOP and FMEA) also include some hazard analysis. In FMEA, this takes the form of examining the hazard's effects at both the local and the system levels. Because in an FMEA the components are examined, the first results may be in terms of reliability - for example, the failure of a component would result in the failure of a subsystem. It may only be when the analysis is extended to the full system level that it becomes clear that these failures could or would result in a 'hazardous event' (e.g. an accident or an irreversible environmental change). In both HAZOP and FMEA the analysis extends not only forwards in the direction of consequence but also backwards in the direction of ultimate cause. This is useful because it helps in the deduction of the probability of occurrence and also may suggest how the hazard might be eliminated or its risk mitigated.

Starting from the identified hazards and working towards a 'top' hazardous event takes a bottom-up approach and creates a chain of cause and effect leading from the initial hazard to its system-level effect - the hazardous event or accident. It is also possible to carry out a top-down analysis, commencing with the top events and determining the chain of cause and effect in search of their causes. This usually takes the form of a fault tree (Veseley et al 1981, IEC 1990, BSI 1991b), a key technique in risk analysis.

In all but the simplest cases, it is advisable to use both bottom-up and top-down analyses, for they complement each other. Although many of the causes of the top hazardous events are the hazards found in bottom-up analysis, it is not unusual for new causes to be discovered in a top-down fault-tree analysis. Nor is it unusual for significant inconsistencies to exist between the two approaches. It is therefore beneficial not only to employ both methods but also to compare them for consistency (Crawford 2001). Yet this is not often done. Indeed, in some modern computer-based tools, fault trees may automatically be derived from the results of a bottom-up technique such as FMEA. Naturally it can be claimed that the fault tree is 'correct' with respect to the model produced from the FMEA. But, being the result of human judgement, that model is subject to human error and bias (e.g. the inside view) and is unlikely to be wholly complete or correct. Moreover, the opportunity for cross-checking between top-down and bottom-up methods is lost. Misplaced confidence in the correctness of the fault tree is likely to result.

The results of a risk analysis depend on the techniques employed, the ways in which the techniques are used, and the consistency with which different techniques are used with respect to each other. There are possibilities for error, omission and inconsistency in and between the techniques.

5.1 Consequences

When it comes to determining potential consequences, there is a decision to be made as to what 'consequence' should represent. Sometimes the worst possible consequence is used, sometimes the worst credible, and sometimes the most likely. The decision on which is appropriate should be taken at the 'definition-of-scope' stage, and, clearly, it has an influence on the estimated risks. Further, what is deemed to be the worst possible or credible or likely consequence is a subjective decision, and the potential consequences of an incident, used in risk estimates, will be greater or smaller depending on how it is made.

Consequence also depends on which event in a chain of events is considered to be 'final' or of interest in the circumstances. For example, in transport a hazard might lead to the loss of control of a vehicle, which could lead to an accident, which in turn could lead to a loss of life. Each of these possible consequences depends on different assumptions, carries a different probability of occurrence, and might be of interest in different instances. For example, in Railtrack's safety management manual (Railtrack 2000) it is the loss of life that is of interest, but in the motor industry's guidelines (MISRA 1994) it is the controllability (or loss of it) of the vehicle.

Thus, the decisions on how to define consequence, at both the definition-of-scope and analysis stages, are subjective. So too are the predictions of what the actual consequences might be. And in this case the subjectivity may represent the objectives or mindset of a particular industry sector rather than that of a single person. The worst possible consequence of a financial investment would be total loss, but in the sinking of a cargo ship it could depend on numerous factors and be open to conjecture. In transport, even the worst possible case is open to opinion, for the number of vehicles involved, the circumstances and location of the accident , the number of lives lost, and the extent of property and environmental damage, cannot be defined in advance with precision. The results of a risk analysis are influenced,

often strongly, by subjective decisions on potential consequence.

5.2 Likelihood

When a quantitative approach to the determination of likelihood is taken, the task is to derive probabilities. In spite of the appearance of accuracy that numeric results might suggest, the derivation is subject to assumptions that are not always made explicit and may be based on guesswork.

One assumption implicit in all statements of probability is that of randomness. Yet, in many components of modern systems - those based on software and human involvement - the assumption may not be valid. Unlike mechanical and electromechanical system components, software does not wear out. Its failures are not random but systematic and would be repeated if the circumstances that gave rise to them were repeated. Yet, once revealed, they may be fixed and therefore will not be repeated - though the repair may introduce other potential causes of failure. Thus, if the rates of future software failure are given in numeric probabilities, they should at least be questioned.

Similarly, the behaviour of humans cannot be assumed to be random. Yet many probabilistic models for human reliability assessment (HRA) exist (see Caccuabue 1997, Moieni, Spurgin and Singh 1994), and their use introduces assumptions that may not maintain in any given case. Indeed, although probabilities are derived, the approach taken in most cases is based principally on human judgement. Thus, the results are at best reasonable approximations and at worst wild guesses - but, in general, they are misleading when presented as 'probabilities'.

But even for hardware, for which failure may indeed be random, it is questionable to what extent the derived probabilities truly represent what they are purported to mean. Crawford (1999) pointed to the fact that some of the probabilities are of incredible magnitude and cited a trial of a weapon system whose component supplier claimed a failure rate of 9.116 in 10^9 operating hours. Two of four devices failed in the first operational hour. The two different modes of failure that were experienced had not been included in the elaborate calculations that resulted in the apparently precise number.

This omission of possible causes of failure (or dangerous failure, if safety is the main criterion) is not unusual and, as will be seen below, no guarantee can be given that it has been avoided. It renders risk calculations spurious. It was highlighted in the case of the failure of the British nuclear submarine *Tireless* in 2000. The problem was a crack in a pipe in the cooling system near to the reactor (and on the wrong side of the valve). Such a crack had not been considered in the probabilistic analysis of submarine failure and, based on the assumption of its infallability, the pipe was never checked during maintenance on any of the twelve submarines of the type in question (hunter killer) at any time in their lives. Yet, when such a crack caused the breakdown of a submarine, and checks of the other submarines were made, seven of the twelve were found to have indications of similar cracking and the other five were not wholly exonerated (BBC 2000). Thus, subjective assumptions and omissions through human judgement or negligence can have enormous implications on the accuracy and relevance of probabilistic calculations. When something fails, or an accident occurs, it may be for a reason not considered in the analysis.

Crawford (2000) enquires into another fundamental issue - whether statistical inference can indeed take us from the past to the future. Such statistical inference, basing predictions of future events on historic data, is the foundation of probabilistic risk analysis. In some cases analyses employ inadequate or inappropriate data and are therefore wrong. But even when the data are statistically valid, what is the justification for basing predictions on them? Crawford calls on Deming (1975), who points to the need for the conditions in which the data are to be applied to be the same as those in which they were collected. Further, Feynmann (1998) says that historical results often rest on crucial conditions that are unrecorded, often unnoticed, and unrepeatable, and that there is therefore no objective way of extrapolating from past to future or for assigning a numerical probability that a prediction will be correct. He says that prediction means applying judgement, which bases numerical analysis on subjectivity.

Consider a simple example of inconsistent conditions. A component of risk in system operation is the probability of operator error. Suppose that after that probability was derived the operators were given advanced training or offered a bonus for not making certain types of errors. It might be expected that the probability of error would be reduced - as it was intended to. Or, suppose that, having produced the probability of operator error, the company involved reduced its operating budget and recruited less qualified staff, or cut back on training. The probability of error would now be increased, but to an unknown figure. This is a trivial example, but the same principle applies to all aspects of probabilistic analysis. Past history is an unreliable guide to the future if consistency of conditions is not guaranteed.

In any case, even if the historic data were extensive, and the conditions remained constant, the past frequency only tends to become an accurate future probability as the time of observation tends towards infinity. This is the law of large numbers. Now, what may be taken to be a reasonable approximation to infinity depends on the application, and in some cases may not be large, but care needs to be taken in making assumptions about the predictive value of historic data. Many risk-takers in casinos have lost their shirts because they have implicitly assumed that the law of large numbers applies to small numbers.

When historic data are not available, as in the case of new technologies and products, mathematical models are often devised for assessing probabilities. For example, if a new drug is found to induce cancer in mice, but cannot be tested in large doses on humans for ethical reasons, how can its safety be determined? How valid is a projection of the drug's effect on mice to its effect on humans? Assumptions must be made. Suppose the drug is found not to be carcinogenic in lions. Should its operational mechanism in humans be assumed to approximate that in lions or that in mice? Some experts may opt for one and some for the other. Ignoring the lion example, what should be assumed when using the drug's effect on mice as an indicator of what it would be in humans? It is usual for experts to differ, for it depends on the assumptions made about the relationship between the observed effect and the administered dose.

At a recent conference, there was a paper that presented a mathematical model, based on sets of probabilistic equations, for determining the likelihood of certain hazardous events. At the end of the talk, a delegate pointed out to the presenter that he had suggested that in some cases there were very sparse data for input into the equations, and he asked what was done in such cases. The presenter dismissed the question. 'That is not a problem,' he replied. 'When there is sparse data we just employ an expert to tell us what he thinks.'

The mathematician neglected to wonder what the expert based his opinion on if historic accident data did not exist. In computing there is an acronym, GIGO, which stands for 'garbage in, garbage out'. It is appropriate not only to software algorithms but also to mathematical risk models. Precision is not the same as accuracy and should not be assumed to imply it.

The foregoing discussion has concerned quantitative risk analysis. In qualitative analysis techniques, subjectivity is an integral and even more obvious part of the process, so in respect of them nothing need be added to what has already been said.

5.3 Fault Tree Analysis

A fault tree is a cause and effect network. It starts with a final hazardous event and works backward logically to ultimate causes. The ways in which events combine to cause higher-level events are represented in the tree by AND and OR logical units (gates). If the occurrence of any of a number of events would cause a higher-level event, the causal events are combined by an OR gate; if a given result would require the occurrence of two or more events, these are combined by an AND gate (see Figure 1). A fault tree may simply model the ways in which higher-level events can be caused by lower-level events, but if probabilities can be attached to the various leaves on the tree, their combinations can be calculated all the way up the tree to the top level. Those combined by an OR gate are added and those combined by an AND gate multiplied.

Fault tree analysis was developed in the field of reliability theory, where, typically,

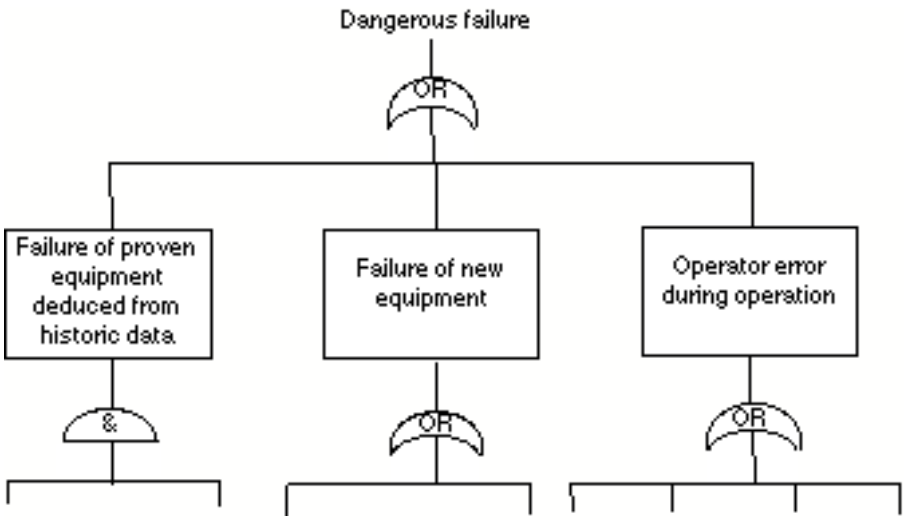

Figure 1: A simple fault tree

systems comprise known and identifiable subsystems and components. In such cases, fault trees may be objectively constructed. Traditionally, too, components were predominantly mechanical or electromechanical, and their fault histories allowed probabilistic determination of the likelihood of the top event. Fault tree analysis thus came to be thought of as an objective technique.

The use of probability theory assumes not only randomness of component failures but also independence from common causes of the failures. However, in using fault tree analysis to

model the faults of modern systems, these assumptions may not be valid. The inclusion of software, as well as complex electronic hardware, means that systematic, rather than random, faults predominate. Human components can introduce common causes. Further, the extension of fault tree analysis to situations of uncertainty, such as policy decisions and safety scenarios, means that the trees are subjectively constructed rather than modelled to reflect directly the combinations of components in systems. Moreover, the probabilities are no longer derived only from historic data but are assembled from various sources, some of 'low pedigree' (see below). Then, the higher the level in the tree, the more combinations of perhaps unreliable numbers it has taken to get there, so the more sceptically the result should be treated. Subjectivity is unavoidable.

In general, the estimation of probabilities may be more or less accurate, depending on the sources of information. Indeed, in any given fault tree, the probabilities will almost certainly be derived from different sources, and risk assessment of any but the simplest systems will depend on inputs from a number of people, all of whom will need to supplement objective analysis with judgement. They will have had different experiences on which to base their judgements and will be more or less competent to make them.

Funtowicz and Ravetz (1990) illustrate the point that a system's probability of failure may be dependent on the combination of various items of (quantitative) information provided by various persons (see Figure 1). One source, they say, may be historic failure data, from which a reliability estimate for certain equipment may be derived. They refer to this source as being of high pedigree. Another item could be the failure probability of a new piece of equipment, estimated by an acknowledged expert, and they refer to this as being of medium pedigree. The third item might be an assessment, made by a recent graduate, of (say) the expected reliability of staff during maintenance, and they refer to this data as being of low pedigree.

Even the data sources that Funtowicz and Ravetz refer to as being of high pedigree may lead to false probabilities. As Feynmann (1998) pointed out, crucial conditions under which data are collected, and the assumptions made in their derivation, are often not recorded, and the conditions under which the derived probabilities are used predictively may be very different from those under which the data were collected. But confidence levels are not commonly assigned to the probabilities, so there is no recognition by users of the fault tree of when confidence is low.

When expert judgement is relied on, in the medium-pedigree case, there is considerable reliance not only on the relevant experience but also on the biases of the expert. When investigating the 1986 disaster of the space shuttle *Challenger*, Feynman (1989) found that engineers at the National Aeronautics and Space Administration (NASA) considered the chance of a shuttle failure to be about 1 in 200 and, at best, 1 in 1000. But NASA management took the figure to be 1 in 10^5 - which, as Feynman pointed out, would mean that a shuttle could be launched every day with an average of almost 300 years between accidents. This led to Feynman's comment: 'If a guy tells me the probability of failure is 1 in 10^5, I know he's full of crap'. The engineers were the 'experts', and historic data suggested that their estimate was probably accurate, but organisational decision-making was carried out as though the management's estimate was correct. Vaughan (1996), in examining the *Challenger* incident, shows in detail how organisations can develop such complacency.

Management believed the incredible because they wanted to. But how can confidence be

derived in very low probabilities? In some cases, such as for the operation of multiple identical systems (e.g. the breaking systems of motor cars), it may be possible to accumulate both the periods of operation and the numbers of failures and thus extrapolate into the future. But in most cases, adequate statistical data are not available for the accurate assessment of the probabilities of rare events, and it could take years to discover if the assumptions on which estimates are based are valid, or even reasonable.

In fault trees, the subjectivity of the derivation of the probability figure for the top event may be ignored or forgotten, but the result is no less subjective and no more accurate for that.

Not only the input probability data is subjective. Fischhoff, Slovic and Lichtenstein (1978) have shown that the construction and use of fault trees are also subject to variability. They point out that in the construction of a fault tree such questions arise as, Which faults should be identified separately and which should be lumped under 'others'? Which items should be grouped together? What sort of graphic display should be used? What level of detail is most appropriate? Answers depend on a number of factors, including the purpose of the analysis and how much of an effect each branch of the tree or each component is thought to have. Thus, the construction of a fault tree is subjective, and a tree prepared for a given purpose by one person is unlikely to be reproduced by another under the same circumstances.

Omissions of relevant pathways in a fault tree are possible because of ignorance, poor memory, and lack of imagination, among other causes, and such omissions would lead to understatement of the relevant probabilities. Fischhoff, Slovic and Lichtenstein (1978) concluded from their study that in the creation of a fault tree humans are likely to be biased in favour of information readily available to them (the availability bias, see below), for example because it has recently been in the newspapers, and that when omissions do occur people are, in the main, insensitive to them. In their study, this was found to be the case not only in tests of groups of college students but also of groups of experts (car mechanics). Such insensitivity to omission occurred in the case of the submarine *Tireless*, as discussed above.

Fischhoff, Slovic and Lichtenstein (1978) also found that the perceived importance of a particular branch of a fault tree was increased if it was represented in pieces (i.e. as two separate component branches). Thus, the probabilities determined were dependent on the construction of the tree, which in turn was subject to human frailties and biases.

But to what extent are numeric inaccuracies significant? In a simple example, Freudenburg (1992) shows that in some cases they can make a huge difference to the results. Suppose, Freudenburg suggests, that in drawing up a fault tree, analysts had arrived at a failure probability of 10^{-6} (one in a million), having identified all but two risk factors - one of which made the system more and the other less safe. And suppose that the system would operate at the 10^{-6} risk level for 80% of the time, but the 'real' risk would in fact include operation at 10^{-3} for 10% of the time and 10^{-9} for the other 10%. It might at first be assumed that the two errors would cancel each other out in the risk calculation, but this is not the case. The actual calculation produces a result of: $(0.1 \times 10^{-9}) + (0.8 \times 10^{-6}) + (0.1 \times 10^{-3}) = 0.0001008001$, which is slightly greater than 10^{-4}. Thus, the omitted higher risk (for 10% of the time) is not cancelled by the omitted lower risk. Rather, the higher risk dominates the result, even though it exists for only 10% of the time, and its omission leads to a distorted belief in the safety of the system.

It should be pointed out that such a result is typical when the probabilities must be combined by an OR gate. If it were necessary for two or more events to occur in order for the top event to occur, then the correct combination would be by an AND gate and the overall probability would be decreased. This is the principle of the 'protection system'.

6 RISK ASSESSMENT (OR EVALUATION)

The hazard identification and hazard analysis stages of the risk-analysis process are concerned with deriving risk values. Risk assessment is concerned with determining the tolerability of the risks. In some cases, tolerability may be assessed only on the basis of the risk values but, in general, a number of other factors are considered, for example the benefits to be gained and the cost of reducing the risk. What is tolerable depends on the circumstances and is a subjective decision based on values as well as technological information. Comparing risks against benefits is certainly a subjective process, for what is a benefit to one person is anathema to another. Gould et al (1988) point out that risk assessment is a matter of public values and therefore political, and Redmill (2000) has shown that many questions must be asked and that they can only be answered subjectively.

But the political process does not, in general, appear capable of arriving at risk-tolerability decisions competently. Risks that are acceptable to one person may be quite unacceptable to others, and the reasons for the differences seem to be psychological, social, and cultural. Risk tolerability depends on how the risk is perceived, and this, in turn, is a function of many variables, such as whether the risk is voluntarily taken, whether there is control over it, and whether it has fearfully large consequences (e.g. Slovic, Fischhoff and Lichtenstein 1985). When risk-tolerability decisions are based only on likelihood and consequence, and imposed on the public, they are often resented and opposed.

Risk-assessment is the most obviously subjective stage of risk analysis, but it is not the purpose of this report to investigate it in detail.

7 PROBLEMS IN THE ASSESSMENT OF HUMAN RELIABILITY

The sources of risks extend over a broad spectrum. At one end are those which arise predominantly out of random hardware failures. Given appropriate historic data, the likelihood of failure of the relevant components may be derived with more or less accuracy. When these are used to determine the probability of system-level failures (top events), accuracy will depend on many subjective variables, such as the way in which fault trees are constructed and which risk factors have been identified.

Moving along the scale, there is complex hardware (including modern electronics) in which systematic failures play a significant role. Systematic failures are those that would always be repeated under the same conditions. They arise not from wear and tear, but from specification and design faults, and they are not random.

Then there is software, in which all failures are systematic. If a systematic fault were discovered during testing, it would be corrected, so why are they not all found? Software is inherently complex, and the high number of logical paths through even relatively simple

programmes means that it would be impossible to test exhaustively in reasonable (or even, in some cases, in finite) time. When systematic faults are significant or dominant, quantitative analysis is inappropriate for it carries the assumption of randomness. Qualitative analysis is necessary and, although it provides a structured, transparent approach, its basis is judgmental and includes considerable subjectivity.

Further along the spectrum are those systems that also include humans. It may be that human error is sometimes random. For example, exhaustion at the end of a long or arduous period of work, or on a night shift, may approximate to the wear-out of a component. Human error can also be systematic, for example when a person repeatedly makes the same mistake as the result of having been given the wrong information. But humans can also cause problems for numerous other reasons, for example by sabotage, acting on a whim, making the wrong decision, neglecting to carry out a task, and failing to deal adequately with complexity. Thus, satisfactory modelling of human behaviour is not trivial or routine. Capricious behaviour is unpredictable.

A great deal has been learned about the propensities for humans to err and the ways in which errors occur - see, for example Reason (1990) and Lucas (1997) - and an influential model for explaining the mechanisms for error is Rasmussen's (1983) framework. This is based on three types of error made by persons supervising operational equipment: skill-based, rule-based, and knowledge-based errors. In fact, this model has been so influential that it is often used, outside of its intended scope, as a general statement about the functioning of humans.

A more general and widely applicable model of human error is used by Kletz (2000). He defines four categories, and points out that causes of an incident may include more than one of them:

- Mistakes - errors that occur because someone does not know what to do;
- Violations - errors resulting from someone who knows what to do deciding not to do it, usually for what appears to be a good reason;
- Mismatches - errors caused by a person's inability to carry out a task;
- Slips and lapses - errors due to lapses of attention.

Given the body of theory on the causes and mechanisms of human error, it was natural that there should be attempts to create models for the derivation of numeric values for its likelihood. Many probabilistic models for human reliability assessment (HRA) exist, and reviews are offered by Caccuabue (1997) and by Moieni, Spurgin and Singh (1994). Probability theory, however, is based on randomness, and, as randomness can, at best, offer only a partial explanation of human behaviour, the application of the techniques to human error introduces assumptions that may not maintain in any given case. Indeed, the approach taken to deriving probabilities is based principally on professional judgement.

One way of removing the possibility of human error is sometimes perceived to be to automate the tasks performed by operators. If this could be achieved, it would remove the need to assess the probability of operator error and its difficulties. But it turns out to be unfeasible. First, there are numerous occasions when the human is needed in the chain of operations, particularly in emergencies when judgements, often based on inadequate information, are required. Another reason for unfeasibility is that even if in some cases the operator's tasks could be wholly automated, the scope for human error would then be passed to the equipment designer and so would not be removed. Moreover, as Bainbridge (1987) has

argued, automation can increase rather than reduce the problems facing a human operator. She points out that the operator is left with those tasks that the designer could not automate. The operator's role becomes supervisory, often with reduced information. Moreover, supervision is a task that is not conducive to humans for they find it boring. Due to less training and less hands-on experience, operators become de-skilled but are still expected to resume control and take time-critical decisions when the automated system fails - often in an emergency.

All these factors contribute to the difficulty of estimating the probability of human error. This difficulty leads not only to inaccurate results but also to the exclusion of human error from many risk analyses. Even when engineers attempt to include human factors, they are likely not to be familiar with the available methods. Further, they often do not have guidance on how to address the subject. For example, the recent international safety standard (IEC 2000) devotes Part 2 to the risk analysis of hardware, Part 3 to that of software, and Part 6 to giving informative advice on using Parts 2 and 3, but, although it exhorts its users to 'consider human factors', it offers no comparable advice on how to do so.

But not only is human risk analysis difficult, also the existing probabilistic models do not cover the full range of human behaviour. Dougherty (1990) challenged the HRA community to answer his question, 'HRA, where shouldst thou turn?' and to examine their first generation human reliability assessment methods for the fundamental flaws to which he drew attention. Hollnagel (1996) complains of a lack of cohesion in the field of human reliability assessment and of the obsolescence of the state of the HRA art. Lydell (1997) agrees and pointed to a reliance on techniques that were developed in the 1970s. He cites a lack of leadership and management competence in the research community for the failure to progress in the field, and calls for a 'broad-based planning group' to be established, with adequate funding from international research administrations.

Swain (1988) had said that all HRA models had serious limitations, that the task of calibrating the models had not been seriously addressed, and that they are often ill-founded relative to human behaviour. Dougherty (1990) agreed and asserted that 'inadequate HRA modelling can lead to increased risk or wasted risk management resources.' He called for second-generation methods to be developed and for advances in error psychology and cognitive science to be accommodated within the HRA framework.

Moieni, Spurgin and Singh (1994) accepted Dougherty's claim that a second generation of HRA methods is required. They said that HRA is at the threshold of change - from the purely expert-judgement methods towards a more balanced approach using 'a combination of experimentally derived data and insights (using both large-scale training simulators and small-scale simulations) coupled with the use of formalized expert judgement methods.' They also recognised the need for research in the area of assessment of influence of management and organisational factors on human reliability.

However, although Hollnagel (1996), having reviewed some first-generation HRA methods, claimed that second-generation methods were emerging, he saw little consistency in approach. The reason, he claimed, was that the emphasis is not on what the second-generation methods should be but, following Dougherty (1990), on what they should not be - they should not be like first-generation methods. Hollnagel (1996) echoed Dougherty (1990) in saying that 'if HRA is going to meet the growing demands from end users, it is necessary that second generation HRA methods take heed of the significant developments in the modelling of

cognition.'

Hollnagel (1997) indicates that no new methods have yet taken the places of the first-generation methods, and he reiterates Lydell's (1997) call for a convergence of researchers and practitioners to establish an improved future. HRA is not contributing fully to risk analysis, and the indications are that it will be some time before it does. It is understandable that early attempts to assess human reliability should have been based on existing reliability theory, which led to the development of probabilistic models. But now there is recognition that these do not adequately represent human behaviour, and there has been a great deal of research into human behaviour, resulting in a deeper understanding of it. Thus, not only is there a need for a new generation of models, based more nearly on human behaviour itself, but it is now also possible to develop them. There are research opportunities in this field, not only in the development of new, more appropriate, human-risk-assessment models, but also in the integration of human risk analysis into the risk analysis of overall systems.

8 HOW RISKY IS MANAGEMENT?

Risk analysis techniques are relatively well developed for equipment which is subject only to random failures, less so when it comes to systematic failures, and much less so in the face of human behaviour. As it is now known that the majority of accidents have a significant element of human cause, the value of risk analyses that do not include the human components of systems may be questioned.

As discussed above, human reliability assessment (HRA) techniques attempt to address human error, but in the main they are limited to the humans who are directly concerned with equipment and whose errors can be final causes of accidents. Of course, equipment operators are important. Having, in his poem, *The Secret of the Machines*, let the machines tell of the wonderful things that they can accomplish, Rudyard Kipling then makes them add:
'But remember, please, the law by which we live,
We are not built to comprehend a lie,
We can neither love nor pity nor forgive.
If you make a slip in handling us you die!'

But the human causes of accidents do not all (or even, perhaps, in most part) stem form the obvious human 'components of systems' - the equipment operators. It is now recognised that the policies, strategies, and cultures, defined (or created by default) and engendered by senior management, predispose accidents to happen or not to happen. In each of the enquiries, into the sinking of the Herald of Free Enterprise (Steel 1987), the Kings Cross Underground fire (Fennell 1988), the Clapham Junction rail crash (Hidden 1989), the Piper Alpha oil rig disaster (Cullen 1990, Pate-Cornell 1993), and many other accidents, senior management failure was concluded to be a primary cause.

Kletz (2000) distinguishes management failure from human error, and says that it occurs because senior managers do not realise that they could do more to prevent accidents. Kletz contrasts the fact that management perceive no need to get involved in the detail of safety issues with their close attention to output, costs, and even product quality. He also contrasts the effort expended on different causes of accidents with their actual importance (see Figure 2) to show the anomaly in the consideration by senior management of their own failures.

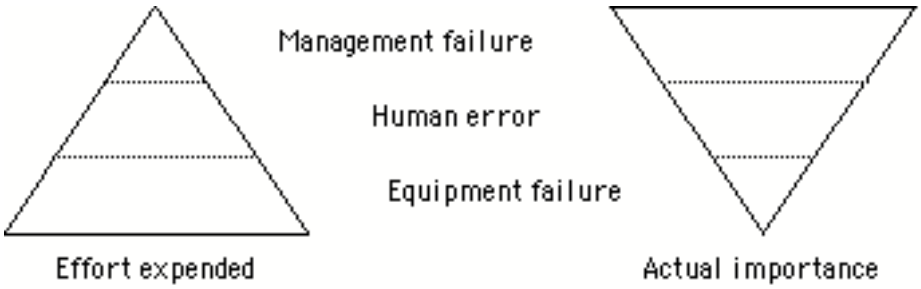

Figure 2: The effort expended on different causes of accidents and their actual importance

Yet, it is almost unheard of for a risk analysis to address the influence of senior management. HRA techniques address only operator error (the probabilities of which the techniques attempt to 'measure'), and Crawford (2000) points out that this bias towards what can be measured must be expected to lead to optimistic results - contrary to what we normally aim to do in risk analysis. Thus, not only is HRA inaccurate in its attempt to model and predict operator error and inadequate in its coverage of human behaviour, but it also fails to address the most significant source of human risk - that of management.

Vaughan (1996) confirms the role of management in accidents, but points to organisational issues and how mistake and disaster are 'socially organised and systematically produced by social structures'. She says that the cause of the Challenger disaster in 1986 was 'a mistake embedded in the banality of organisational life' and she shows how 'deviance in organisations is transformed into acceptable behaviour'. The features of organisational life amount to the culture of the organisation, and the role of senior management is to engender and nurture 'good' and 'strong' culture (Levene 1997) so as to ensure that such transformations do not take place. When they do take place it points to management's ineptitude, negligence, aberrance, or misjudgement. The importance of culture is also emphasised by Reason (1997) who offers useful advice on managing the risks of organisational accidents.

It is also worth remarking that the development of a safety culture, and, indeed, of appropriate processes for reducing risks depends on the ability and willingness of humans - and, in particular, management - to learn lessons from past mistakes and incidents and to apply them. Miller (2000) tells us that the most significant human error is the continuing failure to apply the lessons of past mishaps to prevent future accidents. He says this in the context of aviation systems, but it may easily be taken generally. Kletz (1993) offers a reason for this, saying, 'It might seem to an outsider that industrial accidents occur because we do not know how to prevent them. In fact, they occur because we do not use the knowledge that is available. Organisations do not learn from the past or, rather, individuals learn but they leave the organisation, taking their knowledge with them, and the organisation as a whole forgets.'

Risk analysts go to great effort and expense to determine quantitative estimates of the likelihood of the final triggers of hazardous events - mainly equipment failures. Sometimes they assess operator causes. But almost invariably they totally ignore those predisposing factors, created by management, which in many cases pose the greatest risks. By ignoring management issues - and, in particular, senior management issues - risk analyses are predisposed to be poor predictors of accidents.

There is a great and urgent need, and considerable scope for, research into how management

risks may be modelled, measured and predicted. But with risk analysis being a relatively well established process, why should it require research to improve it rather than merely the introduction of improved procedures? Because working-level engineers and risk analysts are not competent or positioned to analyse the risks introduced by senior management. In general, they could not be expected to be sufficiently familiar with the sources of the risks - company policy, strategic plans, management style, organisational culture - to be able to carry out hazard identification and analysis studies. Indeed, they are unlikely even to have access to some of the essential information. Further, as investigation of the risks under discussion would not be within the ambit of anyone carrying out risk analysis in the traditional way, research is necessary in order to deduce ways of estimating the likelihood of management hazards developing into disasters.

Initial suggestions are that boardroom audits are needed, and that both they and company audits should address the issues of risk and safety insofar as they impinge on the organisation. This implies that auditable issues need to be defined, and these should be the consideration of safety and risk by the most senior management, not only in their policies and strategies but also in the ways in which they ensure and monitor their implementation.

Thus, at board meetings the standing agenda should include topics critical to risk and safety. In some cases there may be a requirement for environmental policy, in others the crucial issue might be the safety of workers, or the public, or the users of particular products. Further, because there is typically a lack of senior management awareness of whether or how their policies and strategies are implemented, there needs to be encouragement for improvement in this deficiency by the inclusion on the standing agenda of how management ensures and monitors their implementation. Then, the minutes of board meetings should be audited to ensure that these topics are properly addressed, and company audits should investigate the way in which actuality accords with the board's minutes. A high-level risk analyst would then bring the boardroom and company audits together with 'traditional' risk analyses so as to produce an improved analysis which would be a better predictor of risks and more in keeping with the requirements of a modern 'risk society' (Beck 1992).

These are only initial thoughts, but they might be contemplated by researchers intent on improving the process of risk analysis. If great technical changes were made to the existing process, they might improve it marginally. For risk analysis to become a better predictor of organisational risk, as well as a better indicator of where risk reduction is needed, the issue of the risks posed by senior management must be addressed.

There is a hint that action may be imminent - not in the assessment of management failure, but in its reduction. Senior management may be pushed into doing something about their lack of awareness of, and lack of involvement in, safety issues. The Institute of Chartered Accountants in England and Wales, acting on the recommendations of the Turnbull Report, have produced a code on internal control, and guidelines on conforming to it, in order to improve corporate governance in the business community (ICAEW 1999). Elliott et al (2000) describe the key features of the code as focusing on:

- The protection of shareholders' investments and company assets through sound systems of control;
- Regular reviews of the effectiveness of controls, specifically on finance, operations, compliance and risk management;
- The need for companies regularly to review the need for an internal audit function, where

one is not present;

That internal control is to be taken seriously is evinced by the fact that the London Stock Exchange has demanded that its members explain in their annual reports and accounts how they have applied these and other principles. As operational risk management is included in the code, there is a chance that senior managers' awareness of safety risks will improve in the coming years.

This may reduce the risks of accidents caused by management failure, but it will not explain how such risks can be assessed and included in system risk analyses. There is a need for research into this, in addition to which there is also scope for research into the motivators and demotivators for organisational learning. The field is not limited to the investigation of serious incidents, for there is considerable scope for learning from small incidents and near hits, as shown by Koornneef (2000).

9 MENTAL HEURISTICS

Psychological research has revealed that humans are subject to biases that result from the ways in which the brain handles information, particularly in the context of decision-making. One example is the tendency to take the 'inside view' as discussed above. Being humans, risk analysts suffer from these biases, and exploring them could provide an understanding of how they introduce subjectivity into the process of risk analysis.

Humans are subject to vast amounts of information from disparate sources. In decision-making it is likely that the volume is too great for a thorough analysis in the time available. Moreover, a thorough analysis (given the implication of the term in the context of modern analytical theory) is likely to be beyond the competence of the majority of people in all but the simplest situations, even if time was not limited.

So humans take mental short-cuts and make approximations in order to cope with disparate information. They use mental rules and strategies, referred to as heuristics, for decision making. The status of humans on the planet suggest that heuristics served us well in our evolution (for example in fight-or-flight decisions), but there is evidence that in more complex situations they can induce biases which distort our perception of risk and our decision-making processes. Particularly because mental heuristics and the biases that they induce are unconscious, it would be useful to summarise the effects that they could have.

9.1 Prospect Theory

Economic and games theories are based on the assumption that human individuals are 'rational', that is, that given a choice, they will always select the option expected to maximises the outcome - they choose the maximum 'expected value'. This is referred to as 'utility theory'. For example, suppose that there is a choice between two gifts: (a) a certain £50, or (b) a 60% chance of getting £100 and a 40% chance of getting nothing. The expected value of (a) is 50 x 1 = £50, and that of (b) is 100 x 0.6 + 0 x 0.4 = £60, so normative theory would predict that every individual would select (b). Yet Kahneman and Tversky (1979) found that most people choose the certain gain and shun the gamble, thus suggesting themselves to be risk-averse (at

least in the face of gain).

In exploring this and other 'anomalies', Kahneman and Tversky (1979) developed their 'Prospect Theory' (extended in Tversky and Kahneman (1992)) to explain them. Prospect Theory does not dismiss utility theory, but modifies it to accommodate the expressed preferences of humans. It expresses outcomes as positive or negative deviations (gains or losses) from a neutral reference outcome with a value of zero, and the value function is roughly S-shaped, with monotonic increase and decrease, but with the slopes being convex above the reference point and concave below it. Because of the convexity and concavity, the difference in subjective value (the value as perceived by an individual) for gains is greater at low values (e.g. between $10 and $20) than at high values (e.g. between $110 and $120) and the same thing holds for losses. Individuals therefore tend to place a higher value on reducing a risk by a given amount at low levels (e.g. to zero - to eradicate it) than on reducing it by the same amount at a higher point on the value curve.

A further attribute of the subjective value function is that the descent of the losses part is steeper than the ascent of the gains part, showing that the displeasure associated with losing a given amount is greater than the pleasure at winning a corresponding sum. It might be added, however, that the excess in value to a person of a given amount of money that might be lost, over the same amount that might be gained, is not merely subjective. Consider the person who has only $10. To keep it offers the person certain options; to lose it would deprive that person of any choice at all. To gain another $10 would offer even more options than before, but the increase from $10 to $20 does not represent as much value as that from $0 to $10. For this reason, an evens bet is not desirable.

Another modification which Prospect Theory makes to expected utility theory is that outcomes are weighted not by their probabilities but by a 'decision weight' which might be considered to be a 'subjective probability'. The result of this is that low probabilities (except zero) are overweighted (i.e. people are over-concerned about unlikely but high-consequence events) and high probabilities (except certainty) are underweighted.

Prospect Theory does not suggest that all people act identically. It does propose, however, to be predictive of the human tendency and thus to indicate the majority choice. It therefore provides not only an indication of the biases that are likely to exist in a risk analyst but also why an analyst might expect to find certain attitudes to risk in others.

9.2 Framing

Kahneman and Tversky (1979) proposed, in their 'Prospect Theory', that rather than leaping from problem to decision in a single step, humans have a two-stage decision-making process. The first stage is concerned with analysing, editing, and perhaps reframing the problem and the possible acts and outcomes, and the second stage with evaluating the options and arriving at a decision. Kahneman and Tversky found that decisions depend heavily on how problem information (e.g. the available options and their outcomes) is expressed, or 'framed', in the first stage of the two-stage decision-making process.

In one now classic experiment, Tversky and Kahneman (1981) achieved a complete preference reversal by framing the possible outcomes of the options for a public health

programme first in terms of lives saved and then in terms of lives lost. First they expressed the options as follows. The country is preparing for the outbreak of an unusual disease that is expected to kill 600 people. Experts' estimates of the outcomes of two alternative programmes to combat it are: if programme A is adopted, 200 people will be saved; if programme B is adopted, there is a one third probability that 600 people will be saved and a two thirds probability that nobody will be saved. It can be seen that the expected values of the two outcomes are the same. 72% of those questioned opted for Programme A and 28% for programme B. Then the experimenters asked the same question, framed in a different way, of another group of people. The country is preparing for the outbreak of a disease that is expected to kill 600 people. Experts' estimates of the outcomes of two alternative programmes to combat it are: if programme A is adopted, 400 people will die; if programme B is adopted, there is a one third probability that nobody will die and a two thirds probability that 600 people will die. It can be seen that the options are the same as those in the first statement of the question, yet 78% opted for Programme B and 22% for programme A. The difference in framing of the question led to a complete reversal of the decision - from 72% for one programme to 78% for the other. From risk aversion, when a positive outcome was perceived, to risk seeking when the perception was of a negative one.

The implications of framing are considerable. Considering frames as communicative rather than cognitive constructs, Drake and Donohue (1996) examined the records of situations in which councillors sought to achieve conflict resolution between disputants. Dake and Donohue found that a disputant would choose (not necessarily consciously) to frame issues, using a choice of language, so as to highlight certain aspects of the communication while ignoring others. The other disputant would then do the same, perhaps ignoring the first disputant's frame. In each case the choice of language provided a verbal cue to the other participant, who could respond by converging or diverging from the other's chosen frame. The researchers found a positive relationship between the convergence of frames and the frequency of agreements.

Shah, Domke and Wackman (1996) tested two groups of subjects for how their decision-making was affected by the framing of information. The groups, of evangelical Christians and university undergraduates, were expected to hold different views. The decisions to be made concerned voting for one or other of election candidates. The information presented to the groups consisted of simulated news stories about the candidates' views, using 'ethical' and 'material' frames. It was found that the media frames had a pronounced influence on the interpretation of the issue in hand and on how the voters perceived other issues within the same environment - the 'priming effect'. If ethical framing was activated, it helped to foster an ethical interpretation of other issues - for both groups of subjects. The material frame had a corresponding effect. Further, voters were likely to put the frame at the centre of their evaluation for decision-making. Thus, the frame was not only influential on the subject's interpretation of the particular issue, but also on subsequent decision-making - in the form of voting. An individual's judgement became biased by the framing of the information.

The role of framing is now well established. It was recognised that in the collection of information for a specification, not one but several (perhaps numerous) frames are necessary, for no single one could represent all the required information. Using the pseudonym of 'perspectives' the CORE (controlled requirements expression) method (SD Scicon 1989) provides a defined set of procedures for acquiring the necessary information from all the relevant perspectives so as to fill the corresponding frames and thus gather all the necessary

information for a complete specification.

The importance of framing on the presentation and communication of risk information is clear and far-reaching. Risk analysts are likely to be influenced by the framing of information and, as Drake and Donohue (1996) showed, the framing is certain to be influenced by the information's source. Thus, a person in favour of accepting a risk is likely to frame the risk information differently from a person who opposes the risk. Consider, for example, the instance of Shell's desire to dispose of their obsolete oil platform *Brent Spa* in the North Sea and Green Peace's opposition to this (Wilkinson 1997). A risk analyst needs to be alert to the possibility of bias, otherwise it is likely to occur unconsciously. Indeed, Crawford (2001) argues that efforts should be made to ensure several perspectives in an analysis.

Similarly, if risk information is be provided to decision makers in a neutral form, risk analysts and others need to take great care in its preparation, presentation and communication.

9.3 Availability

The 'availability' heuristic (Tversky and Kahneman 1973) concerns how readily information that is relevant, or thought to be relevant, to the decision in hand comes to mind. Its effect is to increase the judged likelihood of an event if the event is easy to recall or imagine. When availability is informed by an individual's experience of the frequency of events (for example, in a purely local context), it may be an accurate basis of judgement, but when it is conditioned by factors unrelated to frequency, it can be highly misleading. A vivid news item, or a recently seen film, about a disaster can distort judgement. Similarly, if the media gives wide coverage to every occurrence of a low-frequency high-consequence event, it is likely to be perceived as being more frequent or likely to occur than it is. For example, studies show that the public believes murder to be more frequent than suicide, whereas the reverse is the case.

Discussion of a low-probability, and therefore rarely occurring, event would raise awareness of it and, as Slovic, Fischhoff and Lichtenstein (1980) observed, lead to an increase in its judged probability. Risk analysts and other engineers and scientists must make expert judgements not only on the probabilities or frequencies of events in absolute terms, but also on the relative likelihood of one event compared with another. The availability bias can affect both types of judgement.

Lichtenstein S, Slovic P, Fischhoff B, Layman M and Combs B (1978) report finding the availability bias in their experiments. Ross and Sicoly (1979) explored it in the context of egocentricity and found that spouses overemphasised their own responsibilities for occurrences (including negative ones such as arguments) within their marriages, and basketball players ascribed responsibility for the outcomes of games to the actions or inactions of their team-mates rather than to those of members of the opposing team. Taylor (1982) proposed ways in which the availability heuristic could cause biases, including its effect on the way in which information is retrieved from memory.

9.4 Overconfidence

A lack of awareness of the assumptions on which judgements are based, and a failure to

appreciate what is not known, lead to the bias known as 'overconfidence'. Fischhoff, Slovic and Lichtenstein (1977) asked subjects to place odds of being correct in their judgements of which of two lethal events was more frequent and, although odds of 100:1 or greater were named more than 25% of the time, about one in eight of the judgements was incorrect.

A warning to risk analysts is Slovic, Fischhoff and Lichtenstein's (1980) observation that 'experts seem as prone to overconfidence as lay people', and they cited a number of instances where the bias contributed to accidents. Indeed, Henrion and Fischhoff (1986), in a study of scientific work, found significant underestimation of errors and the likelihood of errors. Oskamp (1965) had previously found that psychologists' confidence in their supposed knowledge and understanding of their patients, as information about them increased, was 'not a sure sign of increasing accuracy' of their conclusions and that 'their certainty about their own decisions became entirely out of proportion to the actual correctness of those decisions.'

Counciling his father, Creon, Haemon said:
'good as it is to have infallible wisdom,
Since this is rarely found, the next best thing
Is to be willing to listen to wise advice.' (Sophocles 441BC)
A feature of overconfidence is that it encourages a belief that the overconfident person's judgement is infallible and that advice from another is unnecessary.

It is often pointed out that the majority of people believe themselves to be above average drivers. Risk analysts are not immune to the overconfidence bias, and experts' trust in their probability estimates is typically a function of how much information they have gathered rather than of its accuracy or its predictive success (Shrader-Frechette 1991). Given this bias, it is difficult to agree that experts have an ability to determine 'real' risk while lay people are limited to being able only to derive perceived or 'subjective' risk.

9.5 Representativeness

A further heuristic is 'representativeness', defined by Tversky and Kahneman (1971) and Kahneman and Tversky (1972) as a subjective judgement of the extent to which the event in question 'is similar in essential properties to its parent population', or 'reflects the salient features of the process by which it is generated'. Representativeness, as Bar-Hillel (1982) indicates, may introduce two kinds of systematic error into judgements. First, it may give undue influence to variables that affect representativeness but not probability, and second, it may reduce the apparent importance of variables crucial to determining probability but unrelated to representativeness. The importance of such a bias in making expert judgements about the likelihood of future events is inescapable.

The law of large numbers offers high confidence that very large samples will be highly representative of the population from which they are drawn. Tversky and Kahneman (1971) suggest that people are likely to believe that a sample is representative - that is, it possesses all the attributes - of its parent population. They say that 'people's intuitions about random sampling appear to satisfy the law of small numbers, which asserts that the law of large numbers applies to small numbers as well.' A consequence of this would be that a scientist would have exaggerated confidence in the validity of conclusions based on small samples. Quite often risk analysts only have small samples at their disposal, particularly of rare

events, and an awareness of the representativeness bias could lead to a re-evaluation of their initial conclusions about what a sample implies. Indeed, Tversky and Kahneman (1971) showed not only that 'experts' are as subject to this bias as lay persons, but also that even when their error is explained to them they do not alter their judgement. So the bias cannot be - or is not easily - unlearned.

Bar-Hillel (1982) investigated and demonstrated the heuristic in three experiments, as did Tversky and Kahneman (1982). Tversky and Kahneman (1973) showed that when provided with information about, for example, a person or situation (the input), we make predictions about the person or situation by selecting outcomes that appear to be most representative of the input.

9.6 Review

Heuristics are a normal aspect of human reasoning processes. As humans do not have full statistical information in their brains at the time that decisions are needed, they must employ what information they do have. Further, even if they did have a great deal of meaningful information, processing it would often be an impossible task, both because they could not do so in the time available and because in many cases they would not be able to deal with its complexity anyway. So rules, or heuristics, for reducing complex problems to simpler, more easily solvable ones, for selecting the information that seems relevant, and for basing solutions on this information, are essential. Thus, heuristics, which evolved in humans long before probability theory was developed, serve them well in many situations. But in more complex decision-making they can turn into biases that distort the judgement of likelihood (Tversky and Kahneman1974). As Bar-Hillel points out, they are 'less than perfectly correlated (if, indeed, at all) with the variables that actually determine the event's probability.' And Covello and Menkee (1982) observe that, ' The need to reduce cognitive strain and anxiety often leads to unrealistic oversimplifications of essentially complex problems and to 'solutions' that are more apparent than real.' Nor are experts less subject to these basic tools of the human brain than lay people. Slovic, Fischhoff and Lichtenstein (1980) observe that 'we have no assurance that experts' judgements are immune to biases once they are forced to go beyond their precise knowledge and rely upon their judgement. Although judgmental biases have most often been demonstrated with lay people, there is evidence that the cognitive functioning of experts is basically like that of everyone else.'

Risk involves uncertainty, so the need for experts to 'go beyond their precise knowledge and rely on judgement' is implicit in the task of risk analysis. Thus, it is important for analysts and those planning analyses to recognise and understand the biases that could, and are likely to, afflict them. When risk analyses are planned, it would be as well to consider introducing not only means of counteracting known biases but also procedures to check and ensure that they have been applied. Examples given earlier include the use of a team for hazard identification and making sure that the team members have appropriate backgrounds and different perspectives. Further research is required to define other anti-bias procedures and build then into the process of risk analysis.

10 COMMUNICATION AND INTERPRETATION OF RISK

The purpose of risk analysis is to derive information to inform decision-making. So the benefit of the analysis is not merely the derivation of risk values, or even of 'accurate' risk values, but also, importantly, of values of appropriate risks. Then, as decision-making is usually done by someone other than the risk analyst, it is necessary to communicate the results of the analysis to the decision maker - and choices of what information to communicate and how to communicate it are made. Further, however 'good' or 'correct' the communicated information, it is subject to interpretation by the person receiving it. Although risk communication is not the principal subject under consideration, it is worth mentioning it briefly.

There are many ways in which risk information may be inappropriate. In the first place, it may simply be wrong. At the Bristol Royal Infirmary in the UK, between April 1990 and August 1993 a surgeon carried out heart operations on thirteen young babies to correct Atrioventricular Septal Defects (AVSD). Seven died (GMC 1998a, 1998b). Yet the surgeon told the parents of the next patient that the risk of mortality was 20 - 25%. Where could such a figure have come from? At the time, not only was the death rate of the surgeon's AVSD-operation patients greater than 53%, but also the trend was not encouraging: not only had the last 3 patients died, but so had 5 of the last 6, and 7 of the last 9. The parents approved the operation and the baby died. Yet the surgeon told the parents of the next patient that the chance of success of the operation by him was 80%. Perhaps the surgeon suffered from the overconfidence bias referred to above and gave a figure based on his confidence in himself rather than on an assessment of the risks based on his past performance. But certainly the communicated information was incorrect with respect to the facts, and inappropriate to the parents' decision of whether or not to allow their child to be operated on by the surgeon.

It is not only to the medical profession that the public looks for information for their decision-making. All professionals hold specialist information and are expected to be competent to provide advice based on it, including advice on risk. In general, borrowers don't know whether their credit is good or bad, plaintiffs don't know whether their case is strong or weak, picnickers don't know what the weather will be like, and they all look to professionals for advice on which to base their decisions. It is therefore important that the professionals provide information that is unbiased and that allows their clients to make informed rather than coerced decisions.

Risk information may be inappropriate if it is not the information required to inform the decision needing to be made. Risk figures are often given in averages, particularly to the public. Yet, average figures can be misleading. The physician who considers the average patient rather than the one before him is likely to be far off the mark in both diagnosis and treatment.

In civil aviation it is reported that there is about one fatal crash per million flying hours (Howard 1991). This may be useful as a reference point against which to monitor the air-worthiness of new aircraft prior to licensing them, on the principle that a new aircraft should be at least as safe as those already in service. But it does not help the public to decide which planes, routes, or airlines to avoid. If it is true that crashes are most frequently associated with takeoffs and landings, crash figures per flight would make it apparent that someone who travels a long distance on a few long flights is exposed to a smaller risk than someone who

travels a much shorter total distance on many flights.

Road-transport risk is often presented in terms of an average mortality rate per year. But this does not inform any decision that an ordinary person may wish to make. It does not distinguish between a driver's chance of being killed, killing another road user, or being in an accident in which a passenger is killed. It does not facilitate a decision on what time of the day or night it is safe to venture out as a driver or pedestrian, what type of person is most likely to have an accident, or which types of roads have the highest and lowest accident rates. To propose an average risk figure is to suggest that there is a common risk to all drivers regardless of their gender or age, the time of day, or their geographical location, but this is not the case, as shown by Adams (1995).

Bier (2001a) concludes that 'Little research has been done on effective methods of communicating risk analysis results to decision-makers.' There is a serious need for such research. At the same time, having reviewed the field, Bier (2001b) offers a number of suggestions on risk communication to the public. It should be remembered that the importance of risk information to the public is to inform decisions, and so 'the public' are decision-makers as much as are more elevated political figures and technological experts.

In addition to providing objective information, it may also be deemed necessary to attempt to influence the behaviour of others (say, vehicle drivers) by the provision of information. In this respect it may be more worthwhile to publicise the number of persons per year who are rendered cabbages in road accidents than the number killed. Not only are the numbers larger, but people are likely to be more fearful of a lifetime of paralysis than of sudden and perhaps painless extinction.

Risk information may also be inappropriate because it lacks meaning. It is not unusual to hear it said (for instance, in the context of bovine spongiform encepalopathy) that 'the risk is one in a million'. But one in a million what? Does this mean that one person in every million people in the country or the world will die of the cause in question? Or that of every million people who come into contact with the source one will die? Or is there another probabilistic implication? Further, even if it is made clear what the probability factor refers to, the risk information is still not meaningful without a 'defined and dimensioned exposure figure' (Clemens 2001). Is the exposure interval a year, or a day, or the lifetime of the persons involved? Or might it be per hour of exposure to the risk?

There is also the issue of how information is presented. The simple information that, on average, one in a million of the population will die of cause A each year, might be presented in a number of forms, for example: 1 per 1,000,000 per year, 1/1000000 per year, 10^{-6} per year, or 1 person in a city the size of Birmingham per year. Which is used should be determined from knowledge of whom it is desired to communicate the information to. Not only will the recipients' understandings depend on how the information is presented (framed), but it is also necessary to be aware of any assumptions made about the ability of recipients to understand the information communicated to them. A scientist may flippantly say that a certain carcinogen in a food or drink, taken each day in moderation, increases the average person's risk of cancer over their lifetime by one in a million. A member of the public may reply, 'Yes, but is it safe?'

Risk information first needs to be the right information for the purpose in hand and then it

needs to be communicated effectively. School pupils are taught to tailor the style, length, tone, and format of documents to suit their intended readership, and not to write an essay on the life of a penny in the same way as a report on a laboratory experiment. These lessons are re-emphasised in further education and industrial training. Yet risk analysts and the producers of risk statistics appear often to be oblivious to any intended use of their output.

Risk information is also inappropriate when it is selected and communicated to give a false impression. Huff (1965) retells a story of statistical 'proof' provided by the (US) Air Force that jet flying was safer than flying in conventional aircraft. The proof consisted of figures that showed that the death rate, in fatalities per 100,000 flying hours, was higher in ordinary planes than in jets. As the air-force jets in question carried one or two persons while the conventional aircraft carried many passengers, many more were at risk per flying hour in the latter. As Huff points out, if genuine information had been the aim, the figures would better have been presented in terms of deaths per million man-hours or man-miles rather than plane-hours.

Although Huff's story shows that risk communication can (successfully) be used to mislead, it should also be pointed out that when the recipients of the information detect bias or an ulterior motive, they are likely to reject the transmitted message. Providing risk information in terms of a comparison - e.g. 'the risk is about the same as that of eating a hundred grams of chocolate per day' - may be enlightening if the recipients understand the risk being referred to. Yet it has been established (see for example Roth et al 1990, Freudenburg and Rursch 1994) that if the recipients believe that the intention is to persuade them of the tolerability of the risk in question, the result is likely to be that they understand the communication but reject the risk.

Thus, not only must risk information be appropriate to its context and purpose, but its framing must also be appropriate and not patronising. The information's purpose needs to be understood, and it should be derived, presented and communicated accordingly. Recipients of risk information should question its origin, its accuracy, the perspective from which it was derived, and whether the person communicating it believes it. They need to ask how appropriate it is for their own purposes. Risk information is unlikely to be accurate, but it needs to be sufficiently accurate for its purpose. The person communicating it needs to know the limits of its accuracy and to communicate them to the recipient in a way that the recipient understands.

This last point has never been insignificant, but its importance is increasing. A great deal of risk discussion, and indeed controversy, falls within areas of scientific uncertainty. Funtowicz and Ravetz (1992) say that when both the uncertainty of facts and the stakes are high, scientists in quest of a solution need to use a language in which the uncertainty and quality of information are made explicit. Yet, as Willis (2001) points out, it is unusual for organisations communicating risk information to the public to be open about the scientific uncertainty involved. Rather, they tend to play it down while pointing to the lack of evidence of risk. However, as Willis affirms, such communication is now not often credible to the public, who loose trust in the organisation as a result of it. The field of dealing with risk under scientific uncertainty is beyond the scope of this study, but it is of increasing importance and offers the opportunity for research.

In the end, there is no single correct mode of risk communication. It has been shown that the

public is quite capable of understanding, and estimating, risk frequencies. Further, although they resent patronisation, they do not automatically reject risk comparisons. Thus, adequate risk communication requires an assessment by the communicator of what the most appropriate method is. And to do this, the communicator needs, above all, to be a good communicator.

11 DISCUSSION

Some see no merit in risk analysis. Others defend it vigorously. Some, such as Shrader-Frechette (1991) see a middle path where it is in some respects objective but at the same time subject to the biases of its practitioners and the inaccuracies of its methods.

As shown above, the process of risk analysis involves subjective judgement at every stage, and its tools, such as hazard and operability studies (HAZOP) and fault tree analysis (FTA), themselves involve considerable subjectivity. Ansell (1992) admits that sometimes the results of risk analysis are founded on assumptions which 'may not be valid', and Ballard (1992) concedes that 'Frequently there will be uncertainty in such information concerning the physical process, equipment reliability, human actions, etc.'

Ballard goes on to say that, 'Risk analysis serves to highlight uncertainties so that their effect can be appreciated rather than hidden in superficially exact rules or judgements', and this, surely, is a crucial point. Risk analysis is not an exact or objective process but a tool for arriving at approximate risk values so as to inform decision-making. It is a tool which, like all tools, should be used within its limitations and with an understanding of its assumptions. It should not be considered an end in itself, and nor should its results be the only basis on which decisions are make.

Judgement is required not only in carrying out risk analysis, but also in using its results - which it would be wrong to portray as definitive. Ansell (1992) says, 'it is only plausible to produce a relative ranking of the likelihood of events, rather than accurate assessments.' But even these estimates can be of considerable value - as long as we accept Schneiderman's (1980) advice and recognise that the numbers 'are likely to be crude'.

Certain risks, for example those posed by genetically modified organisms, carry huge uncertainty, such that numeric values for probabilities and even for consequences are mostly speculative. In such cases attempts to carry out quantitative hazard analysis are at best optimistic, and they can be misleading because numbers are often mistaken for accuracy. But the majority of risks subjected to analysis are concerned with the operation of equipment, the appropriateness of processes and the manner of their performance, and the safety of products. These, in most cases, are better understood. This does not mean that they don't carry uncertainty or that they are immune to subjectivity. They do and they are. However, in these contexts, risk analysis, whether quantitative or qualitative, is, in the main, effective in leading to risk reduction. Evidence for this exists in the vast number of industrial plants in operation, and products on the market, which are not perceived by the public as risk issues. Risk analysis is often carried out badly, and, even when carried out well, its results may only be approximate. But that suggests that its use should be improved, not that it should be discarded.

Thus, carrying out risk analysis is preferable to not doing so. Indeed, international safety standards (e.g. IEC 2000) demand it, and UK law requires most businesses to produce documented risk analyses. The subjectivity identified above is, as well as carrying the many problems discussed, one of the principal strengths of the process. Identifying and analysing hazards and making decisions about risks demand human thought and human probing. If the process were automated, it would not benefit from the human ability to probe and to take the situation as it is rather than as a programmer some time previously generalised it as being. But human delving means that something is always likely to be missed and some things may be wrongly judged. Thus, it is important for risk analysts to understand the subjective influences on the process, including the ways in which their own backgrounds and biases can affect its results.

But, given that subjectivity is an essential component of risk analysis, what might be done to minimise its undesirable effects on the results? There is scope for applying lessons learned in the risk research carried out in the social sciences, for example in the fields of risk perception, decision making, and risk communication.

We might consider setting standards on how risks and benefits should be defined in an analysis and how the definitions should be structured and presented to the decision-making process. As seen above, the framing of information has a considerable influence on decision outcomes. If in risk analysis we defined the framing of information as a requisite component, and we trained analysts to understand the possible influences of framing, it would have a beneficial effect on the way in which the analysts themselves perceived risks. It would require them to seek alternative perspectives before making judgements. It would also affect the way in which they present and communicate information, perhaps with explanations, to decision-makers. There is scope for research and improvement in this field.

It was also seen above that information may arrive from disparate sources with varying degrees of trustworthiness (Funtowicz and Ravetz 1990). It would be useful to pause at this stage, review the information and its sources and their trustworthiness, decide when further information is required and where it should come from, and place trust factors on the information. The HSE (1992, 1999) literature makes no mention of the need for expressions of confidence in the risk values arrived at, but it would be of benefit if, as a matter of course, risk analysts determined and documented the levels of confidence to be attached to their results. Research into this issue could open up a new dimension in risk analysis. In some cases it might also be advisable to reframe the problem, or the options and their expected or believed outcomes, so as to arrive at new perspectives.

Such initial analysis could be designed to raise awareness of various points, including: the importance of perception and its influence on opinion, the framing of questions and possible outcomes, the understanding of the pedigree of data sources and the assessment of confidence, the presentation and communication of information, and the ways in which decisions are arrived at. A guideline could be provided on these and other issues. Then, there would be an increased chance that biases would be neutralised during the analysis and management of risk.

To achieve these aims, it would be necessary to re-examine risk-analysis syllabuses. If risk analysts are to benefit from such disciplines as risk perception and risk communication, these need to be included in both academic and industrial courses. Indeed, the emphasis needs to be shifted from training to education, and the breadth of both risk analysis and its academic basis

needs to be extended to reflect this. Then, not only would engineers and scientists be better educated in the field, and risk analyses better performed, but also debates on the subject would be better informed - at least on the engineering side. Kasper (1980) says, 'The experts' approach to their differences with the rest of the public has been to somehow persuade the public to alter its perceptions so that they more nearly approximate the calculated or projected results. (I know of no serious effort directed toward shifting experts' views so that they will coincide with those of the rest of the public.)' Perhaps, on better understanding the subjectivity that is inherent in risk analysis, and, indeed, the subjectivity that they themselves bring to it, the experts would wish to shift their own views somewhat. In all debates there is need for concession on both sides, and the risk debate is no exception. Engineers and other risk analysts, like the rest of the public, are subject to biases. We need to be seen to want to understand the 'other side' better.

This discussion has so far considered how risk analysis, as it stands, might be improved in response to a better understanding of the subjectivity to which it is prone. However, it is also worth recognising that the risks addressed by traditional risk analysis are often less significant than those introduced by senior management via policy, strategy, management style, organisational inertia, and the culture that is developed intentionally or by default. If the process of risk management is to fit itself better for the modern technological environment, ways must be found to encompass senior management and organisational issues. Some initial thoughts on how that might be addressed have been given above, and research into how they might be brought to fruition is needed.

At the same time, the methodical approach of engineering and scientific risk analysis could usefully be applied in other fields. UK law now calls for risk analyses in all business and public service organisations of more than five employees, and many if not most organisations are struggling to cope with this demand. Technology transfer is called for. Risk-analysis techniques could beneficially be transferred from the scientific and engineering domains to other areas where there is a recognised need for risk analysis.

In particular, some 'soft' industries, for example the medical and prison services, in which the risks at issue are those posed by humans, could benefit from a more methodical and transparent approach to risk analysis. This would not only improve analyses but also facilitate risk-tolerability decisions and make the reasoning in the decision-making process available to subsequent reviews.

While it includes subjectivity at every stage, risk analysis is vital to industry. Indeed, the subjectivity is a principal strength of the process. However, there is much in the research of the social scientists for the engineers and scientists to learn. At the same time, there is much that others could gain by applying the methodical approach taken by engineering risk analysts.

12 ACKNOWLEDGEMENTS

I acknowledge with thanks the comments of Jim McQuaid, Trevor Kletz, Jack Crawford, Gabe Mythen, and Odd Nordland on drafts of this report.

13 REFERENCES

Adams J (1995). *Risk*. London, UCL Press

Ansell J (1992). Reliability: Industrial Risk Assessment. In Ansell J and Wharton F (eds): *Risk Analysis, Assessment and Management*. John Wiley & Sons, Chichester

Bainbridge L (1987). The Ironies of Automation. In Rasmussen J, Duncan K and Leplat J (eds): *New Technology and Human Error*. John Wiley & Sons, Chichester

Ballard G M (1992). Industrial Risk: Safety by Design. In Ansell J and Wharton F (eds): *Risk Analysis, Assessment and Management*. John Wiley & Sons, Chichester

Bar-Hillel M (1982). Studies of Representativeness. In Kahneman D, Slovic P and Tversky A (eds): *Judgement Under Uncertainty: Heuristics and Biases*. Cambridge University Press

BBC (2000). *File On Four*. BBC Radio 4, 12 December

Beck U (1992). *Risk Society*. Sage Publications, London

Bier V M (2001a). On the State of the Art: Risk Communication to Decision-makers. *Reliability Engineering and System Safety*, 71, 151-157

Bier V M (2001b). On the State of the Art: Risk Communication to the Public. *Reliability Engineering and System Safety*, 71, 139-150

BSI (1991a). *BS 5760 Part 5: Guide to Fault Mode and Effects and Criticality Analysis (FMEA and FMECA)*. British Standards Institution, UK, 1991

BSI (1991b). *BS 5760 Part 7: Guide to Fault Tree Analysis*. British Standards Institution, UK, 1991

Cacciabue C (1997). Human Reliability Assessment: Methods and Techniques. In Redmill F and Rajan J (eds): *Human Factors in Safety-critical Systems*. London, Butterworth-Heinemann

Charette R N (1989). *Software Engineering Risk Analysis and Management*. McGraw Hill, 1989

CIA (1977). *A Guide to Hazard and Operability Studies*. Chemical Industries Association Limited (reprinted 1992)

Clemens P L (2001). The Risk Exposure Interval - Too Often an Analyst's Trap. *Journal of System Safety*, 37, 1, First Quarter

Covello V T and Menkee J (1982). Issues in Risk Management. In: Hohenemser C and Kasperson J X (eds.): *Risk in the Technological Society*. Westview Press, Boulder

Crawford J (1999). *What's Wrong With The Numbers? - Part 1*. Safety Systems (The Safety-

Critical Systems Club Newsletter), 9, 1, September

Crawford J (2000). What's Wrong With The Numbers? - Part *2*. *Safety Systems (The Safety-Critical Systems Club Newsletter),* 9, 2, January

Crawford J (2001). Some Ways of Improving Our Methods of Qualitative Safety Analysis, and Why We Need Them. In Redmill F and Anderson T (eds): *Aspects of Safety Management - Proceedings of the Ninth Safety-critical Systems Symposium, Bristol, UK*. London, Springer-Verlag

Cullen, The Hon. Lord (1990). *The Public Enquiry into the Piper Alpha Disaster*. HMSO, London

Deming W E (1975). On Probability as a Basis for Action. *The American Statistician*, 29, 4, 146-152

Department of the Environment (1995). *A Guide to Risk Assessment and Risk Management for Environmental Protection*. HMSO, London

Dougherty E M Jr (1990). Human Reliability Analysis - where shouldst thou turn? *Reliability Engineering and System Safety*, 29, 283-299

Drake L E and Donohue W A (1996). Communicative Framing Theory in Conflict Resolution. *Communication Research*, 23, 5, 297-322

Elliott D, Letza S, McGuinness M and Smallman C (2000). Governance, Control and Operational Risk: The Turnbull Effect. *Risk Management: An International Journal*, 2, 3, 47-59

Fennell D (1988). *Investigation into the King's Cross Underground Fire*. HMSO, London

Feynman R P (1989). *What Do <u>You</u> Care What Other People Think?* Unwin Hyman, UK

Feynman R P (1998). *The Meaning of it All*. Penguin Books

Fischhoff B, Slovic P and Lichtenstein S (1977). Knowing with Certainty: The Appropriateness of Extreme Confidence. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 552-564

Fischhoff B, Slovic P and Lichtenstein S (1978). Fault Trees: Sensitivity of Estimated Failure Probabilities to Problem Representation. *Journal of Experimental Psychology: Human Perception and Performance,* Vol 4, No. 2, 330-344

Freudenburg W R (1992). Heuristics, Biases, and the Not-So-General Publics: Expertise and Error in the Assessment of Risks. In Krimsky S and Golding D (eds): *Social Theories of Risk*. Westport, Praeger

Freudenburg W R and Rursch J A (1994). The Risks of 'Putting the Numbers in Context': A Cautionary Tale. *Risk Analysis*, 14, 6, 949-958

Funtowicz S O and Ravetz J R (1990). *Uncertainty and Quality in Science for Policy*. Kluwer, Dordrech

Funtowicz S O and Ravetz J R (1992). Three Types of Risk Assessment and the Emergence of Post-Normal Science. In Krimsky S and Golding D (eds): *Social Theories of Risk*. Westport, Praeger

GMC (1998a). *Charges, Case No 1A: Wisheart, James Dunwoody. Ameded draft version 21/5/98*. Professional Conduct Committee Hearing. General Medical Council, London

GMC (1998b). *DETERMINATION in the Case Against James Dunwoody Wisheart, Thursday 18 June*. Professional Conduct Committee Hearing. General Medical Council, London

Gould L C, Gardner G T, DeLuca D R, Tiemann A R, Doob L W and Stolwijk J A J (1988). *Perceptions of Technological Risks and Benefits*. New York: Russell Sage Foundation

Henrion M and Fischhoff B (1986). Assessing Uncertainty in Physical Constants. *American Journal of Physics*, 54, 791-798

Hidden A (1989). *Investigation into the Clapham Junction Railway Accident*. HMSO, London

Hollnagel E (1996). Reliability Analysis and Operator Modelling. *Reliability Engineering and System Safety*, 52, 327-337

Hollnagel E (1997). Reply to 'A Practitioner's View of the State of HRA Methodology'. *Reliability Engineering and System Safety*, 55, 3, 261-262

Howard R W (1991). Breaking Through the 10^6 Barrier. *International Federation of Airworthiness Conference*, Auckland, New Zealand

HSE (1992). *Safety Assessment Principles for Nuclear Plants*. HMSO, London

HSE (1999). Health and Safety Executive: *Reducing Risks, Protecting People.* Discussion Document, HSE Books

Huff D (1965). *How to Take a Chance*. Pelican Books

ICAEW (1999). Internal Control: Guidance for Directors on the Combined Code. *The Institute of Chartered Accountants in England and Wales*, London

IEC (1985). *International Standard IEC 812: Analysis Techniques for System Reliability: Procedures for Failure Mode and Effect Analysis*. International Electrotechnical Commission, Geneva

IEC (1990). *International Standard IEC 1025: Fault Tree Analysis (FTA)*. International Electrotechnical Commission, Geneva

IEC (2000). *International Standard IEC 61508: Functional Safety of*

Electrical/Electronic/Programmable Electronic Systems. International Electrotechnical Commission, Geneva

Janis I L (1982). *Victims of Groupthink*. Second Edition, Houghton-Mifflin, Boston

Kahneman D and Tversky A (1972). Subjective Probability: A Judgement of Representativeness. *Cognative Psychology*, 3, 430-454

Kahneman D and Lovallo D (1993). Timid Choices and Bold Forecasts. A Cognitive Perspective on Risk Taking. *Management Science*, 39, 17-31

Kahneman D and Tversky A (1979). Prospect Theory: An Analysis of Decision Under Risk. *Econometrica*, 47, 263-291

Kasper R G (1980). Perceptions of Risk and their Effects on Decision Making. In Schwing R C and Albers W A Jr. (eds): *Societal Risk Assessment — How Safe is Safe Enough?* Plenum Press, New York

Kletz T (1993). *Lessons from Disasters*. Institution of Chemical Engineers, London

Kletz T (1999). *HAZOP and HAZAN*. Fourth edition, Institution of Chemical Engineers

Kletz T (2000). *An Engineer's View of Human Error.* Third edition, Institution of Chemical Engineers, London

Koornneef F (2000). *Organised Learning from Small-scale Incidents* (PhD thesis). Delft University Press

Levene T (1997). Getting the Culture Right. In Redmill F and Dale C (eds): *Life Cycle Management for Dependability*. Springer-Verlag, London

Leveson N G (1995). *Safeware: System Safety and Computers*. Addison-Wesley, Reading, Mass

Lichtenstein S, Slovic P, Fischhoff B, Layman M and Combs B (1978). Judged Frequency of Lethal Events. *Journal of Experimental Psychology: Human Learning and Memory*, 4, 551-578

Lowrance W W (1980). The Nature of Risk. In Schwing R C and Albers W A Jr. (eds): *Societal Risk Assessment — How Safe is Safe Enough?* Plenum Press, New York

Lucas D (1997). The Causes of Human Error. In Redmill F and Rajan J (eds): *Human Factors in Safety-critical Systems*. Butterworth-Heinemann, Oxford

Lydell B (1997). A Practitioner's View of the State of HRA Methodology. *Reliability Engineering and System Safety*, 55, 3, 257-260

Miller C O (2000). The Most Significant Human Error in the Aviation System. *Journal of System Safety,* 36, 1, 11-18

MISRA (1994). *Development Guidelines for Vehicle Based Software*. The Motor Industry Research Association, UK

MOD (1996). *Interim Defence Standard 00-58: HAZOP Studies on Systems Containing Programmable Electronics*. Ministry of Defence, Glasgow, UK

Moieni P, Spurgin A J and Singh A (1994). Advances in Human Reliability Analysis Methodology. Part 1: Frameworks, Models and Data. *Reliability Engineering and System Safety*, 44, 27-55

Okrent D (1998). Risk Perception and Risk Management: on knowledge, resource allocation and equity. *Reliability Engineering and System Safety*, 59, 1, 17-25

Oskamp S (1965). Overconfidence in Case-study Judgements. *The Journal of Consulting Psychology*, 29, 261-265

Pate-Cornell M E (1993). Learning from the Piper Alpha Accident: A Post-mortem Analysis of Technical and Organisational Factors. *Risk Analysis*, 13, 215-232

Railtrack (2000). *Engineering Safety Management Guidance (The Yellow Book)*. Issue 3, Railtrack, on behalf of the UK Rail Industry

Rasmussen J (1983). Skills, Rules, Knowledge: Signals, Signs and Symbols and Other Distinctions in Human Performance Models. IEEE Transactions on Systems, Man and Cybernetics, SMC-13, 3, 257-266

Reason J (1990). *Human Error*. Cambridge University Press

Reason J (1997). *Managing the Risks of Organizational Accidents*. Ashgate Publishing, Aldershot

Redmill F (2000). How Much Risk Reduction is Enough? *Journal of System Safety*, 36, 1 - also translated into Japanese and included on the website of Japan International Center for Occupational Safety and Health (JICOSH)

Redmill F, Chudleigh M F and Catmur J R (1997). Principles Underlying a Guideline for Applying HAZOP to Programmable Electronic Systems. *Reliability Engineering & System Safety*, 55, 3, 283-293

Redmill F, Chudleigh M and Catmur J (1999). *System Safety: HAZOP and System HAZOP*. John Wiley & Sons, Chichester, UK

Redmill F (2001). Subjectivity in Risk Analysis. In Proceedings of *Risk Analysis and Safety Management of Technical Systems*, Gdansk, Poland, 25-27 June

Ross M and Sicoly F (1979). Egocentric Biases in Availability and Attribution. *The Journal of Personality and Social Psychology*, 37, 322-336

Roth E, Morgan M G, Fischhoff B, Lave L and Bostrom A (1990). What Do We Know About Making Risk Comparisons? *Risk Analysis*, 10, 3, 375-387

Schneiderman M A (1980). The Uncertain Risks We Run: Hazardous Materials. In Schwing R C and Albers W A Jr (eds): *Societal Risk Assessment - How Safe is Safe Enough?* Plenum Press, New York

SD Scicon (1989). *CORE - The Method*. SD Software Technology Centre, Issue 2

Shah D V, Domke D and Wackman D B (1996). 'To Thine Own Self Be True': Values, Framing, and Voter Decision-making Strategies. *Communication Research*, 23, 5, 509-560

Shrader-Frechette (1991). *Risk and Rationality*. University of California Press

Slovic P, Fischhoff B and Lichtenstein S (1980). Facts and Fears: Understanding Perceived Risk. In: Schwing R C and Albers W A Jr (eds): *Societal Risk Assessment: How Safe is Safe Enough?* Plenum Press, New York

Slovic P, Fischhoff B and Lichtenstein S (1985). Characterising Perceived Risk. In Kates R W, Hohenemser C and Kasperson J X (eds): *Perilous Progress: Managing the Hazards of Technology*. Westview Press, Boulder

Sophocles (441 BC). *Antigone* (721-723). Translation by Watling E F (1947). Penguin Classics

Steel D (1987). *Formal Investigation into the MV Herald of Free Enterprise Ferry Disaster*. London, HMSO

Storey N (1996). *Safety-Critical Computer Systems*. Addison-Wesley, London

Swain A D (1988). Adapting Risk Analysis to the Needs of Risk Management. Paper presented at *World Bank Workshop of Risk Management and Safety Control*, Washington D.C.

Taylor S E (1982). The Availability Bias in Social Perception and Interaction. In Kahneman D, Slovic P and Tversky A (eds): *Judgement Under Uncertainty: Heuristics and Biases*. Cambridge University Press

Tversky A and Kahneman D (1971). Belief in the Law of Small Numbers. *Psychological Bulletin, 2, 105-110*

Tversky A and Kahneman D (1973). Availability: A Heuristic for Judging Frequency and Probability. *Cognitive Psychology*, 4, 207-232

Tversky A and Kahneman D (1974). Judgement Under Uncertainty: Heuristics and Biases. *Science*, 185, 1124-1131

Tversky A and Kahneman D (1981). The Framing of Decisions and the Psychology of Choice. *Science*, 211, 453-458

Tversky A and Kahneman D (1982). Judgements of and by Representativeness. In Kahneman D, Slovic P and Tversky A (eds): *Judgement Under Uncertainty: Heuristics and Biases*. Cambridge University Press

Tversky A and Kahneman D (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty*, 5, 297-323

UK-ILGRA (1996). United Kingdom Interdepartmental Liaison Group on Risk Assessment, Use of Risk Assessment within Government Departments.

Vaughan D (1996). *The Challenger Launch Decision*. University of Chicago Press, Chicago

Veseley W E, Goldberg F F, Roberts N H and Haasl D F (1981). *Fault Tree Handbook*. US Nuclear Regulatory Commission, NUREG-0492, Washington, 1981

Wharton F (1992). Risk Management: Basic Concepts and General Principles. In: Ansell J and Wharton F: *Risk Analysis, Assessment and Management*. John Wiley & Sons, Chichester

Wilkinson A (1997). Perceptions and Authority. *Science, Policy and Risk*. Royal Society, London

Willis R (2001). *Decision-making Under Scientific Uncertainty: The Case of Mobile Phones*. Green Alliance, London

Wynne B (1980). Technology, Risk, and Participation: On the Social Treatment of Uncertainty. In Conrad J (ed): *Society, Technology, and Risk*. New York, Academic Press

Wynne B (1982). *Rationality and Ritual: The Windscale Inquiry and Nuclear Decisions in Britain*. Chalfont St Giles: British Society for the History of Science